# Understanding the Directed Evolution of *de novo* Retro-Aldolases from QM/MM Studies

**Daria De Raffele, Sergio Martí,* Vicent Moliner.***

Departament de Química Física i Analítica; Universitat Jaume I. 12071 Castellón (Spain)

Corresponding Authors:

S. Martí: smarti@uji.es

V. Moliner: moliner@uji.es

ABSTRACT

In an era in which climatic change puts at risk the planet, the study and develop of alternative green chemistry which can help and improve our life can play an essential role. In this context, the use of artificial enzymes capable of substitute traditional industrial processes by environmental friendly routes is a challenge. Unfortunately, the complete understanding of the catalytic activity and selectivity of enzymes remains still elusive, thus hampering creation and development enzymatic proteins. In this paper, the molecular mechanism of the non-natural multistep retro-aldolase reaction catalysed by a *de novo* biocatalyst, the RA95.5-5, has been investigated by means of multiscale QM/MM methods. The design of a retro-aldolase presents the difficulty to create an enzyme being able to stabilize several transition states, maintaining low energy barriers along the overall reaction. The obtained QM/MM free energy landscape has allowed defining the rate determining step corresponding to the carbon – carbon bond scission of the substrate, which is in accordance with the experimental data. A deep analysis of the electrostatic interactions between the substrate and the different amino acid residues of the protein, as well as the estimation of the electrostatic potential generated on key atoms of the substrate, has been carried out for the key steps of the reaction. The results, compared with previous computational studies on the most efficient *de novo* retro-aldolase, the RA95.5-8F, explains the different activities achieved during the directed evolution process and provides insights for future developments of more efficient enzymes.

*Keywords:* retro aldolase, *de novo* enzyme, free energy, mutation, activation interaction energy

## INTRODUCTION

It was 1947 when Linus Pauling wrote that "no one succeeded in determining the structure of any enzyme nor in finding out how the enzyme does its job".[1] Some time later, James D. Watson stated that "proteins were daunting objectives, not only because of their size and irregularity but because the sequence of amino acids along their polypeptide chains was still unknown".[2] Since then, scientists have flirted with the idea of understanding the principles of biological catalysis, with the purpose to know how enzymes work and, ultimately, to design 'new' ones. After less than a century of Pauling and Watson's assertions, all the research performed has allowed designing new enzymes with catalytic activities close to those exhibited by natural enzymes, accomplished after millions of years of evolution. This milestone was achieved by a revolutionary and robust method to design proteins with desirable functions, by minimizing natural evolution through generating random mutations in the gene of interest.[3–12] The engineering of novel enzymes is a grand challenge, and the past decade has demonstrated that synthetic methods can be combined with computational tools to exploit together the generation of proteins with novel structural and chemical properties.[13–15]The use of theoretical simulations is becoming increasingly valuable due to their ability to provide an enormous quantity of information about chemical reactions in complex media,[13] addressing fundamental questions and evaluating amino acid sequences *in silico* on a scale that is experimentally impossible to achieve. Computational assisted protein design methods can be categorized into three main categories: bioinformatics, or analysis of primary sequences;[16,17] molecular modelling, or computer simulations of tertiary structures;[18] and methods based on the knowledge of the transition state (TS) structure.[19] This last group of methods can be divided according to the employed protein scaffolds: (i) immune-globulins, proteins that have been used to produce catalytic antibodies;[20] (ii) promiscuous proteins showing catalytic activity for more than one reaction that can be re-designed to enhance the activity for the secondary reactions; and (iii) proteins without specific catalytic properties that can be used as scaffolds to support the design of an active site from scratch, or *de novo* design.[21,22] Mayo and co-workers[23] described the power of an iterative approach combining *de novo* design, molecular modeling, and protein crystallography. D*e novo* protein design is experiencing rapid development,[24–37] becoming mainstream, with a substantial amount of articles employing this approach to some extent, many of them highlighting the importance of flexibility and conformational dynamics based in computational simulations.[27,36,37] These efforts have contributed to the protein *de novo* design becoming an accurate and reliable technique in this field,[24] fueled by the development of powerful computational methods.[38,39] For instance, the software SABER[40] has been used to explore and locate proteins with particular 3D structures having active sites that might be redesigned to catalyze new

reactions.[16] A different strategy is used by other programs such as ROSETTA,[41] that try to match an ideal 'theozyme' as coined by Houk and co-workers, to define the set of amino acids placed at adequate positions around the substrate, to provide TS stabilization[42] into an existing protein scaffold.[25,26,43,44] In fact, the first computationally assisted designs of new enzymes were reported in 2008 on retro-aldolases[39] and Kemp eliminases,[38] based on this strategy.

In this paper we are describing the results of a computational study of previously designed enzymes for the chemically demanding retro-aldol (RA) reaction. In particular, we are dealing with the fundamental carbon-carbon bond breaking in a non-natural substrate 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone, methodol, to form acetone and 6-methoxy-2- naphthaldehyde (6-MNA).[39] Since this is a reaction that proceeds in more than one step, the active site had to be modeled in such way that could simultaneously accommodate and stabilize different intermediates and TSs. Indeed, according to the original work of Jiang et al. on *de novo* computational design of RA,[39] four alternative composite active site motifs were modelled but all of them comprising a nucleophilic lysine and general acid/base groups capable of catalyzing various proton transfer steps involved in the reaction. Following further optimization and an assessment of the models, 72 of them with 8–20 amino acid changes in 10 different scaffolds were experimentally tested. 23 had significant retro-aldolase activity, with rate enhancements up to four orders of magnitude over the uncatalyzed reaction.[39]

To improve the designed retro-aldolase, successive studies[22,45–48] have been proceeded, focusing on the design enzyme RA95.0 ($k_{cat}/k_{uncat} = 7.7 \times 10^3$)[48], whose eightfold $(\beta\alpha)_8$ barrel structure belongs to the family involved in molecular or energy metabolism within the cell,[49] and it is considered the most common and versatile fold among naturally occurring enzymes.[50,51] The RA95.0[47] design was created in the $(\beta\alpha)_8$ scaffold by introducing 11 active site mutations,[43] comprising a reactive lysine residue at position 210, a hydrophobic pocket for substrate binding between the L1 and L6 ($\beta_1\alpha_1$ and $\beta_6\alpha_6$ respectively) and a glutamate at position 53 to orient the catalytic water molecule (for promoting proton transfers). Nevertheless, the activity of the enzyme was not perturbed after Glu53 was replaced by an alanine[48] showing that this residue doesn't contribute to the catalysis. After this artificial enzyme, the best variant, RA95.5,[47] contained six mutations in the active site (V51Y, E53S, T83K, M180F, R182M, and D183N). The new tyrosine substituted at position 51 was proposed to act as acid/base species. On the other hand, the lysine at position 83, as in many class I aldolases, would adopt the catalytic role of Lys-210. Moreover, mutagenesis studies showed that Lys-83 is more effective than Lys-210.[48] Just introducing the T83K mutation into the RA95.0 scaffold, the activity was increased by around five-folds[47]. In contrast, the replacement in RA95.5 of Lys-83 with

methionine leads to a reduction of the $k_{cat}$ and $k_{cat}/K_m$ by factors of 3 and 14, respectively. The same mutation on the Lys-210 did not affect the $k_{cat}$ but increased the $k_{cat}/K_m$ by two-folds.[48] Some other mutations (M180F, R182M and D183N) create a deep hydrophobic binding pocket by conformational changes. The most active clone after RA95.5, containing six additional mutations, is the RA95.5-5 (20-fold higher than RA95.5), in which three mutations are in or near the active site (E53T, S110N and G178S), while the rest of the mutations are on the surface of the protein (R23H, R43S and T95M). The asparagine at position 110 was proposed to assist the formation and breakdown of the carbinolamine intermediate through hydrogen-bonding interactions.[48] The lysine at position 210 was mutated into a methionine without any effect on catalytic efficiency. Further mutations led to RA95.5-8 with three additional active site mutations and two distal (K13N, S178V, G212D and S43R, F72Y, respectively) resulting in a new enzyme that is 60-fold more active. Recently, Hilvert and co-workers evolved RA95.5-8 into the highly active RA95.5-8F, whose activity is similar to that of natural class I aldolases.[45] The best variant is 30-fold more active than the predecessor and contains 13 additional mutations. The mutations on RA95.5-5 (close and far from the active site) to get the RA95.5-8F design are shown in Figure 1.

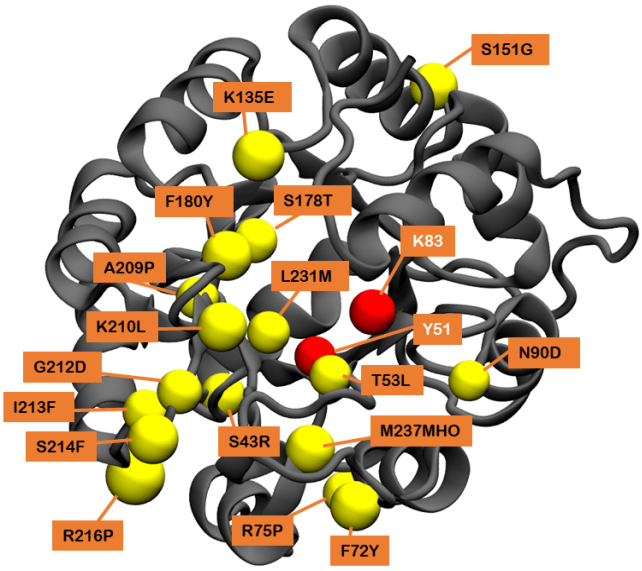
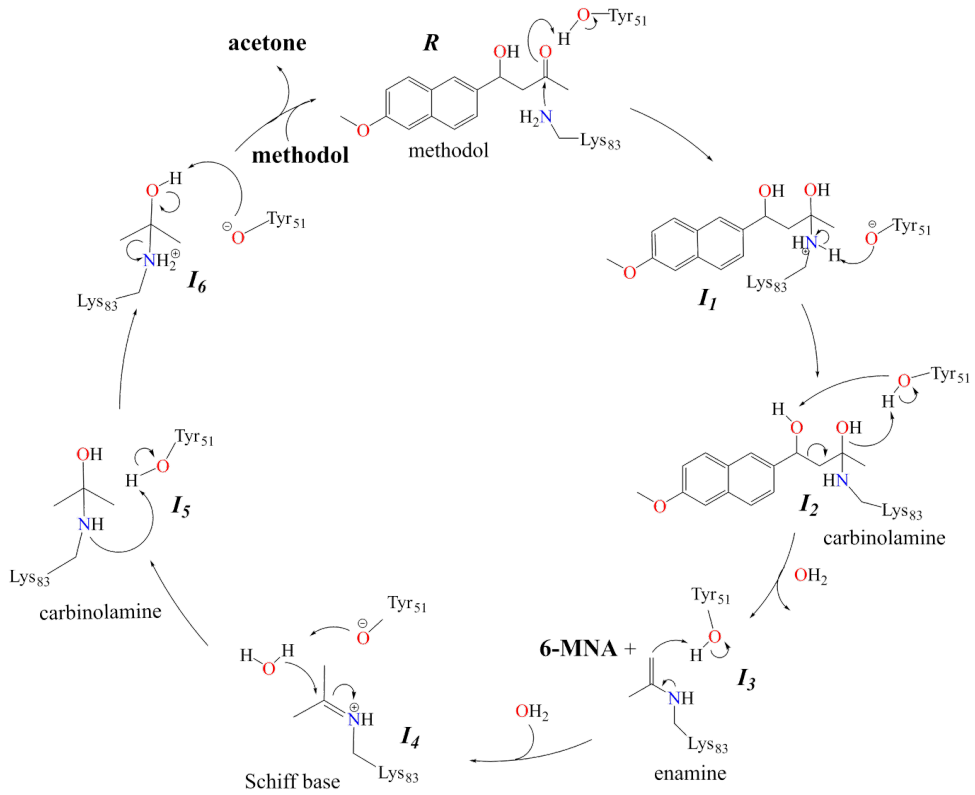



Figure 1. Representation of mutations introduced by Hilvert and co-workers in the RA95.5-5[46] to arrive to the most evolved RA95.5-8F[43] (yellow spheres). The two key active site residues, Tyr51 and Lys83, are indicated as red spheres.

Since the catalytic activity of artificial enzymes are usually much lower than that of natural enzymes,[9] several studies have been recently conducted to understand the behavior of the designed RAs, thus supporting the timeliness of the present study.[52,5354] In particular, we perform a multiscale QM/MM computational study on the reaction mechanism of the RA95.5 designed by Hilvert and co-

workers,[48,51] and compare the results with the more evolved RA95.5-8F design,[48] previously studied in our laboratory.[54] We chose a variant showing an activity clearly lower than RA95.5-8F (see Table S1 deposited in the Supporting Information), in order to confirm the mechanism but, more importantly, to compare a less evolved RA with the most efficient one. This way, key effects on catalysis can be more easily identified. Computational study of these two *de novo* enzymes with very different activity, and not so many different mutations, should render significantly different results to be analyzed. Thus, the goal is to identify the role of those amino acids that are not conserved between both designs, which allows defining their contribution to the activation free energy of the different steps along the full reaction process. The explored molecular mechanism was initiated on the one proposed in our previous QM/MM study of the RA catalyzed by the RA95.5-8F design, shown in Scheme 1.[54] Understanding the origin of the improvement between two designed retro-aldolases at atomistic level can be the bedrock to propose strategies to develop efficient new biocatalysts, especially in the case of multi-step reactions.




**Scheme 1**. Schematic representation of the proposed reaction mechanism of the retro-aldol reaction on methodol, catalized by the *de novo* designs RA95.5-5 and RA95.5-8F.

## COMPUTATIONAL METHODS

The initial coordinates to generate the molecular model of the full enzyme were obtained from the X-ray crystal structure of the *de novo* enzyme RA95.5-5 deposited in the PDB with the code 4A2R. The inhibitor present in the crystal structure, as well as part of the active site, was manually modified using the VMD program,[55] until a favorable (reactive) conformation. This was achieved by, by just adding a hydroxyl group to the C13 carbon atom (see Figure 2). The rest of the setting up of the model was carried out as described in our previous study,[54] Then, potential energy surfaces (PESs) were explored for every possible chemical step, using the fDYNAMO library.[56] Free energy surfaces (FESs) were generated afterwards in terms of potentials of mean force (PMF) using the combination of the weighted histogram analysis (WHAM)[57] and the umbrella sampling (US) methods,[58,59] as implemented in the fDYNAMO library.[56] A description of the collective variables employed to explore every chemical step depicted in Scheme 1 is deposited in the Supporting Information. The estimated error associated to this technique for these kind of systems is usually accepted to be ca 1 kcal·mol$^{-1}$.[60] As an example, a bootstrapping analysis [61] based on 50 samples for the 1D-PMF of the sixth step of the present study renders a rms value for the activation free energy of 0.7 kcal·mol$^{-1}$. More information on the PMF concept and its applications can be found in the literature.[56,62–65] In addition, it is important to point out that several alternative mechanisms have been also explored, rendering activation energies much higher than the ones obtained for the mechanism reported in Scheme 1 (see Figures S8-S10 deposited in the Supporting Information).

During the QM/MM calculations, the AM1 semiempirical Hamiltonian[66] was used to treat the quantum subset of atoms (see Scheme 1 and Figure 2). Once the first product was released (the 6-MNA molecule), the number of atoms treated quantum mechanically was reduced to the enamine, the side chains of Lys83 and Tyr51, and the generated water molecule. The rest of the protein−solvent system (43280 atoms) was represented classically by means of the OPLS-AA[67–69] and the TIP3P[70] water-model force fields. The link atom approach[71] was used for the QM-MM frontier treatment. Finally, an energy cubic spline function,[79–84] was applied in order to correct the possible deficiencies of the AM1 semiempirical Hamiltonian when generating the FESs, with the hybrid M06-2X functional with the 6-31+G(d,p) basis set,[78] using the Gaussian09 program.[79] See Supporting Information for details.

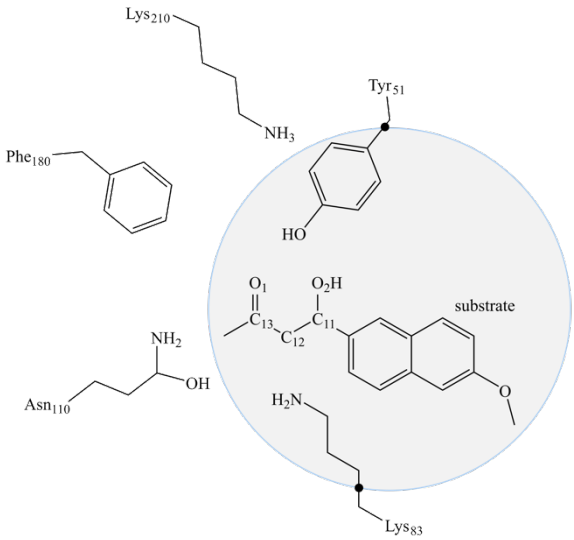



**Figure 2.** Schematic representation of the active site of the RA95.5-5 *de novo* enzyme. The gray region contains the atoms treated quantum mechanically. Link atoms are represented as black dots.

## RESULTS

The first part of this study was devoted to defining and exploring the whole reaction mechanism catalyzed by the *de novo* enzyme RA95.5-5, computing the corresponding FES for each of the proposed chemical steps (see Figure S8, S9 and S10 in Supporting Information for other alternative explored reaction mechanism paths studied). As for the RA95.5-8F design,[80] an acid−base reaction initiates the overall process. The 2D-PMF of the **first step** computed at M06-2X:AM1/MM level is presented in Figure 3, together with representative snapshots of reactants (R), the transition state ($TS_1$) and the intermediate ($I_1$). According to the free-energy surface shown in Figure 3a, the reaction occurs as a concerted process, with an activation free energy for this step of 11.9 kcal $mol^{-1}$. Based on the average key distances of the located structures of the involved states (see Figure 3 and Table S2 on Supporting Information), the reaction is confirmed to take place with the formation of the $N_{K83}$−C bond and the simultaneous protonation of the carbonyl group of the substrate by Tyr-51.

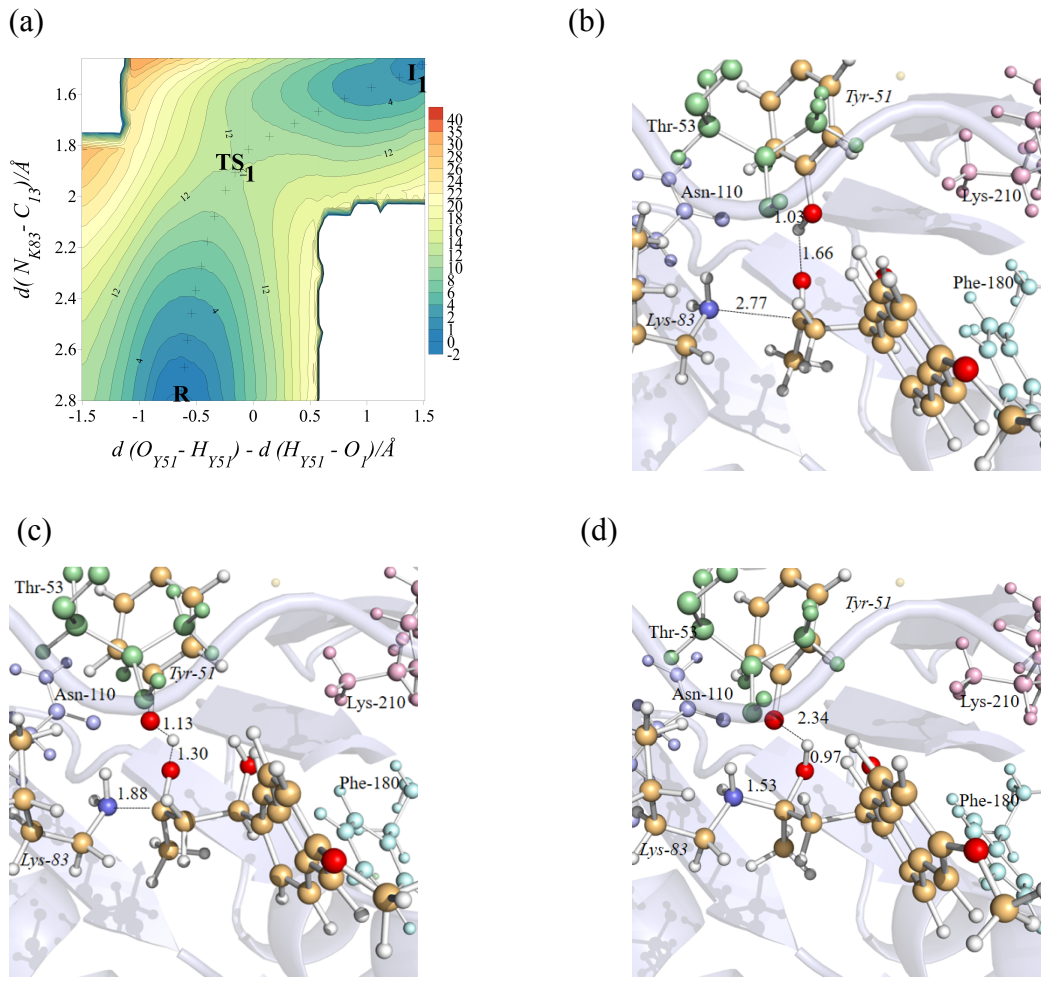



**Figure 3.** (a) M06-2X:AM1/MM FES for the first step of the reaction computed at 300 K. Representative snapshots of the (b) R, (c) $TS_1$, and (d) $I_1$. Values of energies are given in kcal $mol^{-1}$ and distances in Å.

In the **second step** of the reaction, the Tyr-51 deprotonates the positively charged Lys-83. According to the free energy surface shown in Figure 4a, this exergonic reaction proceeds with an activation free energy of 1.6 kcal $mol^{-1}$. The second intermediate, $I_2$, is 1.7 kcal $mol^{-1}$ more stable than $I_1$. The comparison of these results with the corresponding second step in RA95.5-8F,[54] shows that the mechanism is conserved, albeit a slighter higher energy barrier is obtained in the present system. An explanation can be due to the mutation done at position 180, where a phenylalanine presents in the RA95.5-5 is substituted by tyrosine in the RA95.5-8F variant.[45] The Tyr-180 creates a well-defined hydrogen-bond network with Tyr-51, Asn-110 and Lys-83, thus facilitating the formation of the $I_2$. A list of key averaged inter-atomic distances is deposited in Table S3 of Supporting Information.

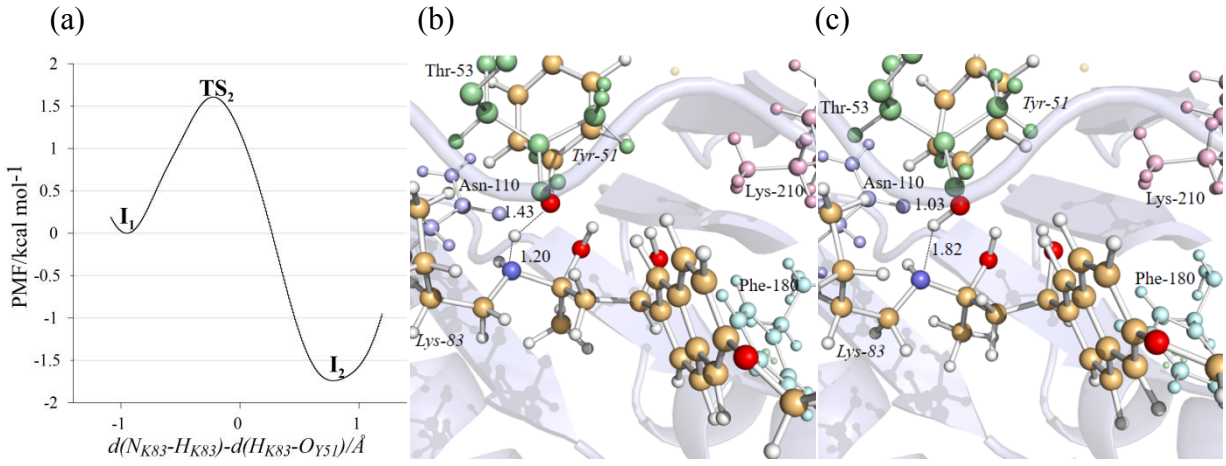



**Figure 4**. (a) M06-2X:AM1/MM FES obtained as a 1D-PMF at 300 K for the second step of the reaction. Representative snapshots of the (b) $TS_2$ and (c) $I_2$. Values of energies are given in kcal $mol^{-1}$ and distances in Å.

The **third step** leads to the formation of an enamine and the elimination of a water molecule, yielding the first product of the reaction, 6-MNA. As described by the free energy surface shown in Figure 5a and confirmed by the analysis of the key inter-atomic distances (see Figure 5b, 5c and Table S4 of Supporting Information), the step occurs in a very asynchronous way. The water molecule has already largely been formed before the transition state takes place. The structure in the TS3 shows how the cleavage of the O-C bond is in a very advanced stage (O-C = 3.50 ± 0.23 Å) and the scissile carbon-carbon bond already broken (C-C = 3.58 ± 0.17 Å). The reaction is strongly endergonic, with a reaction free-energy value of 16.0 kcal $mol^{-1}$ and an activation energy of 29.1 kcal $mol^{-1}$.

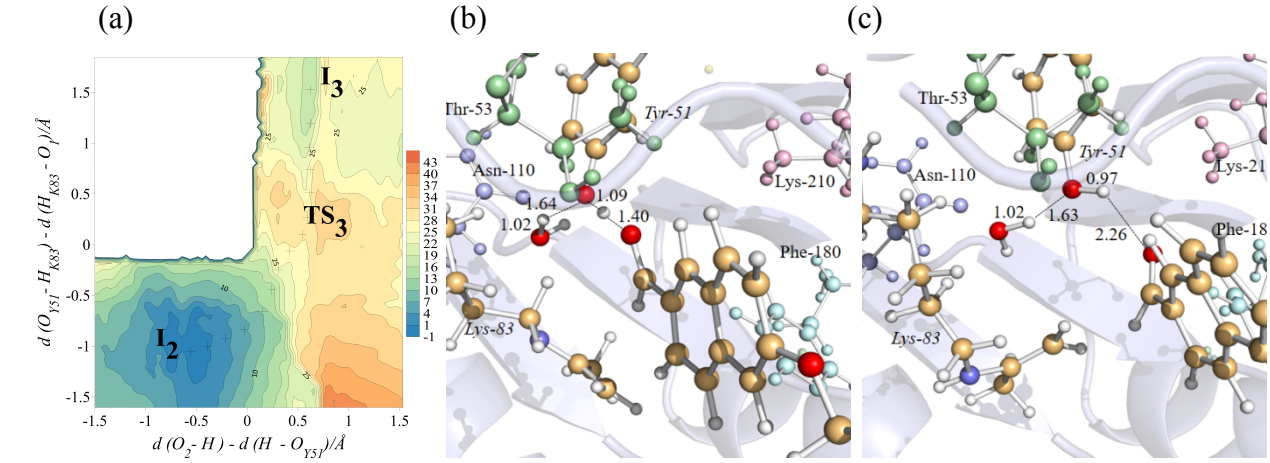



**Figure 5**. (a) M06-2X:AM1/MM FES obtained as a 2D-PMF for the third step of the reaction at 300 K. Representative snapshots of (b) $TS_3$ and (c) $I_3$. All values of energies are given in units of kcal $mol^{-1,}$ and distances are given in units of Å.

The **fourth step** involves the formation of a Schiff base. The free-energy surface is shown in Figure 6a as a mono-dimensional PMF. The activation free energy for this step is 9.6 kcal $mol^{-1}$, while the reaction free energy is -3.1 kcal $mol^{-1}$. This step consists in a proton transfer from Tyr-51 to the C12 atom of the substrate (see Figure 6b and 6c). According to experimental data[81] and our previous QM/MM theoretical results,[54] this is the rate-determining step in the equivalent reaction catalyzed by the RA95.5-8F. The water molecule in the active site, generated during the previous step, remains well-positioned to participate in the following step (see average key distances in Table S5 of Supporting Information).

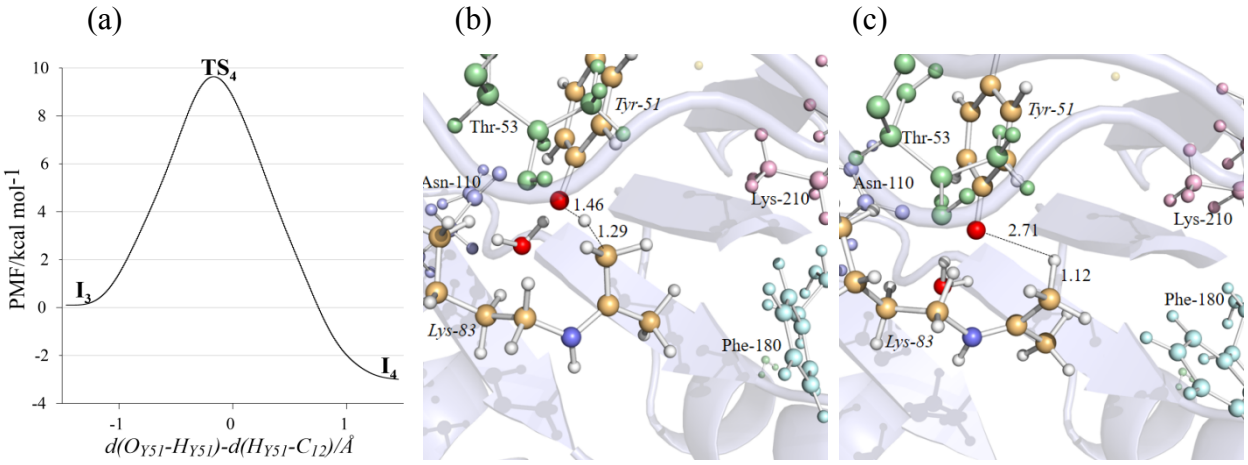



**Figure 6**. (a) M06-2X:AM1/MM FES obtained as a 1D-PMF for the fourth step of the reaction at 300 K. Representative snapshots of (b) $TS_4$ and (c) $I_4$, respectively. All values of energies are given kcal $mol^{-1}$ and distances in Å.

The PMF of the **fifth step** is shown in Figure 7a. An oxygen atom, from a water molecule, attacks the $sp^2$ carbon atom of the Schiff base, and simultaneously, the Tyr-51 deprotonates the water molecule. The product obtained is the intermediate carbinolamine, $I_5$, shown in Figure 7c (the average key distances are reported in Table S6 of Supporting Information). The activation free-energy is 12.9 kcal $mol^{-1}$, while the $I_5$ turns out to be 7.8 kcal $mol^{-1}$ more stable than the $I_4$. The obtained picture of this step is comparable to the one obtained for the reaction catalyzed by RA95.5-8F[54], in which the activation and reaction free-energy were 7.6 kcal $mol^{-1}$ and -6.1 kcal $mol^{-1}$, respectively.

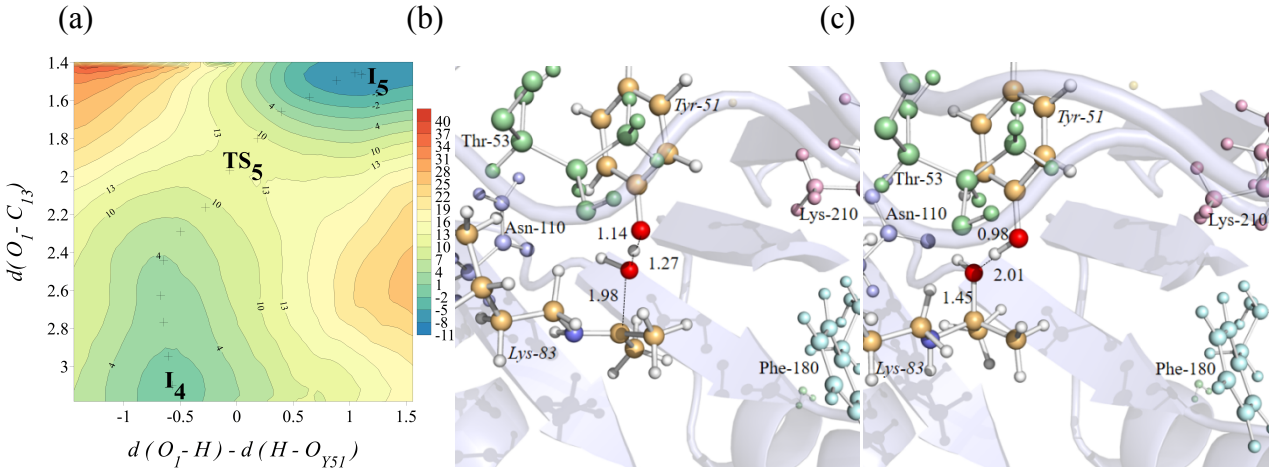



**Figure 7**. (a) M06-2X:AM1/MM FES obtained as a 2D-PMF for the fifth step of the reaction at 300 K. Representative snapshots of (b) $TS_5$ and (c) $I_5$. All values of energies are given in kcal $mol^{-1}$ and distances in Å.

The reaction of the **sixth step** consists in the deprotonation of the Tyr-51 by the nitrogen of carbinolamine. According to the FES depicted in Figure 8a, the production of a charged species is not a favorable process. The activation free-energy, determined by a TS in a very advanced stage, requires 14.7 kcal $mol^{-1}$ while the reaction free-energy is 14.3 kcal $mol^{-1}$. Similar behavior was already obtained for this step in the reaction catalyzed by RA95.5-8F.[54]

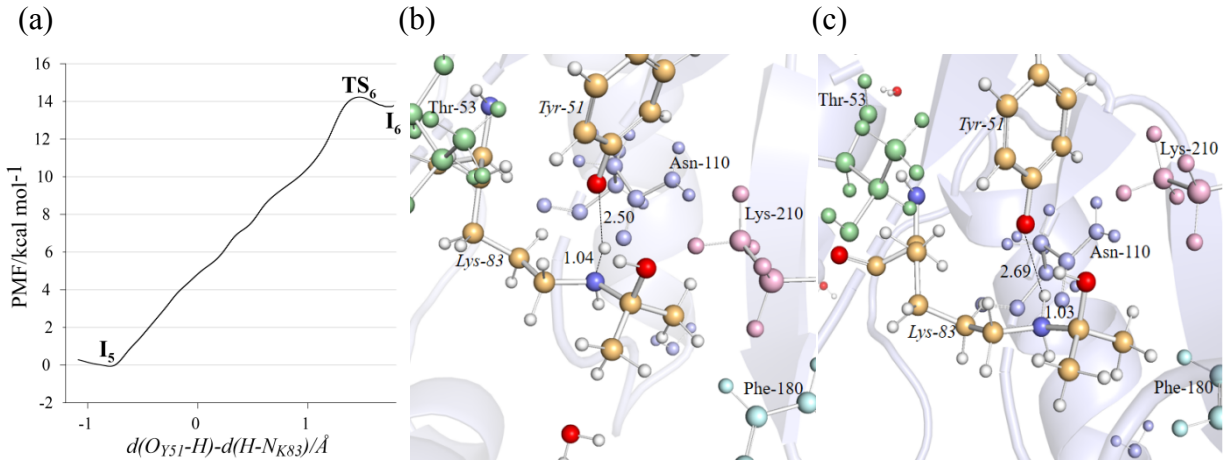



**Figure 8**. (a) M06-2X:AM1/MM FES obtained as a 1D-PMF for the sixth step of the reaction at 300 K. Representative snapshots of (b) $TS_6$ and (c) $I_6$. All values of energies are given in kcal $mol^{-1}$ and distances in Å.

The formation of the acetone and the restoration of the enzyme take place in the **seventh step**. According to the FES shown in Figure 9a, the product is only 0.2 kcal mol$^{-1}$ more stable than the $I_6$, while the activation free-energy is 6.9 kcal mol$^{-1}$. The geometries in the $TS_7$ and the corresponding products state, acetone plus the enzyme, clearly show how the deprotonation of the hydroxyl by Tyr-51 reestablishes the neutral character of the species involved in the reaction (see key inter-atomic distances in Figure 9b and 9c, and in Table S8 of Supporting Information). As in the previous step, the proton transfer and the N-C breaking bond are in a very advanced stage of the process in the $TS_7$.

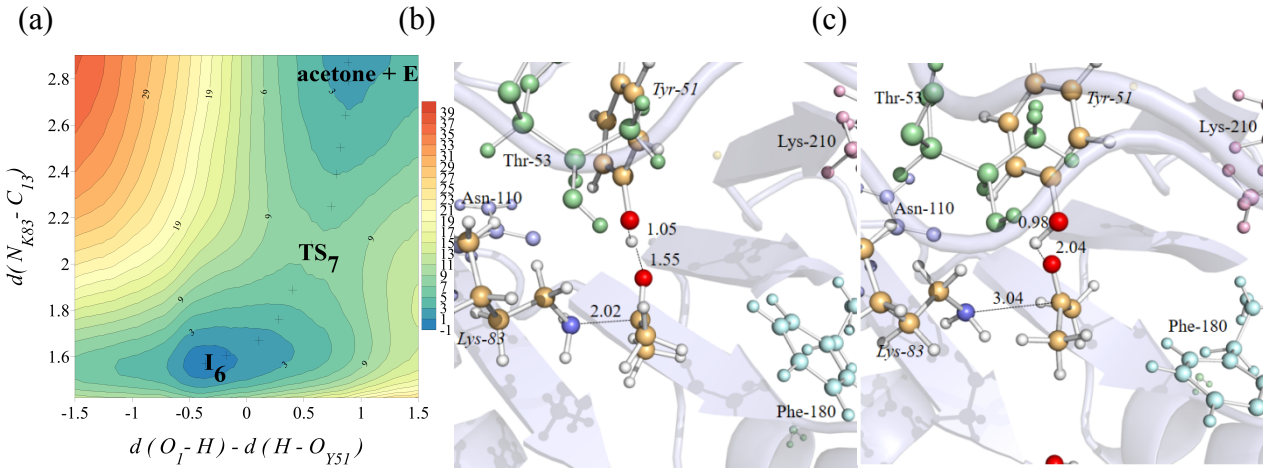



**Figure 9**. (a) M06-2X:AM1/MM FES obtained as a 1D-PMF at 300 K for the seventh step of the reaction. Representative snapshots of the (b) $TS_7$ and (c) acetone + Enzyme. All values of energies are given in kcal mol$^{-1}$ and distances in Å.

## DISCUSSION

The information derived from each of the FESs generated along the 7 consecutive steps provides the full energy landscape of the retro aldol reaction catalyzed by the RA95.5-5 *de novo* enzyme (see Figure 10). Our simulations suggest that the third step, corresponding to the C-C bond cleavage of methodol and the release of 6-MNA, is the rate-determining step of the full process, which is in agreement with experimental data.[81] The detailed study of the individual steps (see Figures 3-9), and the comparison with the corresponding steps of the previously studied reaction mechanism of RA95.5-8F (see Figure 10), allows explaining the consequences on the energetics of each step after mutations. Thus, two general conclusions can be derived by comparing the two energy profiles (see solid and dashed profiles in Figure 10). First, our QM/MM results show how the mutations introduced on the less evolved RA95.5-5 to get the more efficient RA95.5-8F provokes an evident diminution on the overall activation free energy from 29.1 to 20.1 kcal·mol$^{-1}$. This trend is in very good agreement with the kinetic data of Hilvert and coworkers.[81] And second, the comparison between the

full energy profile of the two variants shows how the successful efforts carried out by Hilvert and co-workers to decrease the barrier of the rate-limiting step in the RA95.5-5, the third step, have two noticeable side effect: the increase in the barrier of step 1 and, more noticeable, the rate-determining step is no longer the third step but the fourth one in RA95.5-8F. These two unfavorable effects cannot be considered, nevertheless, as unexpected, since we are dealing with a multi-step reaction and stabilization of the TS of one step can destabilize other TSs.

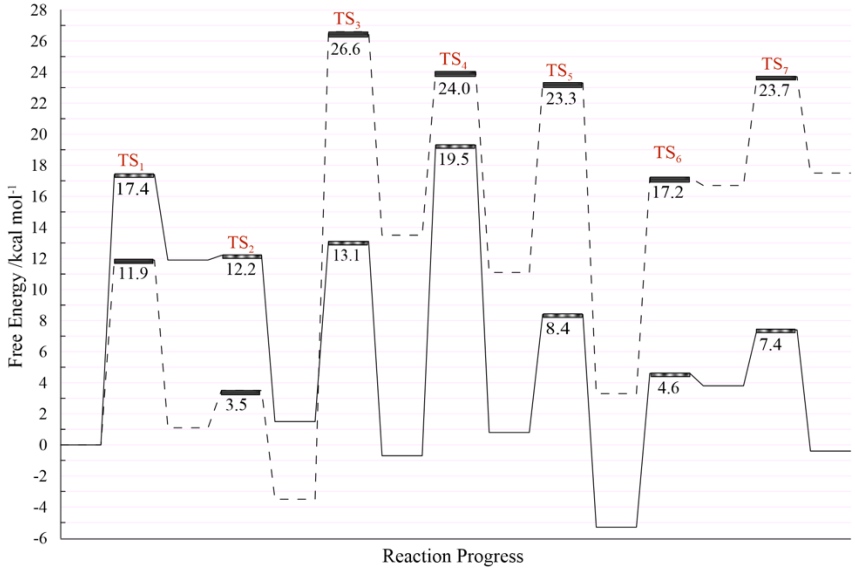



**Figure 10**. M062X/6-31+G(dp):AM1/MM free-energy profile for the reaction mechanism of the Retro-Aldol reaction of the (R)-methodol enantiomer catalyzed by de novo enzyme RA95.5-5 (solid line), compared with our previously QM/MM energy profile obtained for the same reaction catalyzed by the more evolved RA95.5-8F (dashed line).[54]

In order to analyze in detail why the energy profile of the reaction is modified after mutations, an analysis of how the different amino acids of the protein affect the efficiency of the enzyme was carried out. In particular, the catalytic role of the different residues was measured by computing the activation interaction energy by residue, in the TS and the preceding minimum of each step, i.e. the difference in the electrostatic (QM) and Lennard-Jones (MM) interaction energies between the substrate and the protein. The calculations were done in the two *de novo* designed proteins, and the efforts were focused on the key chemical steps, as deduced from the discussion in Figure 10: steps 1, 3 and 4. The decomposition of the averaged activation interaction energy by residue in the three steps of RA95.5 and RA95.5-8F are shown in Figure 11. Substrate-protein interaction energies decomposed by residue for different states along the reaction path are shown in Figures S2-S4 of Supporting Information. The conclusions that can be derived from this comparative analysis on each step are as follows:

***First step***: The results of step 1 for both enzymes are shown in Figure 11A. The reaction appears to be mainly favored in the RA95.5-5 (left panel of the figure) by the presence of Lys-210, Asp-111 and Glu-85, being greater the interaction with the transition state than with the reagent. Asp-111 and Glu-85 are also present in RA95.5-8F (see right panel of Figure 11A) with similar interaction values, but the favorable interaction between the substrate and Lys-210 in RA95.5-5 is completely absent in RA95.5-8F, where a new Leu is located at this position. RA95.5-8F shows a new significant favorable interaction with Asn-110, but also an increase of the number of non-favorable interactions with residues such as Arg-19, Leu-53, Ser-58, Gly-59, Glu-188 and Asp-212. These different patterns of interactions can explain the increase in the energy barrier in the first step after mutations.

***Third step:*** The analysis of the third step, shown in the two panels of Figure 11B, highlights the absence of a proper residue to promote this step in RA95.5-5, apart from Lys-55 (and Tyr-53 and Glu185 at less extent), and the non-favorable effects of Lys-210 and, at less extent, Phe-180. Lys-210 disfavor the reaction by stabilizing the intermediate $I_2$ with respect to $TS_3$. Right panel of Figure 11B, corresponding to RA95.5-8F, shows that while the catalytic effect of Lys-55 is lost, residue at position 210 is no longer destabilizing the $TS_3$ and, in addition, there is an increasing number of stabilizing interactions by different residues. It is remarkable how the Phe180Tyr mutation provokes a series of conformational changes in the surroundings of this position that favor the establishment of new favorable effects of residues such as Ser-81, Asp-111, Glu-135, Glu-188, Met-231 and Ser-233. As a consequence of all of these new favorable interactions, the free energy barrier of the third step in the reaction catalyzed by RA95.5-8F is reduced by 11.6 kcal $mol^{-1}$ with respect to RA95.5-5. This step becomes an exergonic reaction (-2.2 kcal $mol^{-1}$).[54]

***Fourth step:*** Finally, the comparison of the pattern of activation interaction energies of the two *de novo* enzymes in the step 4, depicted in the two panels of Figure 11C, shows a noticeable number of favorable interactions in the less evolved RA95.5-5 enzyme (Tyr-52, Thr-53, Asp-61, Glu-85, Asp-111, Leu-231, Met-237 and, above all, Lys210) as well as some unfavorable interactions (Trp-8, Lys-135, Ile-232 and, above all, Lys55). The more evolved designed enzyme does not present the most intense stabilizing interaction with residue 210, due to the mutation of the Lys into a Leu in this position, despite the destabilizing interaction with residue 55 (Lys-55 in RA95.5-5) is not present. As could be expected, the pattern of stabilizing and destabilizing interactions is significantly different between both enzymes. *Grosso modo*, less stabilizing and more destabilizing interactions are detected in step 4 of RA95.5-8F than in RA95.5-5, which agrees with an increase of ca. 9 kcal $mol^{-1}$ in the free energy barrier of this step.

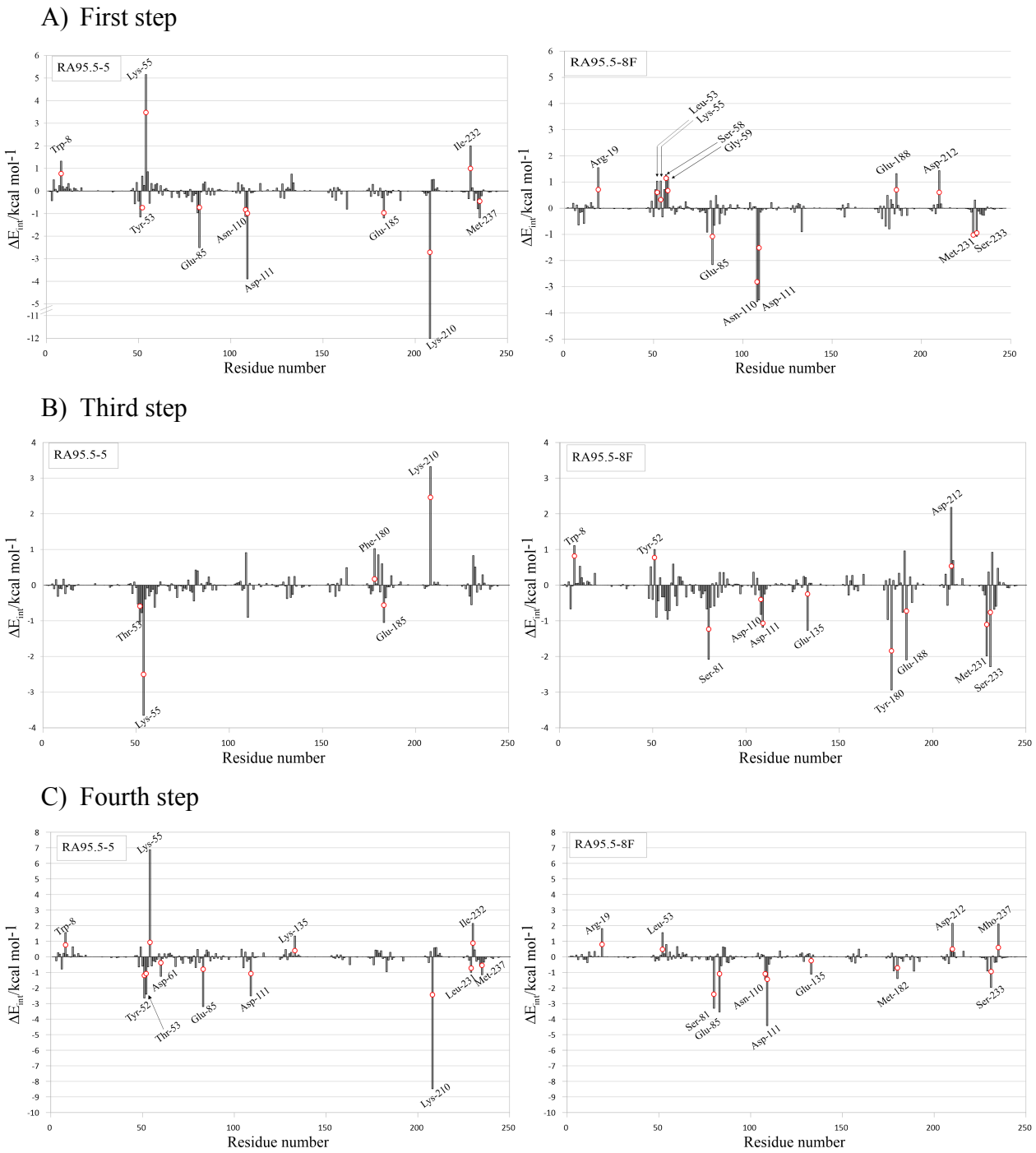



**Figure 11**. Averaged activation interaction energy of RA95.5-5 (left panels) and RA95.5-8F (right panels), for the (A) first (B) third and (C) fourth steps of retro-aldol reaction. Only residues showing interaction energies higher than 1 kcal $mol^{-1}$, in absolute value, are labelled in the panels. Standard deviation of the activation interaction energy per residue is indicated as circles for the most relevant interactions.

In summary, the changes in the activation energies of steps 1, 3 and 4, when comparing RA95.5-5 and RA95.5-8F, are in agreement with the analysis of the pattern of interactions. Thus, it is evident

that RA95.5-5 has more stabilizing interactions than in RA95.5-8F in steps 1 and 4 while, in contrast, RA95.5-8F shows a much clear stabilizing pattern of interactions than RA95.5-5 in step 3. It is interesting to note the role of Lys-210 in stabilizing the $TS_1$ and $TS_4$ in the RA95.5-5 (left panels of Figure 11A and 11C), but not the $TS_3$. The reason can lie in the interaction of the positive Lys-210 with the Tyr-51. This latter, indeed, develop a negative charge in steps 1 and 4, while ends up with a neutral charge in the rate-limiting step, the third step. This behavior of the interaction energies is mirrored when analyzing the electrostatic potential generated by residue on the oxygen atom of Tyr-51 (see Figure S5, S6 and S7 of the Supporting Information). Lys-210 creates a more positive potential on the oxygen atom of Tyr-51 in the $TS_1$ than in the reactants (ca. 44 vs. 39 kcal $mol^{-1}$), which means that Lys-210 stabilizes the negative charge that is being developed in the oxygen atom of Tyr-51 in RA95.5-5 (see Figure S5). In contrast, the electrostatic potential generated by Leu-210, in RA95.5-8F, is dramatically smaller and does not change along this step (ca. 2 kcal $mol^{-1}$). The same behavior is obtained when analyzing the electrostatic role of Lys-210 in step 4 in RA95.5-5 (see Figure S7). In the third step of the RA95.5-5, the electrostatic potential generated by Lys-210 on Tyr-51 decreases (from ca. 49 to 38 kcal $mol^{-1}$), which could be considered conforming to a higher energy barrier in this step (see Figure S6). In RA95.5-8F, the positive electrostatic potential, generated by Asn-110 on Tyr-51, is observed in all the initial steps (ca. 12 kcal $mol^{-1}$); in the fourth, moving from $I_3$ to $TS_4$, the electrostatic potential disappears. This trend is in agreement with the increase in the energy barrier of this step. The geometries analysis shows that the water molecule formed in the active site interacts with Tyr-51, and dramatically modifies the relative orientation of this residue from $I_3$ to $TS_4$.

## CONCLUSIONS

The molecular mechanism of the non-natural multistep retro-aldolase reaction catalyzed by a *de novo* enzyme, the RA95.5-5 designed by Hilvert and co-workers,[46] has been investigated by means of multiscale QM/MM methods using methodol as substrate. The results have been compared to our previous computational study on the more evolved RA95.5-8F retro-aldolase,[52] designed also by Hilvert and co-workers.[43] The obtained free energy landscape has allowed to confirm how the RA95.5-5 presents a higher overall activation free energy barrier than the RA95.5-8F (29.1 vs. 20.1 kcal $mol^{-1}$), as well as to define the details of every of the 7 steps of the full reaction mechanism. According to our results, both designed enzymes catalyze the reaction by means of the same molecular mechanism. Nevertheless, while the rate-limiting step on the RA95.5-5 corresponds to the substrate carbon-carbon bond breaking that generates the first product 6-MNA ($3^{rd}$ step in scheme 1),

the rate-limiting step on the RA95.5-8F appeared to be the decomposition of an enamine intermediate into a Schiff base ($4^{th}$ step in Scheme 1). This is in accordance with pre-steady-state kinetic methods and isotope exchange experiments carried out by Hilvert and co-workers.[43,46]

Our results confirm the difficulties of improving the efficiency of an enzyme that has to be able to maintain low energy barriers along a multi-step chemical reaction, such as the retro-aldolases. The mutation of residues in specific positions (17 amino-acids when evolving from RA95.5-5 to RA95.5-8F, as shown in Figure 1) reduces the barrier of the targeting rate-limiting step, but it brings anti-catalytic effects in other steps. Indeed, our simulations show how the activation free energy barrier of step 3 is reduced from 29 to 10 kcal $mol^{-1}$ after introducing these 17 mutations, but the corresponding free energy barrier of step 4 is increased from 9.5 to 20.0 kcal $mol^{-1}$. In addition, the activation free energy of step 1 is also perturbed during the directed evolution, increasing in ca. 5 kcal $mol^{-1}$ (see Figure 10). The analysis of the electrostatic interaction energy between the substrate and the different amino acid residues of the protein shows how the number of favourable interactions (stabilizing further the TS than the preceding intermediate) in the step 3 is larger in the RA95.5-8F than in the RA95.5-5, and the opposite trend is detected in step 1 and 4. This is in perfect agreement with the corresponding activation energies. Some of the substitutions appear to be exceptionally well connected with the energetics of these steps, such as the mutation Lys210Leu. Thus, Lys-210 appears to have an unfavourable effect in step 3, and a favourable effect in steps 1 and 4 in RA95.5-5. When this residue is mutated to Leu to get RA95.5-8F, the unfavourable effect in step 3 disappears, which is associated to a decrease in the energy barrier, but the catalytic effect that it was producing in steps 1 and 4 also disappears, provoking an increase in their corresponding energy barriers.

The results of the protein-substrate interaction energies are mirrored in the analysis of electrostatic potential generated by residue in one of the key atoms of the system: the oxygen atom of Tyr-51. Our results show how Lys-210 generates a positive potential that is increased in the $TS_1$ and $TS_4$, thus stabilizing the accumulation of negative charge on Tyr-51 that takes place in step 1 and 4 of the full retro-aldol reaction. The corresponding Leu-210 in RA95.5-8F does not produce any favorable electrostatic effect on these steps, which is translated into an increase of the activation free energies. Thus, the relevant role plaid by residue 210 in RA95.5-5, which disappears entirely in the more evolved RA95.5-8F, could be reconsidered in future mutants. It is also remarkable that new residues can provoke changes in the role of conserved residues. This is the case of residues such as Ser-81, Asp-111, Glu-188 or Ser-233, which contribute to lowering the free-energy barrier of step 3 in RA95.5-8F, but they have not clear effect in the RA95.5-5 design.

In all, the analysis of the interaction energies decomposed by residues for each step of an enzyme catalyzed reaction, together with the evaluation of key electrostatic properties, provide a very useful insight on how each amino-acid interacts and participates in the catalysis. This information can be used not only to understand the behavior of every design generated by directed evolution, but also to suggest mutations that can further improve the overall energetic profile.

Daria De Raffele: 0000-0001-5007-9974
Sergio Martí: 0000-0000-0002-1087-7143
Vicent Moliner: 0000-0002-3665-3391

**Author Contributions**

D. De Raffele carried out all the calculations. S. Martí and V. Moliner designed the project. All authors analyzed the results and contributed to writing the manuscript.

**Notes**

The authors declare no competing financial interest.

**Supporting Information**.

The Supporting Information is available free of charge on the ACS Publications website at DOI: Computational details; experimental parameters data of RA variants with (R)-methodol; Time evolution of the RMSD computed along the classical MD simulation for the backbone atoms of the protein; tables of the average key interatomic distances of the different states located in all steps of the reaction; figures of interaction energy by residue for different states along the reaction path; figures of electrostatic potential generated by residue on the negatively oxygen atom of Tyr-51 for different states of the reaction; schematic representation of alternative explored reaction mechanisms;

M06-2X:AM1/MM PESs and M06-2X:AM1/MM FESs for the alternative explored reaction mechanisms.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (Grant PGC2018-094852-B-C21), the Generalitat Valenciana (Grant AICO/2019/195) and the Universitat Jaume I (project UJI·B2017- 31). DDR thanks the Generalitat Valenciana for a Grisolia PhD grant (ref. GRISOLIAP/2016052). Authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I.

REFERENCES:


(1) Pauling, L. *General Chemistry: An Introduction to Descriptive Chemistry and Modern Chemical Theory*, 1st ed.; San Francisco,W. H. Freeman and Company, 1947.

(2) Watson, J. D. *Avoid Borig People*,; Alfred A. Knopf, New York, Ed.; 2007.

(3) Jäckel, C.; Kast, P.; Hilvert, D. Protein Design by Directed Evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153–173.

(4) Brustad, E. M.; Arnold, F. H. Optimizing Non-Natural Protein Function with Directed Evolution. *Curr. Opin. Chem. Biol.* **2011**, *15*, 201–210.

(5) Taylor, S. V.; Kast, P.; Hilvert, D. Investigating and Engineering Enzymes by Genetic Selection. *Angew. Chemie - Int. Ed.* **2001**, *40*, 3310–3335.

(6) Wang, Q.; Parrish, A. R.; Wang, L. Expanding the Genetic Code for Biological Studies. *Chem. Biol.* **2009**, *16*, 323–336.

(7) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chemie - Int. Ed.* **2015**, *54*, 3351–3367.

(8) Todd, A. E.; Orengo, C. A.; Thornton, J. M. Evolution of Function in Protein Superfamilies, from a Structural Perspective. *J. Mol. Biol.* **2001**, *307*, 1113–1143.

(9) Fischbach, M. A.; Lai, J. R.; Roche, E. D.; Walsh, C. T.; Liu, D. R. Directed Evolution Can Rapidly Improve the Activity of Chimeric Assembly-Line Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11951–11956.

(10) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

(11) Babbitt, P. C.; Gerlt, J. A. Understanding Enzyme Superfamilies. *J. Biol. Chem.* **1997**, *272*, 30591–30594.

(12) Bartlett, G. J.; Borkakoti, N.; Thornton, J. M. Catalysing New Reactions during Evolution: Economy of Residues and Mechanism. *J. Mol. Biol.* **2003**, *331*, 829–860.

(13) Chica, R. A. Protein Engineering in the 21 St Century . *Protein Sci.* **2015**, *24*, 431–433.

(14) Zanghellini, A. De Novo Computational Enzyme Design. *Curr. Opin. Biotechnol.* **2014**, *29*, 132–138.

(15) Swiderek, K.; Tuñón, I.; Moliner, V. Predicting Enzymatic Reactivity: From Theory to Design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *4*, 407–421.

(16) Agarwal, P. K. A Biophysical Perspective on Enzyme Catalysis. *Biochemistry* **2018**, *58*, 438–449.

(17) Suplatov, D.; Kirilin, E.; Takhaveev, V.; Švedas, V. Zebra: A Web Server for Bioinformatic Analysis of Diverse Protein Families. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1752–1758.

(18) Mulholland, A. J. Modelling Enzyme Reaction Mechanisms, Specificity and Catalysis. *Drug Discov. Today* **2005**, *10*, 1393–1402.

(19) Swiderek, K.; Moliner, V.; Tuñón, I.; Bertran, J. Computational Strategies for the Design of New Enzymatic Functions. *Arch. Biochem. Biophys.* **2015**, *582*, 68–79.

(20) Yin, J.; Schultz, P. G. *Catalytic Antibodies*; Keinan, E., Ed.; 2005.

(21) Hilvert, D. Critical Analysis of Antibody Catalysis. *Annu. Rev. Biochem.* **2000**, *69*, 751–793.

(22) Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chemie* **2013**, *52*, 5700–5725.

(23) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative Approach to Computational Enzyme Design. *PNAS* **2012**, *109*, 1–6.

(24) Khare, S. D.; Fleishman, S. J. Emerging Themes in the Computational Design of Novel Enzymes and Protein – Protein Interfaces. *FEBS Lett.* **2013**, *587*, 1147–1154.

(25) Kries, H.; Blomberg, R.; Hilvert, D. De Novo Enzymes by Computational Design. *Curr. Opin. Chem. Biol.* **2013**, *17*, 1–8.

(26) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447–470.

(27) Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of Conformational Dynamics in the Evolution of Novel Enzyme Function. *Chem. Commun.* **2018**, *54*, 6622–6634.

(28) Sterner, R.; Merkl, R.; Raushel, F. M. Computational Design of Enzymes. *Chem. Biol.* **2008**,

*15*, 421–423.

(29) Gordon, S. R.; Stanley, E. J.; Wolf, S.; Toland, A.; Wu, S. J.; Hadidi, D.; Mills, J. H.; Baker, D.; Pultz, I. S.; Siegel, J. B. Computational Design of an α  Gliadin Peptidase. *J. Am. Chem. Soc.* **2012**, *134*, 20513–20520.

(30) Huang, P.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537*, 320–327.

(31) Wenda, S.; Illner, S.; Mell, A.; Kragl, U. Industrial Biotechnology—the Future of Green Chemistry? *Green Chem.* **2011**, *13*, 3007.

(32) Behrens, G. A.; Hummel, A.; Padhi, S. K.; Schätzle, S.; Bornscheuer, U. T. Discovery and Protein Engineering of Biocatalysts for Organic Synthesis. *Adv. Synth. Catal.* **2011**, *353*, 2191–2215.

(33) Hammer, S. C.; Knight, A. M.; Arnold, F. H. Design and Evolution of Enzymes for Non-Natural Chemistry. *Curr. Opin. Green Sustain. Chem.* **2017**, *7*, 23–30.

(34) Ebert, M. C.; Pelletier, J. N. Computational Tools for Enzyme Improvement: Why Everyone Can – and Should – Use Them. *Curr. Opin. Chem. Biol.* **2017**, *37*, 89–96.

(35) Bommarius, A. S.; Blum, J. K.; Abrahamson, M. J. Status of Protein Engineering for Biocatalysts: How to Design an Industrially Useful Biocatalyst. *Curr. Opin. Chem. Biol.* **2011**, *15*, 194–200.

(36) Risso, V. A.; Martinez-Rodriguez, S.; Candel, A. M.; Krüger, D. M.; Pantoja-Uceda, D.; Ortega-Muñoz, M.; Santoyo-Gonzalez, F.; Gaucher, E. A.; Kamerlin, S. C. L.; Bruix, M.; Gavira, J. A.; Sanchez-Ruiz, J. M. De Novo Active Sites for Resurrected Precambrian Enzymes. *Nat. Commun.* **2017**, *8*, 1–13.

(37) Petrović, D.; Shina Caroline Lynn, K. Molecular Modeling of Conformational Dynamics and Its Role in Enzyme Evolution. *Curr. Opin. Struct. Biol.* **2018**, *52*, 50–57.

(38) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; Dechancie, J.; Betker, J.; Ro, D.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190–195.

(39) 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.; Hilvert, D.; Houk, K. N.; Stoddard, B.

L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387–1391.

(40) Nosrati, G. R.; Houk, K. N. SABER : A Computational Method for Identifying Active Sites for New Reactions. **2012**, *21*, 697–706.

(41) Leaver-fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Davis, I. W.; Cooper, S.; Treuille, A. ;Mandell, D. J. ;Ban, Y. A.; Fleishman, S. J; Corn, J. E.; Kim, D. ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. **2011**, *487*, 545–574.

(42) Tantillo, D. J.; Chen, J.; Houk, K. N. Theozymes and Compuzymes: Theoretical Models for Biological Catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743–750.

(43) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6*, 1–12.

(44) Damborsky, J.; Brezovsky, J. Computational Tools for Designing and Engineering Enzymes. *Curr. Opin. Chem. Biol.* **2014**, *19*, 8–16.

(45) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. Emergence of a Catalytic Tetrad during Evolution of a Highly Active Artificial Aldolase. *Nat. Chem.* **2016**, *9*, 50–56.

(46) Bjelic, S.; Kipnis, Y.; Wang, L.; Pianowski, Z.; Wang, L.; Pianowski, Z.; Kornhaber, G.; Su, M.; Seetharaman, J.; Xiao, R.; Kornhaber, G.; Hunt, J. F.; Tong, L.; Hilvert, D.; Baker, D. Exploration of Alternate Catalytic Mechanisms and Optimization Strategies for Retroaldolase Design. *J. Mol. Biol.* **2014**, *426*, 256–271.

(47) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; Hilvert, D.; Baker, D. Robust Design and Optimization of Retroaldol Enzymes. *Protein Sci.* **2012**, *21*, 717–726.

(48) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a Designed Retro-Aldolase Leads to Complete Active Site Remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494–498.

(49) Nagano, N.; Orengo, C. A.; Thornton, J. M. One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on Their Sequences, Structures and Functions. *J. Mol. Biol.* **2002**, *321*, 741–765.

(50) Henn-Sax, M.; Höcker, B.; Wilmanns, M.; Sterner, R. Divergent Evolution of $(\beta\alpha)_8$-Barrel Enzymes. *Biol. Chem.* **2001**, *382*, 1315–1320.

(51) Toscano, M. D.; Hilvert, K. J. W. D. Minimalist Active-Site Redesign: Teaching Old Enzymes New Tricks. *Angew. Chemie - Int. Ed.* **2007**, *46*, 3212–3236.

(52) Schafer, J.; Zoi, I.; Antoniou, D.; Schwartz, S. D. Optimization of the Turnover in Artificial Enzymes via Directed Evolution Results in the Coupling of Protein Dynamics to Chemistry. *J. Am. Chem. Soc.* **2019**, *141*, 10431–10439.

(53) Romero-rivera, A.; Garcia-borràs, M.; Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **2017**, *7*, 8524−8532.

(54) De Raffele, D.; Martí, S.; Moliner, V. QM/MM Theoretical Studies of a De Novo Retro-Aldolase Design. *ACS Catal.* **2019**, *9*, 2482–2492.

(55) Humphrey, K.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(56) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The Dynamo Library for Molecular Simulations Using Hybrid Quantum Mechanical and Molecular Mechanical Potentials. *J. Comput. Chem.* **2000**, *21*, 1088–1100.

(57) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free□energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(58) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.

(59) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(60) Kästner, J.; Thiel, W. Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration. *J. Chem. Phys.* **2006**, *124*.

(61) Hub, J. S.; De Groot, B. L.; Van Der Spoel, D. G-Whams-a Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.

(62) Truhlar, D. G. Transition State Theory for Enzyme Kinetics. *Arch. Biochem. Biophys.* **2015**,

*582*, 10–17.

(63) S. J. Klippenstein, V. Pande, and D. G. Truhlar, J. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J Am Chem Soc.* **2014**, *136*, 528–546.

(64) Poulsen, T. D.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. Free Energy Surface Reaction Paths, and Kinetic Isotope Effect of Short-Chain Acyl-CoA Dehydrogenase. *J. Phys. Chem. B* **2003**, *107*, 9567–9578.

(65) Crouzy, S.; Baudry, J.; Smith, J. C.; Roux, B. J. Efficient Calculation of Two-Dimensional Adiabatic and Free Energy Maps: Application to the Isomerization of the C13=C14 and C15=N16 Bonds in the Retinal of Bacteriorhodopsin. *Comput. Chem.* **1999**, *20*, 1644–1658.

(66) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(67) Jorgensen, W. L.; Maxwell, D. S.; Tirado-rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Am. Chem. Soc.* **1996**, *7863*, 11225–11236.

(68) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(69) Pranata, J.; Wierschke, S. G.; 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.

(70) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(71) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(72) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes. *J. Phys. Chem. B* **2006**, *110*, 17663–17670.

(73) Roca, M.; Moliner, V.; Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. Activation Free Energy of Catechol O-Methyltransferase. Corrections to the Potential of Mean Force. *J. Phys. Chem. A* **2006**, *110*, 503–509.

(74) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. Hybrid QM/MM Potentials of Mean Force with Interpolated Corrections. *J. Phys. Chem. B* **2004**, *108*, 8427–8433.

(75) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. Mapped Interpolation Scheme for Single-Point Energy Corrections in Reaction Rate Calculations and a Critical Evaluation of Dual-Level Reaction Path Dynamics Methods. *J. Phys. Chem. A* **1999**, *103*, 1140–1149.

(76) Corchado, J. C.; Coitiño, E. L.; Chuang, Y. Y.; Fast, P. L.; Truhlar, D. G. Interpolated Variational Transition-State Theory by Mapping. *J. Phys. Chem. A* **1998**, *102*, 2424–2438.

(77) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. A Dual-Level Shepard Interpolation Method for Generating Potential Energy Surfaces for Dynamics Calculations. *J. Chem. Phys.* **1995**, *103*, 5522–5530.

(78) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* **2003**, *107*, 1384–1388.

(79) 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.; Sonnenb, D. J. 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, J. M.; 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, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Gaussian, Inc: Wallingford, CT , 2009.

(80) De Raffele, D.; Martí, S.; Moliner, V. QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catal.* **2019**, *9*.

(81) Zeymer, C.; Zschoche, R.; Hilvert, D. Optimization of Enzyme Mechanism along the

Evolutionary Trajectory of a Computationally Designed (Retro-)Aldolase. *J. Am. Chem. Soc.* **2017**, *139*, 12541–12549.

For Table of Contents Only