Published in final edited form as:

J Am Chem Soc. 2019 July 03; 141(26): 10431–10439. doi:10.1021/jacs.9b04515.

Optimization of the Turnover in Artificial Enzymes via Directed Evolution Results in the Coupling of Protein Dynamics to Chemistry

Joseph W. Schafer[†], Ioanna Zoi[†], Dimitri Antoniou, and Steven D. Schwartz^{*}

Department of Chemistry and Biochemistry, University of Arizona, Tucson, Arizona 85721, United States

Abstract

The design of artificial enzymes is an emerging field of research. Although progress has been made, the catalytic proficiency of many designed enzymes is low compared to natural enzymes. Nevertheless, recently Hilvert et al. (*Nat. Chem.* 2017, 9, 50–56) created a series of five artificial retro-aldolase enzymes via directed evolution, with the final variant exhibiting a rate comparable to the naturally occurring enzyme fructose 1,6 biphosphate aldolase. We present a study of this system in atomistic detail that elucidates the effects of mutational changes on the chemical step. Transition path sampling is used to create ensembles of reactive trajectories, and committor analysis is used to identify the stochastic separatrix of each ensemble. The application of committor distribution analysis to constrained trajectories allows the identification of changes in important protein motions coupled to reaction across the generated series of the artificial retro-aldolases. We observed two different reaction mechanisms and analyzed the role of the residues participating in the reaction coordinate of each enzyme. However, only in the most evolved variant we identified a fast motion that promotes catalysis, suggesting that this rate promoting vibration was introduced during directed evolution. This study provides further evidence that protein dynamics must be taken into account in designing efficient artificial enzymes.

Graphical Abstract

^{*}Corresponding Author sschwartz@email.arizona.edu.

[†]JWS and IZ contributed equally to this work.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b04515.

Figures: (S1) Protein scaffold with all the mutations highlighted; (S2) Unsuccessful committor distributions for variants I–V; (S3) Successful committor distributions for variants I–V; (S4, S5, S7) RMSF comparison for important residues; (S6) Committor distribution histograms testing the rc of the OBM to PRM and vice versa; (S8) Protein sequence with mutations and rc residues highlighted; (S9) Charge distribution analysis for variants I, III and V (OBM, PRM); Tables: (S1) Mutations; (S2) Important distances in crystal and equilibrated structures for all enzymes; (S3–S8) Average distances between the substrate and RC residues at key points along the simulation for variants I–V; (S9) Observations on the nature of rc residues (PDF) Movies: (Si) unrestrained TPS simulation for variant V; (S2) restrained TPS simulation for variant V (ZIP)

Notes

The authors declare no competing financial interest.

■ INTRODUCTION

Directed evolution has become a principal technique of the enzyme design community.² The robust nature of directed evolution allows it to be applied to a wide variety of problems without needing to understand the link between sequence and function. The fundamental idea is to take an amino acid sequence, natural or engineered, and subject the sequence to random mutations.³ This generates progeny that can be screened for improvements in a variety of enzymatic functions, and those that have improved function are used as the parent for the next round of mutation. However, while evolution is capable of improving upon what is available, creating an amino acid sequence de novo for a novel chemical transformation remains a significant challenge, and current designs have low efficiency. Directed evolution then provides a mechanism to transform weakly active artificial enzyme designs to more functionally active ones.

Computationally designed artificial enzymes that have been created so far have low rate enhancements compared to naturally occurring counterparts.⁴⁻⁹ Recently, a series of artificial retro-aldolase enzymes¹ have been created and subjected to directed evolution, with the final protein exhibiting rates comparable to natural enzymes. Because the process of directed evolution is agnostic to any arguments regarding mechanism of rate enhancement, these sets of improved artificial biocatalysts provide us with a unique opportunity to

examine in atomistic detail the changes in mechanism that optimize the rate. This, in turn, provides information on what functional deficiencies the original amino acid sequence contained.

Aldolases are broadly useful catalysts for the stereoselective formation and cleavage of C–C bonds.¹⁰ They share a common mechanism involving amine catalysis via a Schiff base-forming lysine¹¹ (Figure 1).

We studied computationally the retro-aldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone in a series of artificial enzymes starting from the RA95.0 variant as reported by Hilvert et al.¹ There have been a number of studies^{12–14} that provided valuable insights on mechanistic and dynamical changes along natural or in vitro evolution trajectories. For example, in the Kemp Eliminase KE07 system conformational plasticity is directly linked to evolvability, showing that remote residues can reshape the active site for efficient catalysis. However, there have not been many computational studies on the effects of rapid dynamics on directed evolution.¹⁵ A recent study on these retro-aldolase enzymes explored the mechanism of only the last variant in the evolutionary process.¹⁶ This is the first study that includes all five variants of the retro-aldolase series.

Our group¹⁷ and others^{18–22} have identified protein motions that are part of the reaction mechanism in several enzymes: horse liver alcohol dehydrogenase,²³ lactate dehydrogenase,^{24–26} purine nucleoside phosphorylase,^{27,28} and aromatic amine dehydrogenase.²⁹

A rate promoting vibration is a motion within the protein body of an enzyme, part of the reaction coordinate and necessary for the reaction to occur. These motions are on the fs time scale and directly coupled to the transition state barrier crossing. Fast motions that are concomitant with the chemical event are not necessarily part of a rate-promoting protein motion: atomic fluctuations can produce a favorable geometry of the active site, but the specific path that led to this formation is not crucial. On the other hand, for residues that are part of a rate promoting vibration, only a specific motion of these residues can lead to the transition state geometry, for example, an organized motion of residues that *causes* a compression along the donor—acceptor axis facilitating the reaction. However, it must be noted that other authors have different opinions on the relevance of protein motions to catalysis.^{30–32}

In this study we found that an important rate-promoting protein vibration was introduced in the final variant of the retro-aldolase series. The retro-aldolase series consists of the enzymes RA95.0, RA95.5, RA95.5–5, RA95–5.8, and RA95.5.8F, which for simplicity we will denote by I, II, III, IV, and V for the rest of the paper. We performed a transition path sampling study (TPS) and analyzed the generated ensemble of reactive trajectories to obtain the reaction coordinate. We found that the enzymes I–IV catalyze the reaction through an intramolecular mechanism. The final enzyme V, however, has an additional reaction mechanism, with a reaction coordinate that had significant overlap with the residues mutated by directed evolution. The appearance of enhanced coupling of protein motion to the chemical step in a process that only optimizes the production of product is intriguing, and suggests that chemical step optimization during an evolutionary process may be concomitant

with such other rate determinants as product release. It is important to emphasize yet again that there is no bias in the nature of this introduction of dynamics into the reaction mechanism—it happens as an undirected result of directed evolution.

■ RESULTS AND DISCUSSION

We performed a transition path sampling study focusing on the C–C bond cleavage step (step 3 in Figure 1) in a series of artificial retro-aldolases. We will explain how the mutational changes introduced by directed evolution alter the details of the catalytic mechanism. We observe two different mechanisms across the generations of these enzymes, that involve separate sets of residues near the active site. Most importantly, the mutations resulting in the last variant introduced three residues that participate in a rate promoting vibration, which is one of the reasons for this successful design. Below, we will describe in detail the different protein scaffold architectures generated by directed evolution, the two mechanisms that can occur, the role of residues in the reaction coordinate identified for each variant, and the rate promoting vibration appearing only in the most evolved enzyme.

Structures.

The computational design for the first variant introduced 16 mutations to a (*β*₂)₈ barrel scaffold (Table S1, Figure S1) and includes Lys210 covalently bound to the substrate and the residue Glu53, assumed to play the role of the catalytic base for the C–C bond cleavage (Figure 2A). This enzyme was further evolved to yield 6 mutations in the second variant where a lysine appeared at position 83, which is conserved in the case of variants II–V and forms the Schiff base intermediate. Variants II–IV have a new tyrosine residue at position 51 suggested to play the role of the catalytic base (Figure 2B). The most evolved variant possesses a catalytic tetrad composed of Lys83, Tyr51, Asn110, with an additional Tyr180 (mutated from a phenylalanine) (Figure 2C). Note that the original catalytic Lys210 is now mutated to a leucine.

Mechanism.

The reaction associated with 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone for the five variants as shown in Figure 1 has multiple steps and it involves the formation of a Schiff base intermediate. Our transition path sampling study examines step 3, the C–C cleavage, known to be rate limiting for chemistry for the original design. Figure 1 shows the mechanism that Hilvert et al. proposed (PRM, proposed reaction mechanism). Through TPS calculations we observe a different mechanism (OBM, observed mechanism) that we discuss later on.

We generated 120 reactive trajectories of length 500 fs for each variant of the RA95 series for the observed mechanism, and 120 reactive trajectories of length 500 fs for only the final variant V, which proceed according to the proposed mechanism.

OBM is possible in all five variants and does not require a catalytic base. As shown in Figure 3A it involves an intramolecular transfer of hydrogen H16 from the Oil hydroxyl group to O17–HZ1 hydroxyl group (Figure 3B). C13–O17 bond breaks and this event results

in the formation of a water leaving group (HI6–O17–HZI) and subsequent Cl2–Cl 1 bond cleavage.

In variant I, Glu53 could have acted as a catalytic base; however, this residue is positioned unfavorably with respect to the substrate: as shown in Table S2, the smallest observed distance was 6.5 Å. In variants II, III, IV, and V the active site has undergone structural rearrangement, including replacement of Lys210 that forms the Schiff base with Lys83. In these variants Tyr51 could act as the catalytic base; however, both the position and orientation are unfavorable for catalysis. Table S2 shows that the distance between Tyr51 and O1 of the substrate in the equilibrated structures for variant II is 5.9 Å, in variant III is 5.7 Å, and in variant IV is 9.1 Å. Despite the structural rearrangement of the original catalytic apparatus and the mutations between variants, the observed mechanism is present in all four variants I–IV. This observed mechanism is also present in the most evolved enzyme V, even though this enzyme has a favorably positioned catalytic base (Table S2). These findings are further supported by a recent computational study³³ that also points out the unfavorable active site construct with respect to catalysis that we mentioned above.

In the mechanism Hilvert et al. proposed the catalytic base can be Glu53 for I, Tyr51 for II–IV, and either Tyr180 or Tyr51 for variant V (Figure 2A–C).³⁴ This proposed mechanism PRM was only observed in the most evolved variant V, where Tyr180 acts as the catalytic base. In variants II–IV residue 180 was a phenylalanine, which cannot act as a catalytic base, and as a result our generated reactive trajectories do not show the PRM mechanism.

Reaction Coordinate.

From the harvested reactive trajectories we identified ten uncorrelated transition state structures for each enzyme. The five different evolved enzymes have structural differences among their transition state structures. The distances for C–C cleavage for OBM were on average 1.66, 1.62, 1.58, 1.96, and 1.59 Å for variants I–V, respectively. The distances for the O–H bond for water formation for OBM were on average 1.41, 1.25, 1.13, 1.04, and 0.96 Å for variants I–V, respectively. We note that there was an overall decrease in bond lengths at the transition state as we go from the least to the most evolved enzyme. The distances for PRM in the most evolved enzyme V were 1.80 Å for C–C cleavage and 0.99 Å for water formation. That is the mechanism progresses from an earlier to a later transition state as evolution proceeds.

We employed committor distribution analysis (described in Methods) to identify the reaction coordinate for all five enzymes. When we constrained only the substrate molecule, the resulting committor distribution did not peak at 0.5 for any of the five designs (Figure S2), implying there are protein residues that play a role in the reaction. Again, this does not mean that there are protein promoting vibrations involved. We identified the reaction coordinate for OBM for all 5 variants and PRM for variant V. The successful committor distributions are shown in Figure S3.

In variant I residues Glu53, Phe89, Leu131, Leu159, and Met 180 are part of the reaction coordinate (Figure 4A). In Table S3 we show average interaction distances between the substrate and the reaction coordinate residues at key points along reactive trajectories. Phe89

lies perpendicular to and underneath the substrate, interacting with C12 and the naphthyl moiety at approximately 3 Å, sterically positioning them for the reaction. Glu53, although not acting as a catalytic base, interacts with the NZ–C13–C14 chain, until C13–O17 and C11–C12 bonds break. Both Leu 159 and Leu131 are sterically interacting with the O17–C13–C14 chain, orienting and holding it until the water leaving group has formed. This is the case for Met 180 too, as it is within 2.5 Å of O17 through its terminal methyl, just before the first bond C13–O17 breaks.

In variant II there is a different active site design (Figure 2B), with Lys83 now forming the Schiff base rather than Lys210. The reaction coordinate includes residues Phe89, Ile133, Leu159, Ser10, and Tyr51 (Figure 4B) and is associated with the committor distribution seen in Figure S2b. We found that Phe89 and Leu 159 play similar roles as in variant I. Interaction distances between the substrate and the reaction coordinate residues are shown in Table S4. Ser10 is keeping O17 in place, while Ile133 is acting as a steric wall to the naphthyl moiety of the substrate, until O17 rotates to face O17 so that the intramolecular hydrogen abstraction can happen. As previously stated, Tyr51 is not in position to facilitate the hydrogen transfer; however, it does move toward the NZ of the Schiff base, which is in the opposite direction from where the reaction takes place.

In variant III residues Asn110 and Phe180 appear in the reaction coordinate (Figure 4C). Residue 110 was mutated from Serine in enzyme II to asparagine, and residue 180 had been mutated to a phenylalanine in enzyme II. RMSF calculations suggest that Ser10 in enzyme II was far less flexible than Asn110 in enzyme III (Figure S4), and the same was true for Phe180. Phe180 is moving closer to O17 and O17, but since it cannot play the role of a catalytic base, its effect is to push O17 to face the same side as O17 for the intramolecular mechanism. Asn110 moves toward O17 to push it against either Phe180 or Tyr51; however, Phe180 is not polarizable, and while Tyr51 attempts to approach the oxygen, it is unfavorably positioned. Again, Phe89 is positioning C12 for catalysis and now Leu9 interacts with the upper part of the naphthyl moiety. Table S5 shows the interaction distances between the substrate and the reaction coordinate residues.

In variant IV the committor distribution analysis showed that Phe89, Asn110, and Phe180 are part of the reaction coordinate and act similarly to variants I–III (Figure 4d). The reaction coordinate also includes Tyr51, Phe112, and Leu159. Table S6 shows the average interaction distances between the substrate and the reaction coordinate residues. Phe112 and Leu 159 now line up along with Phe180 near where the water leaving group forms, keeping the substrate in a favorable orientation. Phe112 blocks the Schiff base from “backing up” and Leu159 serves as a wedge, ensuring Phe112 and Phe180 are properly positioned until the reaction takes place. The reappearance of residue 159 is likely due to mutations making it less flexible in III and then more flexible in IV. Figure S5 shows the RMSF comparison for Leu 159 in variants II–IV.

Finally, we turn our attention to the most evolved enzyme V, which has both the OBM and PRM mechanisms, and discuss the reaction coordinates for both mechanisms. The reaction coordinate for OBM includes residues Leu9, Leu53, Ser56, Phe112, Met182, Leu210, Ser233, Tyr180, and Tyr51 (Figure 5A). Leu53, Ser56, and Leu9 surround the naphthyl moiety and

help orient the substrate within the active site. Phe112 has the same role as in IV, but is now positioned closer to the NZ–C13–C14 chain. Met 182, due to its size, is slightly pushing C11–O11 toward Tyr 180/51. Note that in this final variant V residue 180 was mutated from a phenylalanine to tyrosine. Ser233 is behind Tyr51 and Leu210 is behind Tyr 180, keeping them positioned near the formation of the water leaving group.

The reaction coordinate for PRM includes residues Leu53, Phe89, Leu 159, Leu210, Met231, Met182, Thr178, Tyr180, Tyr51, and Asn110 (Figure 5B). As we can see from the location of these residues for the two mechanisms, the difference lies mostly near the region of the chemical event. Tables S7 and S8 show the interaction distances between substrate atoms and reaction coordinate residues for variant V, OBM, and PRM, respectively. It is reasonable that the reaction coordinate of PRM includes mostly residues that are located near Tyr 180, whereas in OBM includes important residues surrounding the substrate. We used for PRM the reaction coordinate residues identified for OBM, and obtained a committor distribution that peaks at 0.6. On the other hand when we used for OBM the reaction coordinate residues of PRM, the committor distribution shifts to the product region (Figure S6). RMSF calculations clarify the reason why some residues appear in the reaction coordinate of one mechanism and not to the other, e.g., Ser56 and Phe112 are more mobile in OBM (Figure S7). This means in these trajectories, those residues experience greater displacements.

Returning to the role of residues in the reaction coordinate for PRM, Phe89 is interacting with C12 of the substrate preventing this region from backing away during the reaction. Met182 slightly pushes C11–O11, and Thr178 (which was mutated from Valine) along with Leu 159 are interacting with Tyr180, keeping it pointed between O11 and O17. Both help to position efficiently Tyr180, which is the catalytic base. Apart from residue 178, two more residues posterior to Tyr180 were mutated in the most evolved enzyme V: Leu210 from lysine and Met231 from leucine. Both were identified to be important from the committor distribution analysis. The nature of these mutations suggests that the active site has become more hydrophobic and compact. In the next section we will analyze their role in detail, as they are part of a significant rate promoting vibration.

To sum up the results from the reaction coordinate identification, we examined the change of the reaction coordinate residues as we move from variant I to variant V. We found an increase in the number, size, and polarizability of the reaction coordinate residues (Table S9). In all variants, the large majority of reaction coordinate residues are hydrophobic. Comparing, for example, the reaction coordinate residues of variants II and V, the number of residues that are included in the reaction coordinate has doubled from five to ten, the residues are bulkier, and two polarizable residues have been added.

Figure S8 shows the sequence difference between variant I and both variant V PRM and OBM. We note that most of the residues that have been mutated are part of the PRM reaction coordinate, strongly indicating that directed evolution favors the proposed mechanism as an eventual target. Since this series of artificial enzymes has undergone several rounds of mutation, it is reasonable to suspect involvement of electrostatic effects in catalysis. We looked into possible electrostatic effects over the course of directed evolution

by performing a Mulliken charge analysis and present the results for variants I, III, and V (OBM and PRM) for all the atoms of the substrate in Figure S9. All charges have been calculated for the equicommitter point (eq), 10 fs before the equicommitter point (eq – 10), 50 fs before the equicommitter point (eq – 50), and 100 fs before the equicommitter point (eq – 100). The charge distributions are practically the same, and there is no evidence of charge delocalization. The charges change along the reaction event similarly for all the variants. In other words, electrostatic effects on the substrate are not the key to the reaction step; or at a minimum, they are unaffected by the evolutionary process and so not part of the optimization.

Rate Promoting Vibrations.

TPS reactive trajectories contain all the atomistic details and dynamical information about the role of residues associated with the reaction. In the most evolved variant V, there is a rate promoting vibration that involves Tyr180 as the catalytic base. Tyr180 moves close to O17 (Figure 6D) and donates its hydrogen and simultaneously abstracts a hydrogen from O11 as the C11–C12 bond breaks. This motion is assisted by the two residues behind Tyr180, Lys210, and Met231, which push the catalytic base.

As we show in the snapshots (Figure 6A–C) the ring of Tyr180 flips and the residue attacks the substrate's oxygen. Lys210 and Met231 are flexible residues facilitating the push, which allows the PRM mechanism to be realized in this enzyme. We will now show why this motion is indeed a rate promoting vibration. To test this we took a reactive trajectory from the PRM ensemble, restrained residues Lys210 and Met231 and used transition path sampling to generate a new ensemble, which cannot include the rate promoting vibration, since we constrained residues that participate in the promoting motion (Movies in Supporting Information). After 10 generated reactive trajectories the mechanism switched from PRM, which involves the catalytic base, to OBM where we have intramolecular water formation. This ensemble remained on OBM paths for the rest of the transition path generation (total of 50 trajectories). We point out that this is an unnatural ensemble of reactive trajectories; nevertheless, it provides strong evidence that this is a significant rate promoting vibration, and it was introduced by the evolutionary process.

Chemical Barrier Calculation.

We calculated the chemical barrier using a method we have developed,³⁵ where the free energy barrier to reaction is calculated as the ensemble average of the work produced on the bond-cleavage atoms. This should be sufficient for estimating the correct trend for the barriers for the different mechanisms in variant V.

We calculated the barriers for the cleavage of the C11–C12, C13–O17, and O11–H15 bonds in variant V. We used representative reactive trajectories from the OBM and from the PRM ensemble, calculated the barriers, and then averaged them. We also calculated the barriers for the cleavage of the C11–C12, C13–O17, and O11–H15 bonds for the reactive trajectories of the PRM restrained ensemble, where we fixed in space the rate promoting vibration residues, before the mechanism switched to the OBM (Table 1).

Our results indicate that the PRM mechanism has lower barriers than OBM. Comparing the values to the PRM and restrained PRM we see that the contribution of the rate promoting vibration is approximately 1.4 kcal/mol for the O1–H15 bond, 7 kcal/mol for the C13–O17 bond, and –0.6 kcal/mol for the C11–C12 bond.

■ CONCLUSION

The motivation for this work was the successful use of directed evolution¹ to improve an initial computationally designed enzyme into a series of artificial enzymes, with the final variant exhibiting a rate competitive with the naturally occurring enzyme. The present work is the first study incorporating all five variants from the retro-aldolase series, examining in atomic detail how directed evolution changed the series of variants, and how these changes in turn altered the details of the catalytic mechanism.

Our simulations revealed that in all enzymes the chemical step can utilize a different mechanism than the proposed one. In the final, most evolved variant we find that Tyr180 plays the role of a catalytic base, and in this variant we discovered a femtosecond motion that pushes the catalytic base toward the substrate to facilitate catalysis.

We identified the residues that are part of the reaction coordinate in each of the five enzymes and determined their roles. Going from the initial variant I (RA95.0) to the final variant V (RA95.5–8F) we found that the reaction coordinate residues in the active site tend to be hydrophobic, more polarizable, larger, and increased in number. Comparing the original design to the final enzyme V, the large majority of residues that were mutated during directed evolution were found to be part of the reaction coordinate, strongly suggesting that directed evolution favors this PRM design. Residue 180 was mutated to a Tyrosine in the most evolved enzyme V, acting as the catalytic base, which is a reason we did not find the PRM mechanism in variants I–IV. Notably, residues Lys210 and Met231, which are involved in the promoting vibration, were also mutated to their current form only in the final variant V. In other words, the protein motion examined earlier seems to be the dominant factor for the success of the design series as the charge distribution analysis did not point to any electrostatic effects. Finally, the results of this study strongly support the necessity to include rapid protein dynamics as a part of enzyme design, and provides a partial rationalization of the modest activity of artificial enzymes designed using static structural considerations alone.

■ METHODS

The starting point for all simulations were crystal structures 4A29, 4A2S, 4A2R, and 5AN7 from the Protein Data Bank, which correspond to enzymes RA95.0, RA95.5, RA95.5–5, and RA95.5–8F, respectively. Variant RA95.5–8 was generated from 5AN7 using SWISS-MODEL software to create a homology model. We modified the 3NK (1-(6-methoxynaphthalen-2-yl)butane-1,3-dione) inhibitor in the first three crystal structures and LLK ((2E)-1-(6-methoxy-naphthalen-2-yl)but-2-en-1-one) in the last structure into 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone, which is the substrate of the catalyst. We obtained parameters for 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone according to

CHARMM force field using PARAMCHEM. We then deprotonated Lys210 for the variant I and Lys83 in variants II–V in order to form the Schiff base of the mechanism.

All molecular dynamics simulations were performed using the CHARMM molecular dynamics package.^{36,37} Each system was then partitioned into quantum and molecular mechanics regions. The quantum region of enzyme I consisted of the substrate, Lys83, and Glu53. The quantum region for enzymes II–III consisted of the substrate, Lys83, and Tyr51. In enzymes IV–V the quantum regions contains the substrate and residues 51, 83, 110, and 180. We note that residue 180 is mutated from Phenylalanine in enzyme IV to Tyrosine in the enzyme V. The molecular mechanics region for all enzymes includes the rest of the protein, ions, and solvent.

The generalized hybrid orbital (GHO) method³⁸ was used to couple the two regions through the *CB* of residues included in the quantum region. The quantum region was modeled using the PM3 semiempirical method, and the molecular region was modeled using the CHARMM36 force field. The simulations for all the enzymes were performed following the same protocol described below.

Each protein was solvated within a 60 Å sphere using TIP3 waters and neutralized using potassium ions. The system was then minimized for 100 steps using the steepest descent method followed by 2000 steps of adopted basis Newton–Raphson method. The system was then heated slowly from 0 to 300 K for 150 ps and equilibrated at 300 K for 100 ps. During the previously described steps the quantum region was switched on.

We use transition path sampling³⁹ to generate and analyze the ensemble of reactive trajectories. First we define the order parameter to be the distance between the C atoms involved in the cleavage event. We define the reactant state to be a distance smaller than 1.70 Å and the product state to be a distance large than 1.70 Å. To begin the path ensemble we create a biased trajectory by applying a harmonic force and allow the system to propagate for 250 fs. In order to generate new trajectories, we perturbed the momenta of a randomly chosen slice of a reactive trajectory, and using the new momenta we allowed the system to propagate forward and backward in time for 250 fs. Several iterations are needed in order to ensure decorrelation from the initial biased trajectory.

We employed committer analysis^{40,41} to identify the transition state of trajectories within the path ensemble. The committer probabilities are obtained by picking slices along a trajectory within the path ensemble and initializing dynamics by assigning random momenta to the atoms. By initiating several trajectories from each slice we obtain the committer probability. The transition state is defined as the structure in a trajectory with an equal probability to commit either to reactants or products. The collection of transition states within the ensemble yields the stochastic separatrix. This is then used to systematically search for the reaction coordinate, which are the degrees of freedom within the system that are orthogonal to the stochastic separatrix.^{42,43} One makes an initial choice of residues to restrain, then propagates the system and calculates the commitment probabilities from slices along this restrained walk. If the choice of residues to constrain were correct, the system

would remain on the separatrix and the committor probabilities will be peaked at 0.5. If it is peaked elsewhere, we make a different choice of residues to restrain and repeat.

Mulliken charges were calculated with the PM3 method in CHARMM. We selected four time slices of chemical interest along four randomly picked reactive trajectories for which we had identified transition state structures, and averaged the results over trajectories.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

■ ACKNOWLEDGMENTS

All computer simulations were performed at the University of Arizona High Performance Computing Center, on a Lenovo NeXTScale nx360 M5 supercomputer. This research was supported through the NIH program project grant P01GM127594 and NIH grant R01GM127594.

■ REFERENCES

- (1). Obexer R; Godina A; Garrabou X; Mittl PRE; Baker D; Griffiths AD; Hilvert D Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem* 2017, 9, 50–56. [PubMed: 27995916]
- (2). Romero P; Arnold F Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol* 2009, 10, 866. [PubMed: 19935669]
- (3). Brustad EM; Arnold FH Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol* 2011, 15, 201–210. [PubMed: 21185770]
- (4). Rothlisberger D; Khersonsky O; Wollacott AM; Jiang L; DeChancie J; Betker J; Gallaher JL; Althoff EA; Zanghellini A; Dym O; Albeck S; Houk KN; Tawfik DS; Baker D Kemp elimination catalysts by computational enzyme design. *Nature* 2008, 453, 190–195. [PubMed: 18354394]
- (5). Siegel JB; Zanghellini A; Lovick HM; Kiss G; Lambert AR; Clair JLS; Gallaher JL; Hilvert D; Gelb MH; Stoddard BL; Houk KN; Michael FE; Baker D Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 2010, 329, 309–313. [PubMed: 20647463]
- (6). Dwyer MA; Looger LL; Hellinga HW Computational Design of a Biologically Active Enzyme. *Science* 2004, 304, 1967–1971. [PubMed: 15218149]
- (7). Nanda V; Koder RL Designing artificial enzymes by intuition and computation. *Nat. Chem* 2010, 2, 15–24. [PubMed: 21124375]
- (8). Ward T R Artificial Enzymes Made to Order: Combination of Computational Design and Directed Evolution. *Angew. Chem. Int. Ed* 2008, 47, 7802–7803.
- (9). Kiss G; Çelebi Ölçüm N; Moretti R; Baker D; Houk KN Computational Enzyme Design. *Angew. Chem. Int. Ed* 2013, 52, 5700–5725.
- (10). Arakaki TL; Pezza JA; Cronin MA; Hopkins CE; Zimmer DB; Tolan DR; Allen KN Structure of human brain fructose 1,6-(bis)phosphate aldolase: Linking isozyme structure with function. *Protein Sci.* 2004, 13, 3077–3084.
- (11). Gefflaut T; Blonski C; Perie J; Willson M Class I aldolases: Substrate specificity, mechanism, inhibitors and structural aspects. *Prog. Biophys. Mol. Biol* 1995, 63, 301–340. [PubMed: 8599032]
- (12). Hong N; Petrovic D; Lee R; Grynova G; Purg M; Saunders J; Bauer P; Carr P; Lin C; Mabbitt P; Zhang W; Altamore T; Easton C; Coote M; Kamerlin S; Jackson C The evolution of multiple active site configurations in a designed enzyme. *Nature Commun.* 2018, 9, 3900. [PubMed: 30254369]
- (13). Petrovic D; Kamerlin SCL Molecular modeling of conformational dynamics and its role in enzyme evolution. *Curr. Opin. Struct. Biol* 2018, 52, 50–57. [PubMed: 30205262]

- (14). Campbell EC; Correy GC; Mabbitt RD; Buckle AM; Tokuriki N; Jackson CJ Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struct. Biol* 2018, 50, 49–57. [PubMed: 29120734]
- (15). Chen X; Schwartz SD Directed evolution as a probe of rate promoting vibrations introduced via mutational change. *Biochemistry* 2018, 57, 3289–3298. [PubMed: 29553716]
- (16). De Raffele D; Marti S; Moliner V QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catal.* 2019, 9, 2482–2492.
- (17). Schwartz SD; Schramm VL Enzymatic transition states and dynamic motion in barrier crossing. *Nat. Chem. Biol* 2009, 5, 551–558. [PubMed: 19620996]
- (18). Hay S; Scrutton N Good vibrations in enzyme-catalysed reactions. *Nat. Chem* 2012, 4, 161–168. [PubMed: 22354429]
- (19). Pudney C; Guerriero A; Baxter N; Johannissen L; Waltho J; Hay S; N, S. S. Fast protein motions are coupled to enzyme H-transfer reactions. *J. Am. Chem. Soc* 2013, 135, 2512–2517. [PubMed: 23373704]
- (20). Agarwal PK; Billeter SR; Rajagopalan PTR; Benkovic SJ; Hammes-Schiffer S Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A* 2002, 99, 2794–2799. [PubMed: 11867722]
- (21). Benkovic S; Hammes-Schiffer S Biochemistry - Enzyme motions inside and out. *Science* 2006, 312, 208–209. [PubMed: 16614206]
- (22). Kohen A Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res* 2015, 48, 466–473. [PubMed: 25539442]
- (23). Caratzoulas S; Mincer JS; Schwartz SD Identification of a protein promoting vibration in the reaction catalyzed by horse liver alcohol dehydrogenase. *J. Am. Chem. Soc* 2002, 124, 3270–3276. [PubMed: 11916410]
- (24). Basner JE; Schwartz SD How enzyme dynamics helps catalyze a reaction, in atomic detail: a transition path sampling study. *J. Am. Chem. Soc* 2005, 127, 13822–13831. [PubMed: 16201803]
- (25). Qpaytman S; Schwartz S Reaction coordinates of an enzymatic reaction revealed by transition path sampling. *Proc. Natl. Acad. Sci. U. S. A* 2007, 104, 12253–12258. [PubMed: 17640885]
- (26). Masterson JE; Schwartz SD Changes in protein architecture and subpicosecond protein dynamics impact the REaction catalyzed by lactate dehydrogenase. *J. Phys. Chem. A* 2013, 117, 7107–7113. [PubMed: 23441954]
- (27). Nunez S; Wing C; Antoniou D; Schramm VL; Schwartz SD Insight into catalytically relevant correlated motions in human purine nucleoside phosphorylase. *J. Phys. Chem. A* 2006, 110, 463–476. [PubMed: 16405318]
- (28). Saen-oon S; Ghanem M; Schramm V; Schwartz S Remote mutations and active site dynamics correlate with catalytic properties of purine nucleoside phosphorylase. *Biophys. J* 2008, 94, 4078–4088. [PubMed: 18234834]
- (29). Johannissen L; Hay S; Scrutton NS; Sutcliffe MJ Proton tunneling in aromatic amine dehydrogenase is driven by a short-range sub-picosecond promoting vibration: consistency of simulation and theory with experiment. *J. Phys. Chem. B* 2007, 10, 2631–2638.
- (30). Kamerlin SC; Warshel A At the dawn of the 21st century: is dynamics the missing link for understanding enzyme catalysis? *Proteins* 2010, 78, 1339–1375. [PubMed: 20099310]
- (31). Warshel A; Sharma P; Kato M; Xiang Y; Liu H; Olsson M Electrostatic basis for enzyme catalysis. *Chem. Rev* 2006, 106, 3210–3235. [PubMed: 16895325]
- (32). Garcia-Meseguer R; Marti S; Ruiz-Pernia JJ; Moliner V; Tunon I Studying the role of protein dynamics in an SN2 enzyme reaction using free-energy surfaces and solvent coordinates. *Nat. Chem* 2013, 5, 566–571. [PubMed: 23787745]
- (33). Romero-Rivera A; Garcia-Borras M; Osuna S Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* 2017, 7, 8524–8532. [PubMed: 29226011]
- (34). 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. [PubMed: 28783336]

- (35). Dzierlenga M; Antoniou D; Schwartz SD Another Look At The Mechanisms Of Hydride Transfer Enzymes With Quantum And Classical Transition Path Sampling. *J. Phys. Chem. Lett* 2015, 6, 1177–1181. [PubMed: 26262969]
- (36). Brooks BR; et al. CHARMM: the biomolecular simulation program. *J Comput. Chem* 2009, 30, 1545–1614. [PubMed: 19444816]
- (37). Brooks B; Bruccoleri R; Olafson B; States D; Swaminathan S; Karplus M CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem* 1983, 4, 187–217.
- (38). Gao J; Arnara P; Alhambra C; Field M A generalized hybrid orbital (GHO) approach for the treatment of link-atoms using combined QM/MM potentials. *J. Phys. Chem. A* 1998, 102, 4714–4721.
- (39). Bolhuis P; Dellago C Practical and conceptual path sampling issues. *Eur. Phys. J.: Spec. Top* 2015, 224, 2409–2427.
- (40). Antoniou D; Schwartz S Protein dynamics and enzymatic chemical barrier passage. *J. Phys. Chem. B* 2011, 115, 15147–15158. [PubMed: 22031954]
- (41). Zoi I; Suarez J; Antoniou D; Cameron D; Schramm VL; Schwartz SD Modulating enzyme catalysis through mutations designed to alter rapid protein dynamics. *J. Am. Chem. Soc* 2016, 138, 3403–3409. [PubMed: 26927977]
- (42). Zoi I; Motley M; Antoniou D; Schramm VL; Schwartz SD Enzyme Homologues Have Distinct Reaction Paths through their Transition States. *J. Phys. Chem. B* 2015, 119, 3662–3668. [PubMed: 25650981]
- (43). Antoniou D; Schwartz SD The stochastic separatrix and the reaction coordinate for complex systems. *J. Chem. Phys* 2009, 130, 151103. [PubMed: 19388729]

Figure 1.
Reaction scheme showing the major steps in amine-catalyzed retro-aldol reaction.

Figure 2.
(A–C) Crystal structures for variant I (A), variant III (B), and variant V (C) showing the different active site construct with catalytic lysine and active site residues proposed to act as the catalytic base.

Figure 3.
(A) OBM scheme. (B) Substrate with lysine covalently bound and key atoms labeled for the OBM mechanism description.

Figure 4.
(A–D) Reaction coordinate residues in variant I (A), variant II (B), variant III (C), and variant IV (D).

Figure 5.

(A) Reaction coordinate residues in RA95.5-8f OBM. (B) Reaction coordinate residues in RA95.5-8f PRM.

Figure 6.

Snapshots (A–C) along a representative trajectory of enzyme V, going from reactants to the transition state; (D) the promoting motion from Leu210 and Met231 to Tyr180, as Tyr180 approaches the substrate oxygen to facilitate the hydrogen donation and abstraction and the Asn110 interaction with the substrate in (i) a reactant configuration, (ii) at 100 fs, (iii) at 230 fs.

Table 1.

Average Barrier Heights of Bond Cleavages in kcal/mol for the OBM, PRM (Variant V), and Restrained PRM Mechanism

	O11-H15	C13-O17	C11-C12
OBM	30.6 ± 5.3	10.8 ± 1.7	2.1 ± 0.7
PRM	15.5 ± 0.3	2.8 ± 0.2	1.6 ± 0.7
restrained PRM	16.9 ± 1.4	10 ± 4	1 ± 0.2