Origins of catalysis by computationally designed retroaldolase enzymes

Jonathan K. Lassila*, David Baker*, and Daniel Herschlag†

*Department of Biochemistry, Stanford University, Stanford, CA 94305; and †Department of Biochemistry, University of Washington, Seattle, WA 98195

Edited* by Stephen J. Benkovic, The Pennsylvania State University, University Park, PA, and approved January 27, 2010 (received for review November 25, 2009)

We have investigated recently reported computationally designed retroaldolase enzymes with the goal of understanding the extent and the origins of their catalytic power. Direct comparison of the designed enzymes to primary amine catalysts in solution revealed a rate acceleration of $10^{10}$-fold for the most active of the designed retroaldolases. Through pH-rate studies of the designed retroaldolases and evaluation of a Bronsted correlation for a series of amine catalysts, we found that lysine pK_a values are shifted by 3–4 units in the enzymes but that the catalytic contributions from the shifted pK_a values are estimated to be modest, about 10-fold. For the most active of the reported enzymes, we evaluated the catalytic contribution of two other design components: a motif intended to stabilize a bound water molecule and hydrophobic substrate binding interactions. Mutational analysis suggested that the bound water motif does not contribute to the rate acceleration. Comparison of the rate acceleration of the designed substrate relative to a minimal substrate suggested that hydrophobic substrate binding interactions contribute around $10^9$-fold to the enzymatic rate acceleration. Altogether, these results suggest that substrate binding interactions and shifting the pK_a of the catalytic lysine can account for much of the enzyme’s rate acceleration. Additional observations suggest that these interactions are limited in the specificity of placement of substrate and active site catalytic groups. Thus, future design efforts may benefit from a focus on achieving precision in binding interactions and placement of catalytic groups.

B because natural enzymes catalyze reactions with tremendous rate accelerations and specificities, a long-standing goal in enzymology and protein engineering has been to reliably design new enzyme catalysts for chemical reactions of interest. Computational protein design offers a promising tool for achieving this goal. Computational protein design methods use specialized potential energy functions in combination with search algorithms to optimize an amino acid sequence for a given protein structure and function (1–3). These methods have been adapted to the challenge of designing enzyme active sites (4–8).

Although new enzymatic activities have been designed computationally, the resulting catalysts have catalytic efficiencies ($k_{cat}/K_M$ values) of about 1–100 M$^{-1}$ s$^{-1}$ (4,7,8), considerably less than values of $10^7–10^9$ M$^{-1}$ s$^{-1}$ typical of natural enzymes, and similar to those of early catalytic antibodies. In contrast to catalytic antibody methods and other stochastic processes, however, the potential for success of computational enzymatic design is tied to the predictive power of the computational model. Thus, future improvement of our ability to computationally design enzymatic activity will require ongoing rigorous assessment of the successes and failures of the design process.

We therefore sought to investigate the origins of the catalytic power of recently reported computationally designed retroaldolases (7). One of few examples of catalytic activities designed using computational modeling alone, these enzymes were developed to catalyze the retroaldol cleavage of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butane (Fig. 1), a fluorogenic substrate developed for catalytic antibody studies (9). This retroaldol reaction is catalyzed by amines through the formation of an iminium (or Schiff base) intermediate. Analogous to the strategy used by type I aldolases (10), formation of the iminium intermediate with the enzymatic lysine side chain provides an electron sink, facilitating the retroaldol cleavage. Several catalytic antibodies and peptide systems have been developed that utilize this lysine iminium strategy (11–16).

The lysine side chain provides a key catalytic element in the enzymatic reaction, yet the free lysine side chain in solution can also catalyze the retroaldol reaction. Thus, to evaluate the contribution of the computationally designed active sites to catalysis, we first determined the rate acceleration of the enzymes beyond that of the lysine side chain alone. Next, we asked how the computational design procedure facilitates additional catalysis beyond that of the lysine side chain alone. We investigated the catalytic contribution of each of three elements used in designing the enzymes: a hydrophobic pocket intended to lower the pK_a of the catalytic lysine, a stabilized, positioned water molecule for facilitation of proton transfer, and binding interactions with the substrate (Fig. 2) (7). We provide an accounting for much of the catalysis by the most active of the computationally designed retroaldolases and use these results to evaluate strengths and limitations of current enzyme design methodology.

Results and Discussion

Conditions for Kinetic Studies. To investigate mechanisms of catalysis in an enzymatic system, it is necessary to identify assay conditions in which the enzyme is stable and the substrate is soluble. We found that substrate solubility was exceeded in previously used conditions for the designed enzymes (7,17). Thus, we tested buffers and cosolvents and identified conditions that provided improved solubility while protein stability was maintained (see Materials and Methods). For the variants tested herein, the enzymes could not be saturated with substrate under conditions where the substrate remained soluble. We therefore focused exclusively on second-order rate constants ($k_{cat}/K_M$), measured under subsaturating conditions where initial rates increased linearly with both enzyme and substrate concentration ($v_0 = k_{cat}/K_M[E][S]$). The resulting values of $k_{cat}/K_M$ are within 2-fold of those from the original report for the variants tested (Table 1) (7).

The second-order rate constant $k_{cat}/K_M$ reports on the difference in energy between the free enzyme and substrate in solution and the rate-limiting transition state in the steps up to and including formation of the 6-methoxy-2-naphthaldehyde product that is monitored by fluorescence (steps 1–3, Fig. 1). These rate constants are especially useful for our purposes because they can be directly compared to the second-order rate constants for

Author contributions: J.K.L. and D.H. designed research; J.K.L. performed research; and J.K.L., D.B., and D.H. analyzed data.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

†To whom correspondence should be addressed. Email: herschlag@stanford.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0913638107/DCSupplemental.
the corresponding nonenzymatic reactions of primary amines in solution. This comparison provides a direct evaluation of the transition state stabilization provided by the designed enzyme active sites.

For practical reasons described below, we obtained $k_{\text{cat}}/K_M$ values from initial rates under pre-steady-state conditions, before the first turnover. Under subsaturating conditions, $k_{\text{cat}}/K_M$ values obtained before the first turnover are the same as those obtained under multiple-turnover conditions. For the designed retroaldolases, however, tight binding of the fluorescent product creates a special complication that can make multiple-turnover conditions difficult to achieve. This complication is described in the following paragraphs because it results in unexpected assay behavior that could be misinterpreted as a kinetic burst, and because it has interesting implications for the specificity of the active site. However, the present studies used initial rates measured under conditions where only low concentrations of product were formed, so the complication described below does not affect the kinetic results reported herein.

As the 6-methoxy-2-naphthaldehyde product builds up, it can return to the active site to form a covalent Schiff base species with the active site lysine (SI Appendix). This covalent enzyme-product species is not a part of the reaction pathway, and its stability and rate of formation may be enhanced by the aldehyde functionality of the product relative to the ketone of the substrate (18).

Formation of the covalent enzyme-product species reduces the reaction rate in the fluorescence-based assay in two ways. First, the concentration of free enzyme available to react is reduced; this reduces the actual rate of the reaction. Second, the measured rate is reduced because the fluorescent signal of the product is lost upon binding to the enzyme. For all of the retroaldolase variants tested, an exponential loss of fluorescent signal was evident when 6-methoxy-2-naphthaldehyde was incubated with enzyme in the absence of substrate (SI Appendix).

During a reaction time course beginning with substrate, this rebinding of product can manifest as an apparent burst of fluorescence (SI Appendix). Although the reductions in the rate of fluorescence development arising from this mechanism can appear burst-like, they need not follow normal expectations for burst kinetics. Reductions in apparent rate arising from this mechanism can occur under second-order conditions where the enzyme is almost entirely unsaturated by substrate, when the enzyme is in excess, and well after or before a full turnover of the enzyme.

Thus, there are two sources of burst-like kinetics to consider in this system. First, traditional burst kinetics may arise under saturating conditions due to the occurrence of a rate-limiting step subsequent to product formation, such as proton transfer or release of acetone in steps 4 and 5 (Fig. 1). Second, the apparent burst described above relates to product rebinding to the enzyme in an off-pathway state and does not provide kinetic information about steps 4 and 5. For example, RA61 shows continuous reaction progress curves over multiple turnovers at low enzyme concentration but curvature in the fluorescence time course at higher enzyme concentration where product rebinding is more pronounced (SI Appendix). We estimate a $K_p$ for product dissociation from RA61 of about 10–30 μM, considerably lower than the $K_M$ for product rebinding, which is greater than 500 μM (SI Appendix).

Because the present studies were performed under conditions where little product builds up during the reactions and with enzyme concentrations below the $K_p$ for product rebinding, the fluorescence loss due to product binding does not complicate our

---

**Fig. 1.** Major steps in amine-catalyzed retroaldol reaction. Proton transfers, binding events, and some intermediate steps are omitted for clarity.

**Fig. 2.** Three elements comprise the catalytic motif used for the retroaldolases investigated in this paper (7), illustrated here with the designed model for the most active enzyme, RA61. The experimental crystal structure for RA61 did not contain ligand, and the binding orientation of the substrate is not known. (A) A hydrophobic pocket was intended to lower the $pK_a$ of the catalytic lysine side chain. (B) Hydrogen-bonding interactions to a bound water molecule were incorporated into the design models. In RA61, the side chains modeled in contact with the bound water are Tyr78 and Ser87. (C) Hydrophobic side chains line the active site cavity, providing a binding surface for the hydrophobic substrate.
efforts to investigate contributions to catalysis of the retroaldol reaction (steps 1–3). Nevertheless, this unexpected and tight product inhibition illustrates how one hallmark of natural enzymes, their exquisite specificity, remains a challenge for enzyme engineering.

Rate Acceleration Relative to Lysine Side Chain in Solution. In the protein engineering literature, rate acceleration is frequently reported as $k_{cat}/k_{uncat}$, where $k_{uncat}$ often represents the first-order, nonenzymatic water-mediated reaction. However, mechanistically appropriate small-molecule catalysts in solution can provide another useful comparison (19–21). In the case of the retroaldolases, the enzymatic reaction is catalyzed by a lysine side chain that is required for activity (7), whereas the reaction in water is specific-base catalyzed through a different chemical mechanism. Thus, for understanding the catalytic contribution of the designed active sites, a more direct measure of rate acceleration is provided by comparing the second-order rate constant for the enzymatic active site lysine with that of the lysine side chain alone, free in solution.

For comparison to the retroaldolase rate constants, Table 1 includes a value for the nonenzymatic, second-order rate constant of the lysine side chain in solution at pH 7.5, taken from the value for a primary alkylamine of $pK_a$ 10.6, as described in a later section. The most active of the reported retroaldolases has a rate acceleration of $2 \times 10^5$ relative to the lysine side chain in solution, and the other enzyme rate constants range from $10^2$- to $10^4$-fold above lysine in solution.

How do the designed enzymes achieve these rate accelerations beyond those of the catalytic lysine side chains within their active sites? The computational design strategy included three elements intended to contribute to catalysis: a hydrophobic pocket to lower the $pK_a$ of the catalytic lysine, a bound water motif designed to facilitate proton transfers, and binding interactions with the substrate (Fig. 2) (7). We investigated the contributions provided by each of these elements to the rate acceleration of the designed retroaldolases, with a focus on the most active of these enzymes, RA61.

Catalytic Contribution of Shifted Lysine $pK_a$. One of the strategies employed in the computational design process was to attempt to lower the $pK_a$ of the catalytic lysine residue by placing it within a hydrophobic pocket (Fig. 2A). Because the deprotonated, neutral form of lysine is required for formation of the iminium intermediate and progress to products, lowering the $pK_a$ of the catalytic lysine effectively increases the concentration of active enzyme available to react, thereby increasing the observed second-order rate constant for reaction of free enzyme and substrate (Fig. 3A).

To determine whether the computational design process was successful in lowering the $pK_a$ of the catalytic lysine, we measured pH-rate profiles for several of the retroaldolases (Fig. 3B and Table 2). The assignment of the observed $pK_a$ values to the catalytic lysine is discussed further in SI Appendix and is supported by a pH-binding study for RA61. Relative to the solution $pK_a$ of the lysine side chain of 10.6, observed $pK_a$ values are considerably shifted in the retroaldolases, with measured values from 6.8–7.5. These perturbed $pK_a$ values are consistent with prior observations of lysine $pK_a$ shifts in similar enzymes, including acetoacetate decarboxylase and catalytic antibodies and helical peptides that catalyze the same or related reactions (11, 22–24).

Our goal was to estimate the catalytic contribution of these observed $pK_a$ shifts, and for this we needed more information. Lowering the $pK_a$ by one unit can increase the reactivity by raising the concentration of active, neutral lysine by up to 10-fold. However, lowering the $pK_a$ also reduces the reactivity of the neutral amine species, as the amine becomes less nucleophilic. To determine how much the reactivity decreases, we determined the relationship between $pK_a$ and reactivity for a series of nonenzymatic amines in solution. This information, combined with the measured enzymatic $pK_a$ values, allowed us to estimate the catalytic contribution from lowering the $pK_a$ of the enzymatic lysine side chains.

We measured rate constants for the retroaldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone catalyzed by a series of primary amines of the structure $RCH_2NH_2$ across a broad range of $pK_a$ values (Table 3). Plotting the log of the maximal rate constants for the amines against their $pK_a$ values yields the linear relationship with slope of +0.54 ± 0.03 (Fig. 4A). This slope represents a Brønsted value, a quantity related to changes in the rate of a reaction with changes in $pH$.
in charge distribution between the ground state and transition state of the reaction (25, 26). The Brønsted slope provides an empirical relationship between reactivity and pKa for the amine-catalyzed retroaldol reaction. The relationship can be replotted, as in Fig. 4B, to account for the concentration of neutral amine present at the pH value of the reaction conditions, here pH 7.5. The plot in Fig. 4B relates the observed rate constant (kobs) at pH 7.5 to the pKa of the amine and illustrates that the optimal amine catalysts have pKa values similar to the pH of the reaction conditions.

The relationship in Fig. 4B shows that the dependence of kobs on pKa is not steep, and the difference in observed rate constant between an amine of pKa 10.6 and 6.8, as in RA61, is modest, about 10-fold. Thus, an enhancement of ∼10-fold is estimated for RA61, assuming that the enzymatic and nonenzymatic Brønsted values are similar, and similar rate enhancements are estimated for the other designed retroaldolases (Table 2).

Several observations suggest that the nonenzymatic Brønsted value provides a reasonable estimate of the relationship between reactivity and pKa in the enzyme. First, studies of nucleophilic substitution reactions have suggested that changes in pKa values arising from changes in substituent groups (as used to obtain the nonenzymatic Brønsted value here) have similar effects on reactivity to changes in pKa values arising from different solvation environments (as the enzymatic pKa shift presumably entails), and equivalent Brønsted values have been measured for solvation and substituent pKa perturbations (27, 28). Second, although differences in the rate-limiting step between nonenzymatic and enzymatic reactions are possible, a Brønsted value of +0.8 reported for iminium formation with primary amines (29) suggests that if iminium formation rather than the retroaldol step limited either the enzymatic or nonenzymatic reactions, a value larger than +0.54 might be expected. A larger Brønsted value, whether arising from a different rate-limiting step, or as is sometimes observed, from placement in a low dielectric environment, would result in smaller catalytic contributions from the pKa differences in the rate-limiting step between nonenzymatic and enzymatic reactions are possible, a Brønsted value of +0.8 reported for iminium formation with primary amines (29) suggests that if iminium formation rather than the retroaldol step limited either the enzymatic or nonenzymatic reactions, a value larger than +0.54 might be expected. A larger Brønsted value, whether arising from a different rate-limiting step, or as is sometimes observed, from placement in a low dielectric environment, would result in smaller catalytic contributions from the pKa shift presumably entails), and equivalent Brønsted values have been measured for solvation and substituent pKa perturbations (27, 28). Second, although differences in the rate-limiting step between nonenzymatic and enzymatic reactions are possible, a Brønsted value of +0.8 reported for iminium formation with primary amines (29) suggests that if iminium formation rather than the retroaldol step limited either the enzymatic or nonenzymatic reactions, a value larger than +0.54 might be expected. A larger Brønsted value, whether arising from a different rate-limiting step, or as is sometimes observed, from placement in a low dielectric environment, would result in smaller catalytic contributions from the pKa shift presumably entails), and equivalent Brønsted values have been measured for solvation and substituent pKa perturbations (27, 28). Second, although differences in the rate-limiting step between nonenzymatic and enzymatic reactions are possible, a Brønsted value of +0.8 reported for iminium formation with primary amines (29) suggests that if iminium formation rather than the retroaldol step limited either the enzymatic or nonenzymatic reactions, a value larger than +0.54 might be expected. A larger Brønsted value, whether arising from a different rate-limiting step, or as is sometimes observed, from placement in a low dielectric environment, would result in smaller catalytic contributions from the pKa shift presumably entails), and equivalent Brønsted values have been measured for solvation and substituent pKa perturbations (27, 28).

### Table 3. pH-independent second-order rate constants for amine-catalyzed retroaldol reactions of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone

<table>
<thead>
<tr>
<th>Amine</th>
<th>pKa</th>
<th>kmax (M⁻¹ s⁻¹) *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanomethylamine</td>
<td>5.3</td>
<td>5.8 × 10⁻⁶</td>
</tr>
<tr>
<td>Trifluoroethylamine</td>
<td>5.7</td>
<td>7.1 × 10⁻⁶</td>
</tr>
<tr>
<td>Propargylamine</td>
<td>8.2</td>
<td>1.0 × 10⁻⁴</td>
</tr>
<tr>
<td>Bromoethylamine</td>
<td>8.5</td>
<td>1.5 × 10⁻⁴</td>
</tr>
<tr>
<td>Chloroethylamine</td>
<td>8.8</td>
<td>4.0 × 10⁻⁴</td>
</tr>
<tr>
<td>Ethanolamine</td>
<td>9.5</td>
<td>6.9 × 10⁻⁴</td>
</tr>
<tr>
<td>Butylamine</td>
<td>10.6</td>
<td>5.3 × 10⁻³</td>
</tr>
<tr>
<td>Methylamine</td>
<td>10.6</td>
<td>2.6 × 10⁻³</td>
</tr>
</tbody>
</table>

* kmax refers to the rate constant for the fully deprotonated, neutral species, whereas kobs is used elsewhere for the observed rate constant for an amine at a given pH.

### Table 4. Second-order rate constants for RA61 mutants at pH 7.5

<table>
<thead>
<tr>
<th>Variant</th>
<th>(kcat/KM)obs (M⁻¹ s⁻¹)</th>
<th>Mutant/WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type RA61</td>
<td>0.49</td>
<td>(1)</td>
</tr>
<tr>
<td>Tyr78Phe</td>
<td>1.56</td>
<td>3.2</td>
</tr>
<tr>
<td>Tyr78Ala</td>
<td>0.98</td>
<td>2.0</td>
</tr>
<tr>
<td>Ser87Ala</td>
<td>0.79</td>
<td>1.6</td>
</tr>
<tr>
<td>Ser87Gly</td>
<td>1.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Tyr78Phe/Ser87Ala</td>
<td>2.6</td>
<td>5.3</td>
</tr>
</tbody>
</table>
solvent may act to facilitate proton transfers. It is possible that mutations of Tyr78 and Ser87 increase activity slightly by allowing greater access to bulk solvent, although this model remains to be tested. Regardless, we can conclude that the designed water motif does not contribute to catalysis by RA61.

**Catalytic Contribution of Hydrophobic Binding Interactions.** A third and final element included in the design calculations was binding interactions with the hydrophobic surface of the substrate molecule. While specific binding interactions were not explicitly required in the design calculations, hydrophobic binding interactions with substrate are favored by van der Waals and solvation terms in the energy function (7). The design model of RA61 shows close proximity of the hydrophobic naphthyl rings with hydrophobic side chains (Fig. 2C). These types of interactions are expected to stabilize the transition state bound within the active site and would thus contribute to catalysis. In particular, binding interactions with the transition state would be expected to increase the enzymatic second-order rate constant examined here, which corresponds to the difference in energy between the transition state and the free enzyme and substrate in solution.

To estimate the catalytic contribution of binding interactions to the rate acceleration of RA61, we compared the reactions of the full substrate to those of an alternative substrate that lacks the hydrophobic naphthyl side chain, 4-hydroxy-4-methyl-2-pentanone (Table 5). The second-order rate constants for the nonenzymatic reactions differ by 4- to 5-fold from those of the full substrate, suggesting that intrinsic differences in reactivity are small. In contrast, the rate constants for enzyme-catalyzed reactions of the larger and smaller substrates differ by more than 105-fold. The differential reactivity between large and small substrates in RA61 compared to the amine-catalyzed reactions provides an estimate of around 500-fold for the contribution of the enzyme’s interaction with the naphthyl side chain. As the smaller substrate has two additional methyl groups that could make hydrophobic interactions in the active site and both substrates retain the hydroxybutanone scaffold, the full catalytic contribution from binding interactions is expected to be somewhat larger.

**Conclusions and Implications.** Computational enzyme design methods represent a promising tool for engineering new enzymes. Yet the success of these methods depends on their predictive power, so it remains crucial to experimentally evaluate computationally designed enzymes and to determine the strengths and limitations of computational design procedures.

The results presented here illustrate that the computationally designed retroaldolases provide around 105-fold rate acceleration beyond catalysis by the lysine side chain alone in solution. Studies of pH-rate profiles and a Brønsted analysis of nonenzymatic reactions suggest that about 10-fold of this rate acceleration of RA61 can arise through a combination of binding interactions and stabiliza-

<table>
<thead>
<tr>
<th>Catalyst</th>
<th>$k_{obs}$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_{obs}$ (M$^{-1}$ s$^{-1}$)</th>
<th>Ratio$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propargylamine</td>
<td>1.7 × 10$^{-5}$</td>
<td>3.3 × 10$^{-6}$</td>
<td>5</td>
</tr>
<tr>
<td>Trifluoroethylamine</td>
<td>7.0 × 10$^{-6}$</td>
<td>1.8 × 10$^{-6}$</td>
<td>4</td>
</tr>
<tr>
<td>RA61</td>
<td>0.49</td>
<td>2.3 × 10$^{-4}$</td>
<td>2.100</td>
</tr>
</tbody>
</table>

$^*$Ratio for the full substrate divided by $k_{obs}$ for the smaller substrate.

Thus, our results suggest that around 5,000-fold of the rate acceleration of RA61 can arise through a combination of binding the substrate into the active site and increasing the amount of reactive primary amine available to catalyze the reaction. The 40-fold remaining rate acceleration may arise through additional binding energy with the hydroxybutanone scaffold or the 4-hydroxy-4-methyl-2-pentanone substrate used for comparison herein. Other elements of catalysis not directly tested herein such as desolvation of the substrate and the reactive amine as well as entropic effects of localizing the substrate and the amine can also be linked to binding energies (30).

**Binding substrate and providing reactive chemical groups represent fundamental strategies of natural enzymes. However, natural enzymes offer highly specific interactions that provide optimal stabilization in the transition state of the reaction. Two observations suggest that the designed retroaldolases lack the precision and specificity in their interactions with substrate that are seen in natural enzymes. The observed tight inhibition by the naphthaldehyde product indicates that the enzymes favorably accommodate the naphthyl rings in a position that is translated toward the enzyme by 2–3 Å (SI Appendix). Furthermore, the substrate 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanol has a chiral center but we see no evidence of significant stereospecificity in RA61 (SI Appendix), suggesting that this enzyme has little preference for specific orientations of the naphthyl rings with respect to iminium bond.

**Materials and Methods**

**Materials.** The compounds 4-hydroxy-4-methyl-2-pentanone, 6-methoxy-2-naphthaldehyde, 2-aminoacetonitrile, 2,2,2-trifluoroethylamine, propargylamine, 2-bromoethylamine, 2-chloroethylamine, ethanolamine, methylamine, and butylamine were obtained from Sigma-Aldrich/Fluka at the highest available purity.

The fluorogenic substrate 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butano
tone was synthesized from acetone and 6-methoxy-2-naphthaldehyde as previously described (9) and further purified by silica gel chromatography in hexanes/ethyl acetate. Substrate quantitation was checked by comparison of $^1$H NMR signal to an internal standard.

Proteins were obtained using the previously described procedures (7), including an autoinduction expression protocol followed by purification.

Lassila et al. 2010

PNAS | March 16, 2010 | vol. 107 | no. 11 | 4941
by His-tag affinity to Ni-NTA resin (QIAGEN). Proteins were extensively dialyzed into a storage buffer of 25 mM sodium phosphate, pH 7.5, 100 mM NaCl or 50 mM sodium phosphate pH 7.5, 300 mM NaCl. Protein concentration was determined by absorbance at 280 nm in 6 M guanidine HCl. Mutants of RA61 were constructed using inverse PCR mutagenesis.

**Cosolvent and Buffer Conditions.** The low solubility of substrate 4-hydroxy-4-[(6-methoxy-2-naphthaldehyde)-2-butane in water dictates that cosolvents be included in kinetic assays. Initial investigations of the designed proteins used 2.7% acetonitrile in assays (7). However, under these conditions, we observed substantial departure from linearity in plots of nonenzymatic rate versus substrate concentration suggestive of a loss of substrate through insolubility. We therefore tested other cosolvent conditions. Several prior studies of antibody and peptide catalysis of the retroaldol reaction of the same substrate used 5% cosolvent (15). In 5% DMSO, substrate concentrations of ~500 μM could be reached before significant curvature in nonenzymatic rates was evident.

In testing conditions for pH-rate profiles, we found that many organic buffers either inhibited the enzymatic reaction or accelerated the nonenzymatic reactions. Enzymatic and nonenzymatic reactions were not affected by varying the concentration of phosphate, carbonate, and sulfate, so these buffers were used in all assays. Reaction pH values were checked by tests of mock reactions with a pH meter.

**Kinetic Assays.** Second-order rate constants were determined under conditions in which rates remained linear with enzyme and substrate concentration over at least 10-fold ranges. Standard assay conditions were 25 °C, 50 mM sodium phosphate, 300 mM NaCl, and 5% DMSO at pH 7.5. Stocks of the 4-hydroxy-4-[(6-methoxy-2-naphthaldehyde)-2-butane substrate were made in DMSO and stored at ~20 °C. Reaction kinetics were monitored by following fluorescence of the product 6-methoxy-2-naphthaldehyde on a FluoroLog 3 spectrofluorometer (HORIBA Jobin Yvon) with excitation at 330 nm and emission at 452 nm. Product concentration was calibrated by measurement of fluorescence of the product 6-methoxy-2-naphthaldehyde on a FluoroLog 3.

Reactions of 4-hydroxy-4-methyl-2-pentanone were monitored by proton NMR on a 600 MHz system with PRESAT water suppression. Reaction conditions were 25 °C, 50 mM sodium phosphate, pH 7.5, 300 mM NaCl, and 5-8% D2O for lock signal. Reactions were kept in a temperature-controlled water bath between data collection time points. The acetone peak was integrated and compared to an internal standard, sodium 2,2,3,3-tetradetero-3-trimethylsilylpropionate and a standard curve of acetone integration under identical buffer conditions and delay times. All reactions were first-order with respect to both enzyme or amine and substrate.

**pH-Rate Profiles.** Second-order rate constants were measured at different pH values using the buffer conditions described above (50 mM sodium carbonate, phosphate, or acetate, 300 mM NaCl, and 5% DMSO). Control experiments with enzyme incubated at pH extremes and subsequently assayed at pH 7.5 indicated that the protein was not irreversibly denatured during the course of the experiments. Data were fit to the following relationship, derived from the kinetic scheme shown in Fig. 3:

\[
\frac{k_{cat}}{K_M} = \frac{(k_{cat}/K_M)_{max}}{1 + [P]_{cat}^{p_{K_a} - pH}}.
\]

**Nonenzymatic, Amine-Catalyzed Reactions.** Second-order rate constants for amine-catalyzed reactions were measured under conditions in which rates remained linear with both substrate and amine concentration over at least 10-fold ranges. Spot checks were performed to ensure that no significant third-order effects were observed with buffer components. pH values of aqueous amine stock solutions were adjusted to match the reaction pH before addition to buffer solutions, and pH values of the final reaction conditions were tested in mock reactions. Amine-catalyzed rate constants were measured at several different pH values and average values for maximal rate constants were determined using the above equation. While the presence of DMSO cosolvent in reaction conditions has effects on both pH and pKₐ values, at 5% DMSO, these effects will be small (<0.5 units) and will not significantly affect the correlation (31–34).

**Acknowledgments.** We thank Jason Schwans, Stephen Lynch, Eric Althoff, and members of the Baker and Herschlag labs for helpful discussions. This work was supported by a National Institutes of Health postdoctoral fellowship to J.K.L. (F32 GM080865), by NIH Grant GM64798 to D.H., and by the DARPA Protein Design Processes program.

Supporting Text for:

Origins of Catalysis by Computationally Designed Retroaldolase Enzymes

Jonathan K. Lassila¹, David Baker², and Daniel Herschlag¹,³

¹Department of Biochemistry, Stanford University, Stanford, California 94305, U.S.A.
²Department of Biochemistry, University of Washington, Seattle, Washington 98195, U.S.A.
³To whom correspondence should be addressed: Department of Biochemistry, Beckman Center, B400, Stanford University, Stanford, CA 94305-5307. Phone 650-723-9442. Fax: 650-723-6783. Email: herschla@stanford.edu
Binding of naphthaldehyde product to enzyme

The 6-methoxy-2-naphthaldehyde product can return to the active site and form a covalent iminium species with lysine (Figure S1). Formation of this iminium state is not part of the reaction pathway, and it requires free lysine, which is regenerated after loss of acetone in steps 4 and 5.

The fluorescent signal of 6-methoxy-2-naphthaldehyde exhibits an exponential decrease when incubated with enzyme, and the rate and extent of loss depends on the concentration of enzyme (Figure S2). Fluorescence was monitored as described in the main text, with excitation at 330 nm and emission at 452 nm.

Figure S1. Formation of covalent iminium species with the product 6-methoxy-2-naphthaldehyde.

Figure S2. An exponential decay of fluorescence was observed when 6-methoxy-2-naphthaldehyde was incubated with enzyme. Standard assay conditions were used (see main text). A. Three variants at the concentrations indicated were incubated with 0.1 µM naphthaldehyde and monitored by fluorescence at 452 nm. Fluorescence background of enzyme and cuvette was not subtracted. B. Loss of fluorescence observed at the indicated concentrations of RA61, corrected for fluorescence background of enzymes and cuvettes. Initial naphthaldehyde concentration was 0.1 µM.
The fluorescence of bound 6-methoxy-2-naphthaldehyde is lower than that of the free species, but because of limitations in the solubility of some of the enzymes, saturating concentrations of enzyme were not attainable and we could not obtain a value for the fluorescence of the enzyme-bound naphthaldehyde species. For further analysis of RA61, we made the assumption that the fluorescence of the naphthaldehyde species is zero when it is completely bound to enzyme. The assumption is based on the observation that the Y78F/S87A double mutant could be concentrated slightly more and showed fluorescence loss of ~80% at achievable concentrations. While this assumption may not be correct and the enzyme-naphthaldehyde complex may show some fluorescence, the results and conclusions described here would not be significantly affected by a low level of fluorescence remaining for the enzyme-bound species, as discussed below.

The final observed fluorescence was converted to a concentration of free naphthaldehyde, using a standard curve. The final concentration of free naphthaldehyde was plotted against enzyme concentration, and the following binding equation was used to obtain a $K_d$ value:

$$[N]_f = \frac{K_d [N]_t}{[E]_t + K_d}$$

In this equation, $[N]_f$ represents the final free naphthaldehyde concentration, $[E]_t$ and $[N]_t$ are the total enzyme and naphthaldehyde concentrations, respectively. Figure S3 shows results for RA61, giving a $K_d$ value from curve fitting of 26 µM. Because the enzyme concentration could not be raised to obtain data points above the $K_d$, the value represents an approximate upper limit for $K_d$. If, for example, the naphthaldehyde product retained 20% of its fluorescent signal when fully bound to enzyme, the resulting $K_d$ would be about 17 µM. If 40% of the fluorescence was retained, the resulting $K_d$ would be about 10 µM.

![Figure S3. Binding of 6-methoxy-2-naphthaldehyde to RA61. Final apparent free naphthaldehyde concentration was plotted against enzyme concentration, and fit using a binding equation to give $K_d = 26$ µM. The initial concentration of naphthaldehyde was 0.1 µM.](image)
The loss of fluorescence of 6-methoxy-2-naphthaldehyde in the presence of enzyme has implications for following reaction progress curves. As the product builds up in the reaction, it increasingly binds to free enzyme. Binding of naphthaldehyde to enzyme can reduce the apparent fluorescence signal as well as the amount of free enzyme available to react. Because the rate of fluorescence production is slowed as more product is formed, the loss of fluorescence can result in burst-like behavior during reaction time courses.

This phenomenon is pronounced in RA34, for which a rapid and extensive loss of fluorescence is observed when enzyme is incubated with 6-methoxy-2-naphthaldehyde (Figure S2A). As shown in Figure S4, RA34 also displays burst-like behavior in fluorescence time courses, in which the degree of curvature increases with increasing enzyme concentration.

**Figure S4.** Reaction progress curves for RA34. Substrate (270 µM final concentration) was mixed with the indicated enzyme concentrations under standard conditions (see main text). The fluorescence signal was converted to free naphthaldehyde concentration with a standard curve, and the assumption was made (as discussed above) that the fluorescence of the enzyme-naphthaldehyde bound species is zero.

With RA61, only minor curvature is observed over several turnovers at low enzyme concentration (Figure S5), consistent with less pronounced rebinding of naphthaldehyde as seen in Figure S2. As with RA34, the degree of curvature is greater with higher enzyme concentration.
Figure S5. Reaction progress curves over multiple turnovers for RA61. Substrate (500 µM final concentration) was mixed with the indicated enzyme concentrations under standard conditions (see main text). The fluorescence signal was converted to free naphthaldehyde concentration with a standard curve, and the assumption was made (as discussed above) that the fluorescence of the enzyme-naphthaldehyde bound species is zero.

Assignment of pKₐ values to the catalytic lysine

As described in the main text, pH-rate profiles were determined for RA34, RA45, RA60, and RA61. Figure S6 shows the profiles for RA34, RA45, and RA60, while that of RA61 was included in the main text.

pH-rate profiles obtained under subsaturating conditions reflect ionizations of both the free enzyme and free substrate. Thus, although the pH-dependencies of multistep reactions can be complicated given the possibility of differential pH effects for different steps, under subsaturating conditions, these complications simplify to a consideration of the reactive forms of free enzyme and substrate and the rate-limiting transition state: \( v_0 = \frac{k_{cat}}{K_M} [E_{active}][S_{active}] \) (References S1-S3). Because the substrate 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone does not have titratable groups within the range of protein stability, observed pKₐ values are expected to arise from the free enzyme. In particular, because the deprotonated, uncharged form of the lysine side chain is required for reaction, the reaction rate has a first-order dependence on the concentration of deprotonated amine under subsaturating conditions.

Because the deprotonated form of lysine is also required in the model for naphthaldehyde fluorescence loss shown in Figure S1, we tested the simple model that the observed pKₐ in the retroaldol reaction with RA61 represents a simple titration of the enzyme by determining the pH-dependence of 6-methoxy-2-naphthaldehyde binding to RA61 (Figure S7). The resulting pKₐ value of 6.9 for binding is within error of the value of 6.8 obtained from the retroaldol reaction pH-rate profile. These results suggest most simply that the pKₐ observed in the pH-rate profile for RA61 arises from titration of the active site lysine.
**Figure S6.** pH-rate profiles for RA34, RA45, and RA60. The profile for RA61 is included in the main text. Fitting the curve (see Materials and Methods) gave pKₐ values of 7.5 for RA45 and 7.2 for RA60. Errors are estimated to be about ± 0.2 units.

**Figure S7.** pH-binding profile for RA61 with 6-methoxy-2-naphthaldehyde. Note that the Kₒ values determined are subject to the uncertainties described in the prior section, *Binding of naphthaldehyde product to the enzyme*. The fit gives a pKₐ value of 6.9, well within error of the pKₐ value obtained from the pH-rate profile for RA61 of 6.8. The error is estimated to be about ±0.5 units.
Estimation of enantioselectivity of RA61

The enantioselectivity of RA61 was estimated by following a reaction time course to about 40% of completion. If the enzyme reacts with significant selectivity for one enantiomer over the other, a measurable slowing in the time course should be evident as the favored enantiomer becomes fully depleted at 50% of overall completion.

The retroaldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone was followed at low substrate concentration to avoid large effects on fluorescence from the naphthaldehyde product rebinding. Aliquots were taken from the reaction mixture and assayed separately to confirm full activity of the enzyme after the reaction time course.

The reaction time course was compared with simulated curves calculated using different degrees of enantioselectivity. Simulated curves were produced using KinTek Explorer (Reference S4) with the following simple model for subsatrating kinetics and product rebinding:

\[ \begin{align*}
    E + S & \rightleftharpoons E + P \\
    & \uparrow \\
    & EP
\end{align*} \]

Enantiomeric preference was modeled by assuming equal concentrations of two reactants with 3-fold, 6-fold, and 10-fold differences in rate constant.

As shown in Figure S8A, when the measured value of \( K_d = 26 \mu M \) for product rebinding is used, the reaction progress curve fits best to a simulated progress curve with no enantioselectivity. The low degree of curvature suggests a preference of less than 3-fold for one enantiomer over another. If lower \( K_d \) values are used, greater curvature results and the reaction progress curve becomes even more inconsistent with the extent of curvature that would be seen if the enzyme were enantioselective.

To evaluate whether the degree of estimated enantioselectivity is dependent on the model for product rebinding and our measured \( K_d \) value, we modeled curves in which product rebinding is completely neglected (Figure S8B). When the loss of naphthaldehyde fluorescence in the presence of enzyme is not accounted for at all, the reaction progress curve still remains consistent with low enantioselectivity. The observed curvature in the reaction progress curve fits well to simulated curves with between 3-fold and 6-fold enantioselectivity, but does not fit well to a simulated curve with 10-fold enantioselectivity. This observation suggests that the enzyme does not exhibit enantioselectivity of more than 10-fold, regardless of the kinetic model used for product rebinding to the enzyme.
Figure S8. Observed reaction progress curve for RA61 (black points) compared to simulated curves for varying levels of preference for one enantiomer over another. Enzyme concentration was 10 µM and total substrate concentration was 0.5 µM. Free naphthaldehyde concentration was calculated using a standard curve for fluorescent signal and assuming, as described earlier, a zero fluorescence for the enzyme-product species. Observed data is indicated in black curves, while simulated curves are colored as labeled. A. Curves simulated using the simple model for subsaturating kinetics and product rebinding with a $K_d$ value of 26 µM. B. Curves simulated with a simple model for subsaturating kinetics and completely neglecting the effect of fluorescence loss in the presence of enzyme.
Catalytic contributions of shifted lysine pK\(_a\) calculated using different Brønsted values

Catalytic contributions of the shifted lysine pK\(_a\) were calculated with different Brønsted slope values (Table S1). For each enzyme, the difference between the measured pK\(_a\) and the pK\(_a\) of the lysine side chain, 10.6, was used to calculate the shift in maximal rate constant as follows:

\[ \Delta \log(\frac{k_{\text{cat}}/K_M}{K_M}_{\text{max}}) = \beta \Delta pK_a \]

Given the values for maximal rate constant for the lowered pK\(_a\) values, observed second order rate constants at pH 7.5 were then calculated using this expression to account for the quantity of deprotonated amine at pH 7.5:

\[ \left( \frac{k_{\text{cat}}/K_M}{K_M} \right)_{\text{obs}} = \left( \frac{k_{\text{cat}}/K_M}{K_M} \right)_{\text{max}} \frac{1 + 10^{pK_a - \text{pH}}}{10^{pK_a - \text{pH}}} \]

The ratio of observed rate constants at pH 7.5 for an amine of pK\(_a\) 10.6 and an amine of the measured enzymatic pK\(_a\) value was taken as the catalytic contribution of the pK\(_a\) shift.

Table S1. Values for the catalytic contribution of lysine pK\(_a\) shifts at pH 7.5, calculated using different Bronsted values.

<table>
<thead>
<tr>
<th>Retroaldolase</th>
<th>Fold Effect of pK(_a) shift using different Brønsted values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>RA45</td>
<td>151</td>
</tr>
<tr>
<td>RA60</td>
<td>175</td>
</tr>
<tr>
<td>RA61</td>
<td>182</td>
</tr>
</tbody>
</table>

References


