Supporting Online Material for

Robust design and optimization of retro-aldol enzymes

Eric A. Althoff, Ling Wang, Lin Jiang, Lars Giger, Jonathan K. Lassila, Zhizhi Wang, Matthew Smith, Sanjay Hari, Peter Kast, Donald Hilvert, Daniel Herschlag, David Baker

Part I: Materials and Methods

Part I: In vitro optimization of RA designs

Supplementary Tables

Supplementary Figures

1. Computational enzyme design methodology

The new designs described in this paper were produced using computational enzyme design methodology for the multi-step retro-aldol reaction shown in Figure S1 similar to that described previously (*1*). Given a schematic representation of a catalytic motif, the first step is to generate three-dimensional coordinates of the corresponding idealized active site. In our previous design, a mechanism involving bound water molecules yielded a number of active designs. In the motif, shown in Figure S2, a water molecule is still used and placed to coordinate the two alcohol groups on the carbinolamine, and then situated on the composite TS based on TS-carbinolamine superposition. In this paper, one of the two residues holding the water in place is a glutamate or aspartate, designed to facilitate water elimination and carbon-carbon bond-breaking as shown in Figure S2. An additional hydrogen bonding sidechain (Ser, Tyr, or His) is incorporated with the goal of hydrogen bonding and better positioning the water molecule.

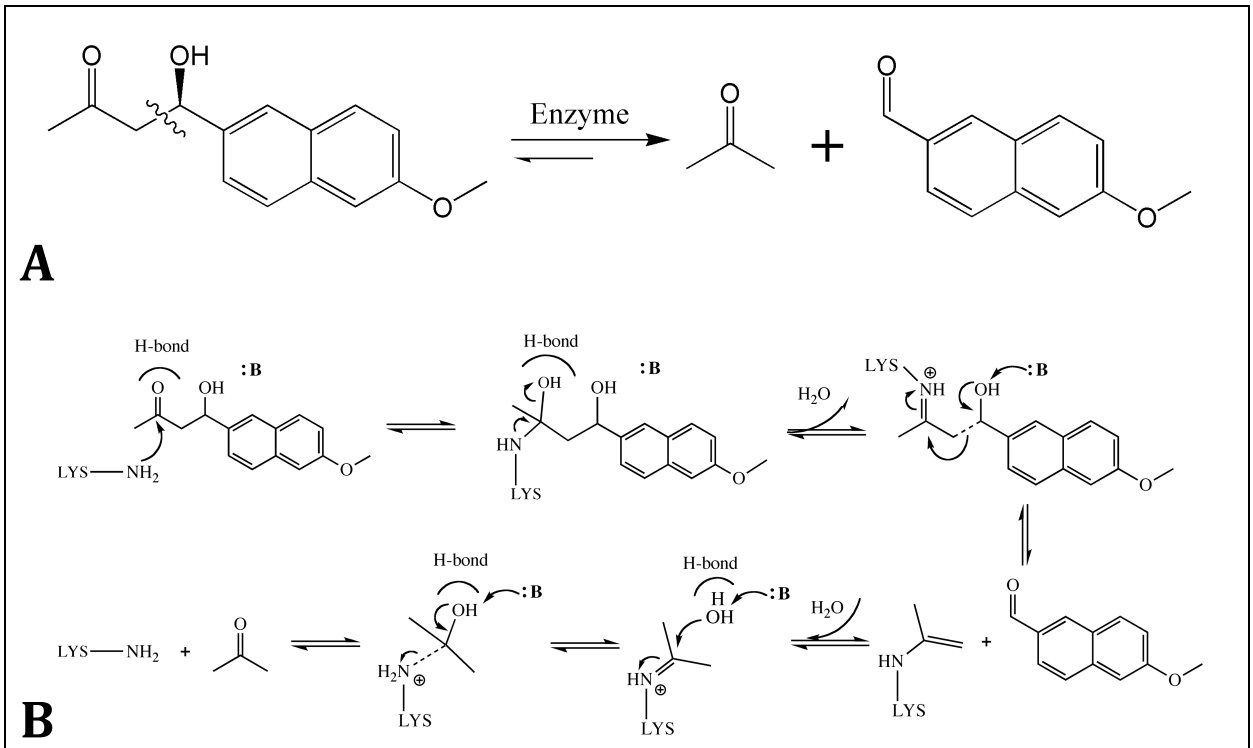


Figure S1. Retro-aldolase reaction: (**A**) Retro-aldol cleavage of methodol yields a fluorescent naphthaldehyde derivative (λ_{ex} = 330 nm, λ_{em} = 452 nm). (**B**) General description of retro-aldol reaction pathway through immonium or imine (Schiff-base lysine) formation and resulting enamine (*2*). The main catalytic steps are shown. The catalytic lysine needs to be deprotonated initially in order to attack the ketone to form the necessary covalent bonds. The general base serves to deprotonate the hydroxyl group. The hydrogen bond donor/acceptor is provided to further stabilize the carbinolamine hydroxyl group- and for this study the hydrogen bond network is replaced by a more specific sidechain combination depicted in Figure S2.

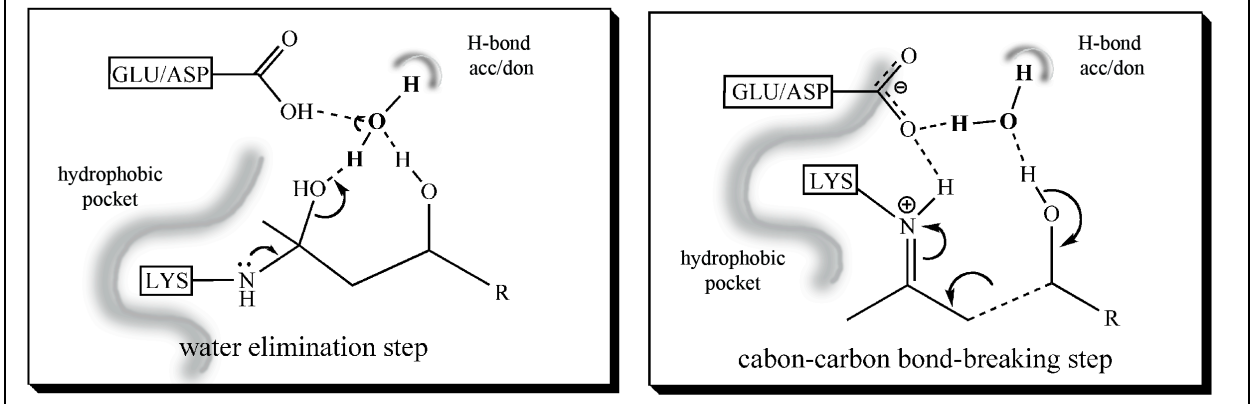


Figure S2: Catalytic motif using Glu/Asp as a general base. We used the minimal active site shown above for all designs. The catalytic glutamate or aspartate is designed to act as a general acid-base to facilitate the water elimination (left panel) and carbon-carbon bond-breaking transition state (right panel). The bridging water positioned by a hydrogen bonding residue (Ser, Tyr or His) is designed to allow proton shuttling on and off the substrate and product.

It is desirable to generate as large and diverse a set of three-dimensional realizations as possible for a given catalytic motif. We accomplish this by finer sampling along each of the degrees of freedom of the transition state-functional group complex, and by including an additional intermediate/transition state to represent the conformation change of the catalytic GLU/ASP residue when facilitating different key intermediates and transition states. In the calculations described in this paper, for each torsion angle we sample at +/- 10 degrees from the mean in addition to +/- 5 degrees as in our previous paper. Due to the significant conformational changes of catalytic glutamate or aspartate as a general base in facilitating several steps along the reaction pathway, we use the more consolidated composite substrate/intermediate/transition state by using the key intermediates and transition states of water elimination and Schiff-base lysine forming in addition to those of carbinolamine intermediate and carbon-carbon bond breaking step in our previous paper.

The RosettaMatch method (*3*) employs geometric hashing to identify positions in a set of input protein scaffolds, which support the construction of a specified constellation of catalytic residues. The initial active site configurations are often suboptimal due to slightly strained catalytic residue geometry and/or small clashes between the composite transition state and the protein backbone. To eliminate these imperfections and to ensure a near ideal geometry for the subsequent full site design calculations, the rigid body degrees of freedom of the composite transition state and the chi angles of the catalytic sidechains were subjected to quasi-Newton optimization using a force field consisting solely of repulsive interactions, a constraint function representing the relevant catalytic constraints and a sidechain torsional potential. We used RosettaDesign (*4*) to carry out full sequence optimization of the remaining (not included in the minimal catalytic site description) positions surrounding the docked composite TS model.

To optimize interactions (hydrogen bonding or packing) that are still missing at the end of the design process, we next carried out iterative refinement through small rigid body perturbations,

design of active site residues and structure relaxation. The resulting designs were filtered and evaluated based on the criteria described in the previous paper (*1*).

The filtering during design has been described in the supplementary materials in the previous paper. Simply the designs were filtered by each of important energy terms (LJ, LK, Hbond,sasa pack, and etc), that is, the design with a good score for each individual energy term was selected. And thus designs with the best total score and the best total binding energy were never selected for further refinement steps as the probability of satisfying all energetic constraints and having the best total scores is very low. After filtering, the designs were inspected using human intervention and comparison to naturally occurring active sites. For each design, we checked that every proposed mutation answered in a favorable way:
1. Does this mutation directly contribute to the binding or catalysis?
2. If the mutation is necessary, is the conformation in the design model stabilized by its neighbors? If no, is the conformation stable in the *apo* structure?
3. If the wild type sequence in the pocket has an alternative conformation, do the same as question 2.
Every unnecessary mutation was reverted back to the wild type sequence of the original scaffold. Based on the sequence profiles of scaffold homologues, additional mutations are introduced to stabilize the core of the active site (ie. important interactions, key catalytic residue). For each "polished" design, the model was evaluated by energy function again and the ligand docking was preformed. We discarded the design if the design can not pass the filter or fail the ligand docking.

47 new designs in 13 different scaffolds at 30 different Lysine positions were selected for experimental testing. 33 of the 47 designs had at least 10-fold activity over background as shown in the main text Table 1, and their amino acid sequences are shown in Part II, Table S4-S6. The active designs are based on 13 different scaffolds. Including the previous designs, among the total of 119 designs tested there are 65 active designs on 17 different crystal structures representing 14 different protein backbones (Figure S5). The relatively high success rate and the large number and structural diversity of the active designs suggest that we have sufficient understanding of this reaction and mechanism to consistently design active catalysts for this retro-aldol reaction.

Catalytic constraint energy calculation

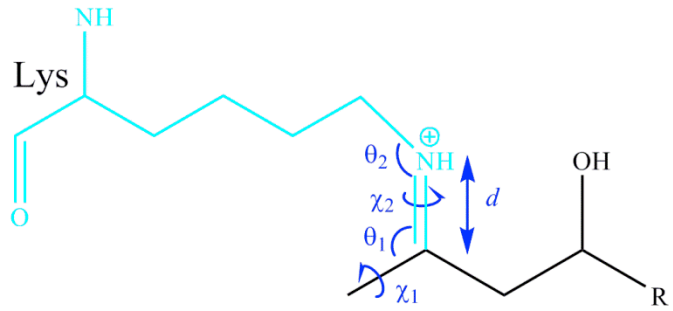
As the same as the previous paper (1), the constraint function consists of a flat bottom well with a width equal to the deviations with a harmonic penalty for violations greater than these values. For instance, the distance constraint energy is calculated by the following function:

$$E_{cst} = E(d) + E(\theta_1) + E(\chi_1) + E(\theta_2) + E(\chi_2)$$

$$E(d) = \begin{cases} k_d (d - d_0 - \Delta d)^2 & d > d_0 + \Delta d \\ 0 & d_0 - \Delta d \le d \le d_0 + \Delta d \\ k_d (d - d_0 + \Delta d)^2 & d < d_0 - \Delta d \end{cases}$$

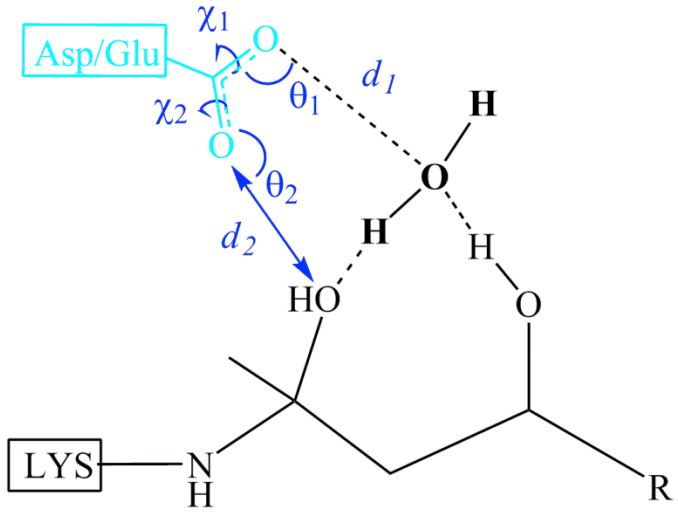
Based on the new catalytic motif using Glu/Asp as a general base, the catalytic geometries of Schiff base lysine and a general base of Glu/Asp are specified in the Figures S3 and S4. For each geometrical parameter, such as distance, angle and torsion, its ideal value (d_0), deviation (Δd) and harmonic penalty weight (k_d) are listed.

Figure S3. Geometric parameters for the imine-forming lysine



	ideal value	deviation	weight
d	1.4Å	+/-0.15Å	150
θ_1	125°	+/-15°	40
θ_2	120°	+/-15°	40
χ_1	180°	+/-25°	20
χ_2	0°,180°	+/-10°	10

Figure S4. Geometric parameters for the general base Glu/Asp



	ideal value	deviation	weight
d_1	2.8Å	+/-0.5Å	100
d_2	4.5Å	+/-1.5Å	100
θ_1	125°	+/-20°	20
θ_2	125°	+/-30°	10
χ_1	180°	+/-30°	10
χ_2	180°	+/-40°	5

Table S1. Calculated Energetic Contribution of all 43 Design Models.

The table is sorted by the rate enhancements (RE) of these retro-aldolase designs (Table 1), from low or no activity to high activity. The darkness of each row is dependent on the RE of the designed enzyme: 9 designs with no/week enzyme activity (RE of <10) are in a light gray color; 4 designs with RE of 10-100 are in a light-medium gray color; 22 designs with RE of 100-1000 are in a medium gray; 7 designs with RE of >1000 is in black. For each enzyme, the TS binding energy was calculated by summing up interactions cross the protein-TS interface, and each of important energy terms, such as LJ, LK and Hbond, was listed. The packing scores (SasaPack) and catalytic constraint energy were also listed. And for the interactions between the Schiff-base lysine and TS, only the catalytic constraint energy was calculated.

		Calculated TS Binding Energy (kcal/mol[1])					
Design	**Scaffold**	**LJ_{atr}[2]**	**LK_{sol}[3]**	**Hbond[4]**	**E_{bind}[5]**	**$Pack_{sasa}$[6]**	**E_{cst}[7]**
RA87	1PVX	-15.8	5.8	-2.8	-13	3.0	0.1
RA88	1F5J	-14.2	5.8	-3.4	-12	1.2	0.2
RA103	1DL3	-15.4	5.0	-1.6	-12	2.4	0.2
RA104	1DL3	-14.0	4.8	-1.8	-11	2.7	0.4
RA105	1H2J	-12.6	6.0	-2.8	-9	0.5	0.0
RA106	1H2J	-17.6	6.0	-2.8	-14	0.1	0.3
RA107	1GYE	-13.4	4.0	-0.8	-10	1.8	0.1
RA108	1E1A	-14.8	4.6	-1.4	-12	0.1	0.1
RA109	2IZJ	-12.4	5.2	-2.4	-10	2.3	0.3
RA101	1THF	-16.6	6.0	-3.2	-14	1.3	0.0
RA93	1LBL	-17.2	4.2	-2.4	-15	3.5	0.2
RA77	1M4W	-13.6	5.4	-2.2	-10	1.7	0.3
RA100	3B5V	-10.0	2.2	-0.4	-8	2.2	0.1
RA99	**1LBF**	**-14.4**	**5.6**	**-1.6**	**-10**	**2.2**	**0.2**
RA111	**1SJW**	**-12.0**	**4.4**	**-1.6**	**-9**	**1.0**	**0.4**
RA83	**1F5J**	**-17.6**	**7.0**	**-1.6**	**-12**	**-0.1**	**0.1**
RA97	**2C3Z**	**-15.0**	**5.0**	**-2.2**	**-12**	**-0.3**	**0.5**
RA84	**1M4W**	**-18.8**	**6.0**	**-2.2**	**-15**	**1.4**	**0.2**
RA81	**1M4W**	**-17.6**	**6.2**	**-3.2**	**-15**	**1.0**	**0.0**
RA94	**1LBL**	**-14.0**	**3.2**	**-1.6**	**-12**	**3.2**	**0.2**
RA80	**1PVX**	**-15.0**	**7.8**	**-2.2**	**-9**	**2.1**	**0.0**
RA71	**1DL3**	**-17.8**	**6.0**	**-3.6**	**-15**	**3.1**	**0.3**
RA96	**1LBL**	**-15.0**	**3.6**	**-2.2**	**-14**	**4.0**	**0.3**
RA85	**1PVX**	**-13.6**	**5.4**	**-2.2**	**-10**	**2.3**	**0.2**
RA79	**1F5J**	**-15.4**	**6.6**	**-3.8**	**-13**	**2.2**	**0.1**
RA92	**1A53**	**-15.4**	**3.2**	**-2.8**	**-15**	**2.2**	**0.4**

RA76	**1M4W**	**-15.0**	**5.0**	**-1.0**	**-11**	**2.1**	**1.6**
RA112	**1ILW**	**-15.0**	**4.4**	**-1.4**	**-12**	**1.7**	**0.1**
RA98	**1A53**	**-13.6**	**3.8**	**-0.2**	**-10**	**2.5**	**0.8**
RA113	**1V04**	**-17.0**	**4.0**	**-0.6**	**-14**	**2.1**	**0.0**
RA78	**1PVX**	**-14.0**	**5.6**	**-1.2**	**-10**	**2.4**	**0.7**
RA67	**1A53**	**-17.0**	**8.0**	**-1.6**	**-11**	**2.3**	**1.4**
RA102	**1THF**	**-14.6**	**5.0**	**-2.6**	**-12**	**1.0**	**0.0**
RA75	**1F5J**	**-16.6**	**6.0**	**-2.0**	**-13**	**0.5**	**0.0**
RA90	**1A53**	**-13.4**	**3.2**	**-2.6**	**-13**	**2.3**	**0.1**
RA72	**1M4W**	**-13.6**	**4.8**	**-0.8**	**-10**	**1.7**	**2.1**
RA82	**1M4W**	**-18.0**	**6.6**	**-1.8**	**-13**	**1.0**	**0.4**
RA89	**3B5L**	**-15.2**	**6.8**	**-1.6**	**-10**	**0.5**	**0.2**
RA110	**1OHO**	**-16.4**	**5.4**	**-1.0**	**-12**	**1.2**	**0.9**
RA95	**1LBL**	**-18.6**	**5.4**	**-1.2**	**-14**	**3.4**	**0.1**
RA86	**3B5L**	**-13.4**	**5.2**	**-2.4**	**-11**	**0.0**	**0.2**
RA91	**1LBL**	**-14.6**	**8.0**	**-0.6**	**-7**	**3.4**	**0.1**

[1.] TS binding energy is calculated by summing up interactions cross enzyme-TS interface.

[2.] attractive portion of LJ potential for non-bond interaction

[3.] LK solvation energy

[4.] hydrogen-bonding interactions

[5.] total TS binding energy

[6.] sasa pack score to measure the packing quality

[7.] catalytic constraint energy

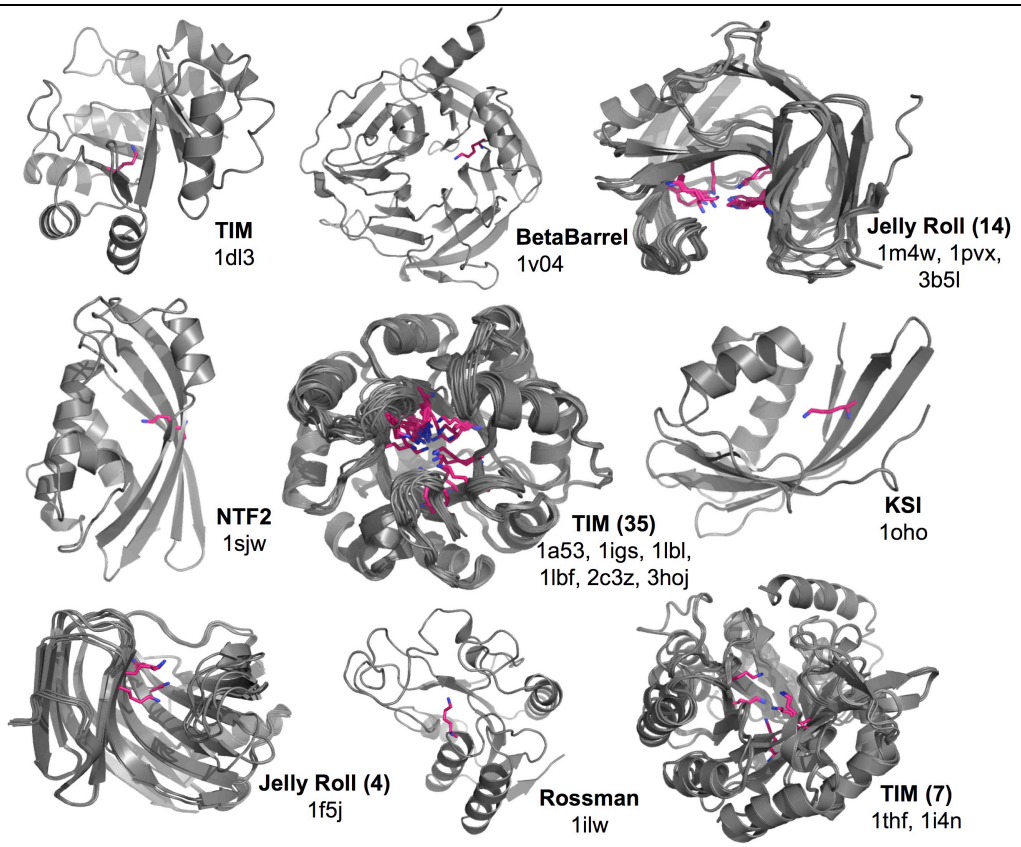


Figure S5: Scaffold and catalytic lysine positions used in active designed retro-aldolases including previous study (*1*). The number in parentheses indicate the number of active retro-aldolase designs in the indicated scaffold (if more than one). PDB scaffolds with identical sequences (1igs, 1lbl, 1lbf, 1i4n) or high structural homology (3b5l, 3hoj, 1pvx) are group grouped with their similar structures. Catalytic lysine positions are indicated in pink. 3hoj and 3b5l are crystal structures solved for other designed retro-aldolases (*1*); however, different lysine positions and substrate binding orientations within the pocket were used in making these active designs.

2. Experimental active site optimization

Saturation mutagenesis was performed at each active site position using Kunkel mutagenesis with degenerate oligonucleotides (*5*), and the resulting mutants were screened for increased catalytic activity using a cell lysate assay. In 96 well plates (or can be scaled up for larger volumes in culture tubes if signal was low for early rounds), 1 mL cultures of the mutants were grown at 37˚C in LB with 25 mg/mL Kan until OD ~ 0.6. Expression was then induced by the addition of IPTG to a concentration of 1.0 mM. The cells were then grown 4 h at 37˚C before being pelleted by centrifugation and having the supernatant media removed by aspiration. The pellets were frozen at -20˚C overnight. The cells were lysed with 5 cycles of dry ice/ethanol bath for 2 minutes and then 5 minutes in 10˚C water. The pellet was then resuspended in 250 μL 25 mM HEPES, 100 mM NaCl, pH 7.5, and incubated on ice for 15 minutes before centrifugation. 180 μL of supernatant was transferred to a 96 Well Black Flat Bottom Polystyrene Non Binding Surface Microplates (Corning, Lowell, MA). 5 mL of 10.0 mM 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (*6*) in CH_3CN was than added and the enzyme activity was monitored by fluorescence in a Spectramax M5^e (Molecular Devices, Sunnyvale, CA) in using λ_{ex} of 330 nm (9 nm bandwidth) and λ_{em} of 452 nm (15 nm badwidth) with an additional filter at 435 nm for increased noise reduction. Quartz cuvette measurements were used to verify the plate measurements. The reactions were controlled for evaporation and were generally stable up to 4-5 hours with minimal evaporation- though most kinetic measurements were done within the first hour.

For most positions that were examined with mutagenesis the original Rosetta-designed side-chain gives the highest activity. So in general, Rosetta appears to find the best solution; however, several positions could be further optimized by point mutation, demonstrating how to improve future designs.

2a. Optimization of the retro-aldolases RA45 and RA60

For positions around the active site, Kunkel mutagenesis (*5*) using degenerate oligos were used at each position to make a library. Each position library was sequenced and colonies encoding as many different amino acids as could be identified were assayed for a difference in catalytic activity. The best "hits" were combined together in multiple combinations to look for additive effects and then additional rounds of single-position Kunkel degenerate saturation mutagenesis were performed. For RA60, three rounds were done this way. For RA45, several additional rounds were done including a round of mutagenesis at "consensus" positions where the template scaffold sequence varied from the consensus sequence when using BLAST to compare it to other similar sequences (*7*). These consensus mutations increased protein expression levels and stability. Several rounds of active site mutagenesis were completed after the consensus mutations were added. After the first or second rounds of mutagenesis, the mutants were not sequenced until after being assayed and only mutants that had increases in reaction rate were sequenced.

2b. Optimization of the retro-aldolases RA95 and RA110

RA95 and RA110 were mutagenized very similarly to the others; however, only mutations resulting in increased reaction rate were sequenced. Although the identity of deleterious mutations is not known, larger libraries with redundancy were screened. Additionally, important mutations were discovered early in the screening and used immediately for screening the entire active site M180F and R182I/L/S for RA95 and T114R for RA110. RA95 also underwent a consensus mutation round of mutagenesis like RA34 and RA45. See Part II, Tables S9-S12.

3. Conditions for enzyme expression, purification, and catalytic activity assays

Genes were synthesized by GenScript (Piscataway, NJ) and cloned into pET29b+ with the resulting protein expressed having only an N-terminal formyl-methionine but a C-terminal Gly-Ser (linker) and Leu-Glu (cloning site, XhoI) followed by a six His tag. Enzymes are expressed in BL21-star cells using auto-induction media 8 hours at 37˚C and 24 hours at 18˚C. The cells were sonicated (Bug-Buster-like methods left a very high aldolase background signal in negative control experiments) and purified over Qiagen Ni-NTA resin. After elution the proteins were dialyzed at least three times into 100-fold buffer 25 mM HEPES, 100 mM NaCl at pH 7.5 or by desalting column to remove the imidazole.

For the catalytic assays, fluorescence of the product was measured with an excitation wavelength of 330 nm and an emission wavelength of 452 nm in either 96 Well Black Flat Bottom Polystyrene Non Binding Surface Microplates (Corning, Lowell, MA) or quartz cuvettes and compared to known product concentration curves for quantitation. The measurements were taken every 20-60 seconds. (More frequent measurements led to increased noise over the course of long assay runs presumably due to photo-bleaching or side reactions of the photoactive molecules.)

4. Multiple-turnover kinetics for RA34.6, RA95, and RA110

4a. Fluoresence assay

Because of complications in the fluorescence assay caused by tight binding of the fluorescent naphthaldehyde product to the enzyme (*8*), rate constants reported in this paper were determined from initial rates measured before the first turnover. As described in this section, multiple-turnover rates do not depart significantly from single-turnover rates, suggesting that the enzymes are not limited by a buildup of the enzyme-acetone species, and making it possible to use initial rates before the first turnover to obtain k_{cat} and K_M values.

As previously described (*8*), the binding of the naphthaldehyde product to the enzyme results in a loss of its fluorescence, leading to effects in the fluorescence-based activity assay that can appear similar to a kinetic burst caused by a slow reaction step. For RA95.4 and RA110.4, multiple-turnover kinetics could be monitored in a straightforward manner, provided that low enzyme

concentrations were used to minimize naphthaldehyde binding to enzyme (Figure S6). These time courses show some curvature, consistent with naphthaldehyde product binding to enzyme, but they indicate that the reaction progress continues over multiple turnovers without an evident burst. Multiple-turnover rate constants were within error of single-turnover rate constants obtained with higher enzyme concentrations.

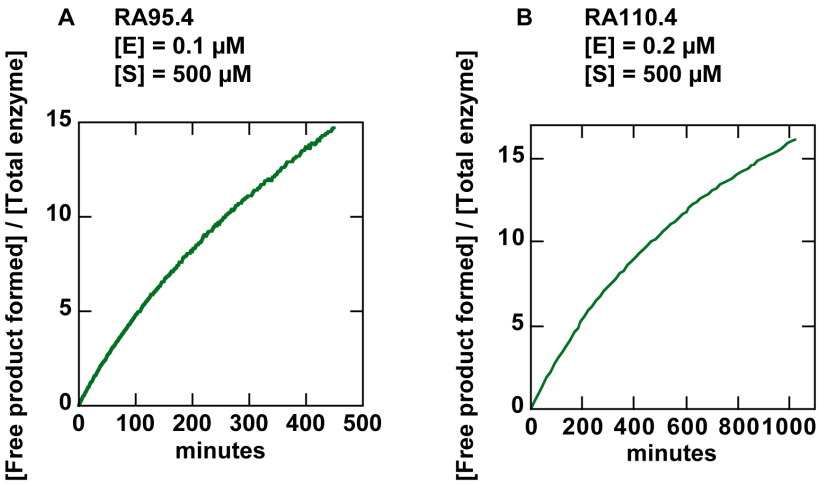


Figure S6: Multiple turnover kinetics of RA95.4 and RA110.4. Reaction progress continues over multiple turnovers in (A) RA95.4 and (B) RA110.4, suggesting that the multiple-turnover rates are not significantly limited by a buildup of the enzyme-acetone species. The curvature observed is consistent with loss of free naphthaldehyde product fluorescence through rebinding to the enzyme (*8*). Reaction conditions were: 25 °C, 50 mM sodium phosphate, pH 7.5; 300 mM NaCl; 5% DMSO.

The RA34.6 retro-aldolase shows tighter binding of the naphthaldehyde product to enzyme, as was previously seen in its parent retro-aldolase RA34 (*8*). In RA34.6, lowering the enzyme concentration was not sufficient to eliminate the burst of fluorescence attributed to naphthaldehyde binding to enzyme, so additional measures were required. A variant of RA34.6 was used in which six of twelve surface lysine residues (not the catalytic lysine) were mutated, in order to reduce loss of naphthaldehyde fluorescence due to binding of these surface residues. In addition, a discontinuous assay was used in which the enzyme was treated with guanidinium, presumably leading to denaturation of the protein and release of the bound naphthaldehyde product. The additional product observed suggested that upon addition of guanidinium HCl there was at least some weakening of bound and fluorescently quenched product allowing its release back into solution. The similar fluorescent increases observed upon addition of 3.3 or 5 M guanidinium suggests that dissociation is complete under these conditions. When these measures were used, reaction progress was seen to continue linearly through at least two turnovers of enzyme, suggesting that the acetone loss steps do not significantly limit reaction rates over multiple turnovers in the RA34.6 active site (Figure S7).

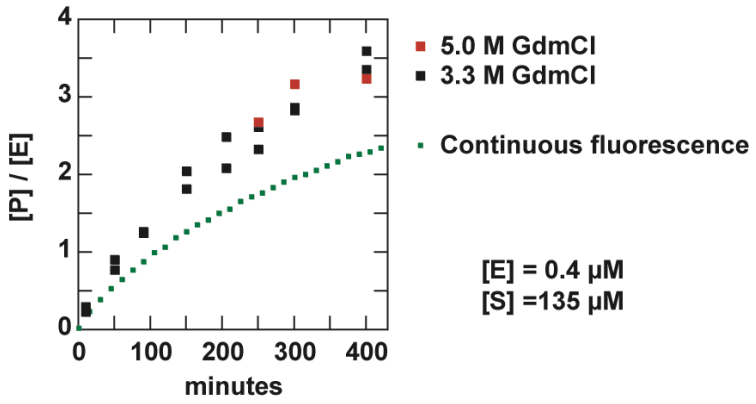


Figure S7: Multiple turnover kinetics with a RA34.6 variant. In a variant of RA34.6 in which six of twelve surface lysines were mutated, reaction progress continued over multiple turnovers without significant reductions in rate, when guanidinium HCl was added to release bound naphthaldehyde product. Reaction conditions were: 25 °C, 25 mM HEPES, 100 mM NaCl, 2.7% acetonitrile, pH 7.5.

4b. Mass spectrometry assay.

A comparison between the fluorescence assay, which detects the evolution of the product, and a mass spectrometry assay set up to detect the rate of substrate disappearance. In a direct comparison (Figure S8), the mass spectrometer-based assay gives a rate (k_{cat}) that is almost two-fold higher than that of the fluorescence-based assay.

Table S1: Kinetic parameters for RA34.6

RA34.6 kinetics	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
Mass spectrometer assay	7.60×10^{-4}	57 ± 6	13
Fluorescence assay	4.25×10^{-4}	40 ± 7	10

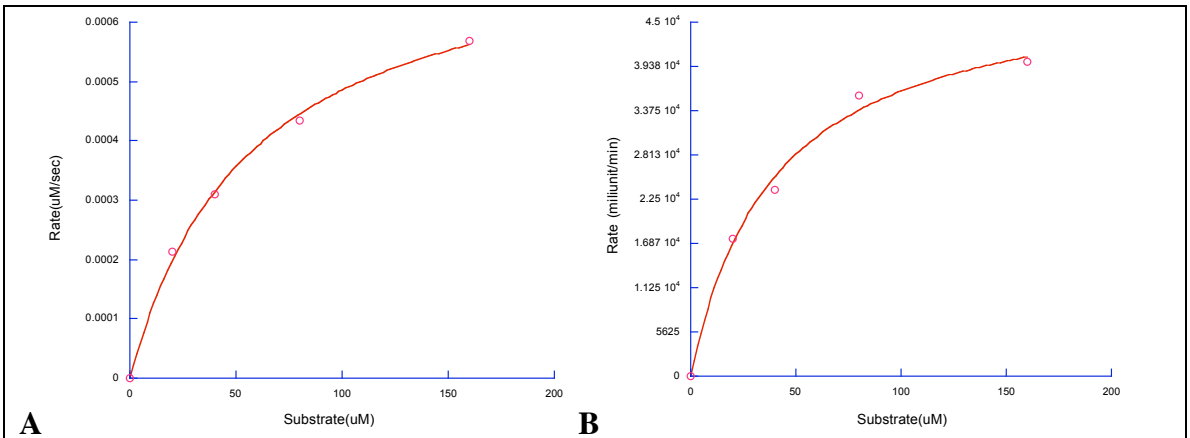


Figure S8: Comparison of RA34.6 catalytic rate by mass spectrometry and fluorescence-based assays. 5 μM of RA34.6 was incubated with 20 μM, 40 μM, 80 μM, and 160 μM of substrate for 2, 5, and 10 min. **A.** Then the reaction mixture was diluted into 75% acetonitrile to stop the reaction and filtered to remove the protein from the solution. The mixture was separated on a C18 column and analyzed with LC-MS-MS mass spectrometry. The concentration of the substrate was determined by the mass spectrometry and fitted to a standard curve (0.25-8 mM) of the substrate under the same condition. The rate of substrate concentration decrease for each substrate concentration was plotted against the substrate concentration to make the Michaelis-Menten plot. Substrate turnover was below 10% for each point, and the rate was still in the linear region when under 10 minutes. **B.** Simultaneously, the identical assay was run using the fluorescence assay used elsewhere for measuring the enzymatic rate of the retro-aldol reaction.

5. pH-rate profile for RA34.6

The pH-rate profile under subsaturating conditions is shown for RA34.6 in Figure S9. The reaction conditions were those previously discussed (*8*) and used for RA34 (50 mM phosphate, acetate, or carbonate, 300 mM NaCl, and 5% DMSO, pH 7.5), and the data were collected and fit to a single pK_a value as previously described (*8*). Note that RA34.6 showed about 4-fold higher activity in the presence of acetonitrile, the cosolvent used to obtain data reported in the main text. The other enzymes tested did not show this effect of cosolvent on rate constants.

The pH-rate profile for RA34 was included in prior work (*8*). Due to low protein solubility, low activity, and high nonenzymatic reactivity at high pH, enzymatic data was not obtained above pH 8.0 for RA34. The data that could be obtained for RA34 gave the best fit to a pK_a value of 7.2, and the pK_a was reported as a lower limit of 7.0. The best-fit values of 7.2 for RA34 and 7.3 for RA34.6 are within error, with error estimated to be ± 0.2 for RA34.6. If these two values are taken as the pK_a shift that occurs as a result of sequence changes from RA34 to RA34.6, the resulting estimated effect on rate constant, calculated as previously described using a value of β= 0.54 (*8*), is about 3-fold. If the pK_a value for RA34 were 7.5, the estimated effect of the pK_a shift would be a roughly 3-fold *reduction* in rate constant in RA34.6. If the pK_a value for RA34 were 8.0, the estimated effect of the pK_a shift would be a 3-fold increase in rate constant in RA34.6.

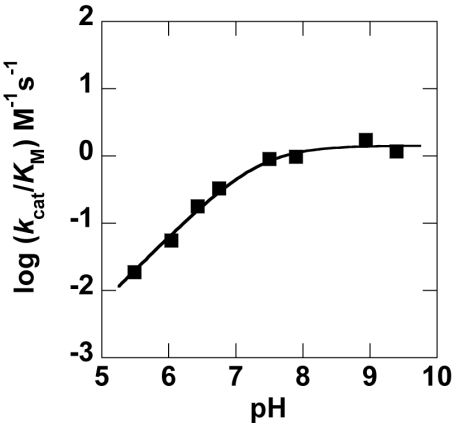
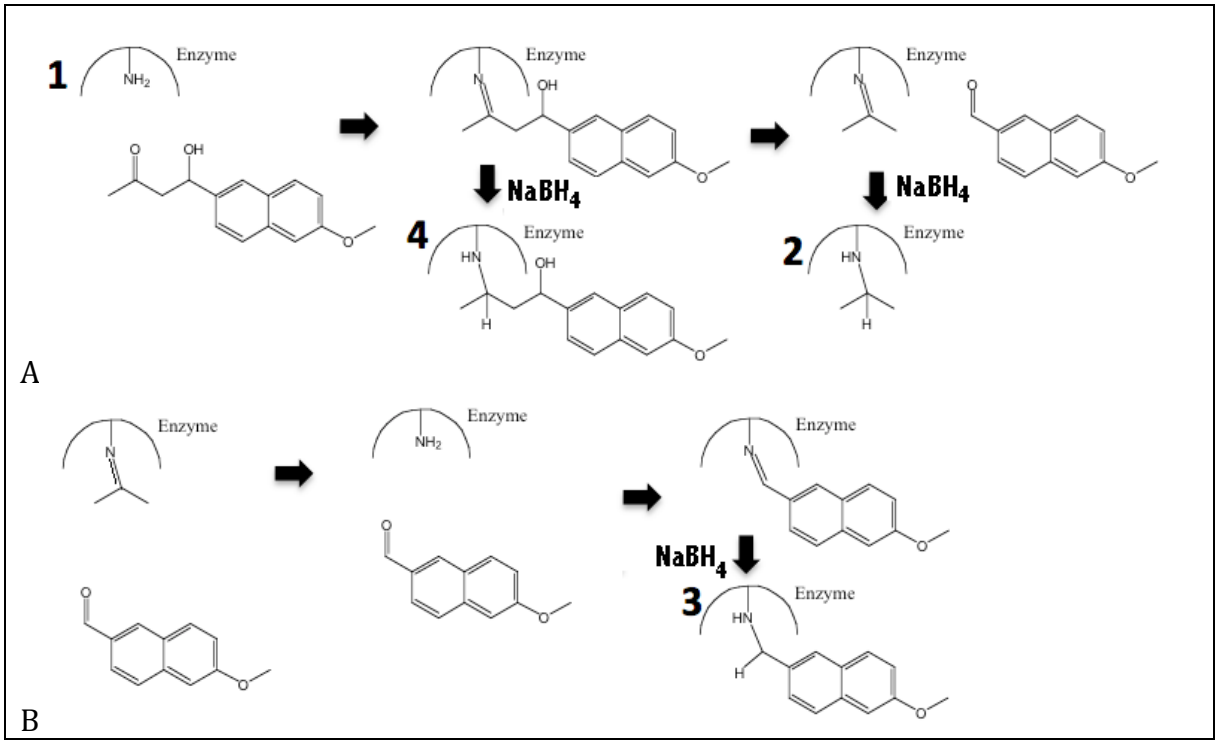


Figure S9: pH-rate profile for RA34.6. Fitting the curve to a single pK_a (4) gave a pK_a value of 7.3. Reaction conditions were 50 mM phosphate, acetate, or carbonate, 300 mM NaCl, and 5% DMSO, pH 7.5, conditions that were previously used in mechanistic studies of the retro-aldolases in order to reduce the problem of substrate insolubility (*8*).

6. Borohydride trapping of reaction intermediates

Covalent enzyme-substrate with substrate and products accumulate during the retro-aldol reaction. 10 μM RA34.6, RA95.4, or RA110.4 were incubated with 200 μM substrate in 25 mM HEPES (pH7.5), 100 μM NaCl, 5% acetonitrile and after the indicated time, $NaBH_4$ was added to a final concentration of 8mM for 1 minute to trap reaction intermediates bound to the active site lysine. Samples were then injected into a small C4 column (BioBasic4 30x1), washed with H2O+0.1% formic acid for 1 minute, and eluted with a 5 min acetonitrile gradient (up to 5:95 H_20:acetonitrile in 0.1% formic acid) into a Fourier transform mass spectrometer. The protein peak elutes from the column in about 2.5 minutes; the panels shown in Figure S10 are the deconvoluted spectra of this peak. The major peaks in the spectra are labeled based on their masses: 1 is the mass of the unmodified protein, 2 is the protein mass plus 42 Da, which corresponds to an acetone adduct on a Lys side chain or acetylation, 3 is the protein mass plus +170 Da, the mass of the product 6-methoxy-2-naphthaldehyde bound to a lysine side chain, and 4 is the protein mass +226 Da, which corresponds to the substrate bound to a lysine. No covalent adducts were observed in an RA34.6 catalytic lysine knockout. Similar distributions of species were observed when the borohydride reduction incubation period was increased from 1 to 10 minutes, and when the borohydride concentration was reduced to 4 mM.

Rough quantitation of the acetone adduct peak for RA34.6 (Figure S11) is consistent with a reaction rate 2-fold greater than expected based on the fluorescence assay, but close to that expected from the mass spectrometry data in Fig S8. The fluorescence data likely underestimates the rate due to quenching through product association with the protein (8).



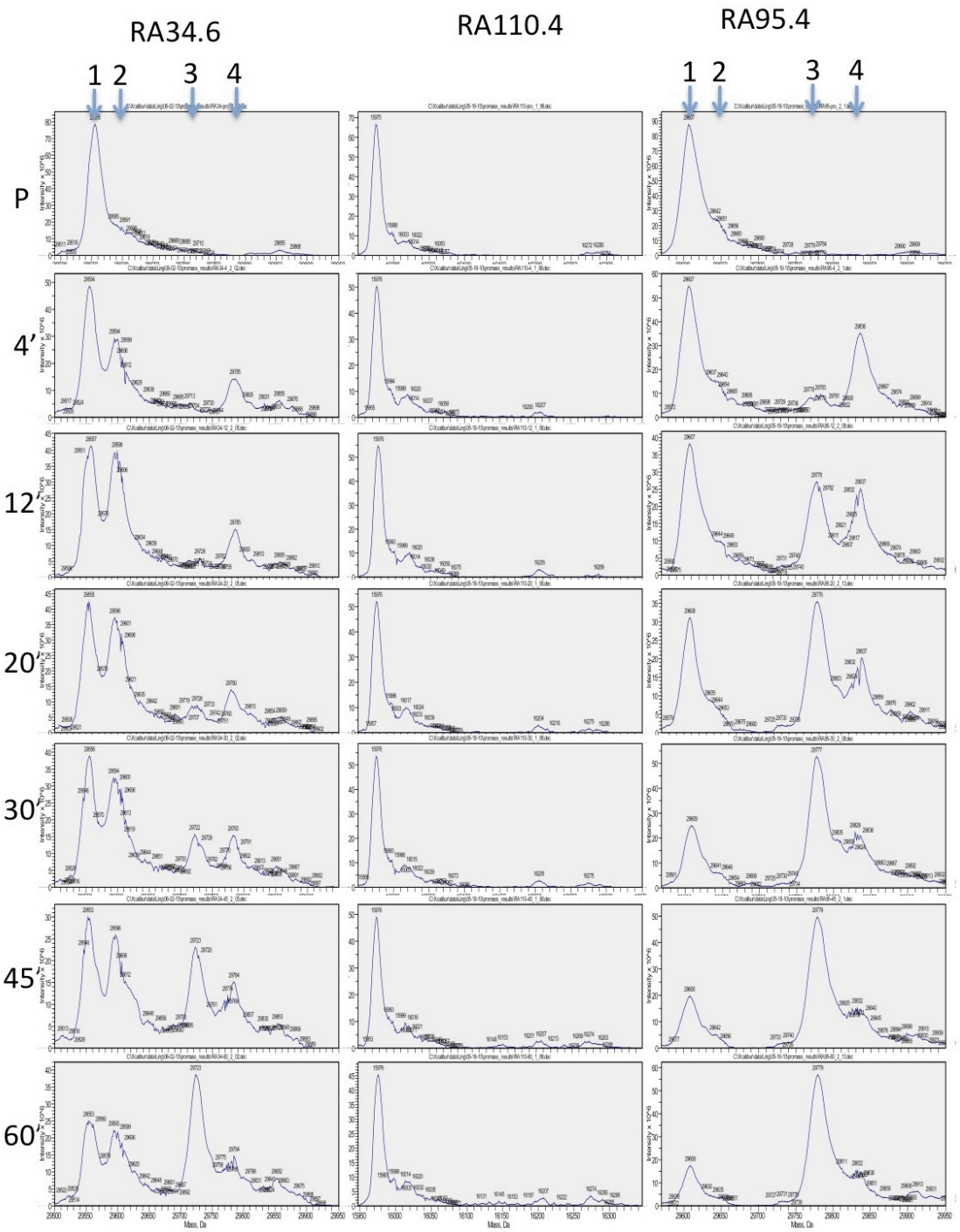


Figure S10: Formation of covalent enzyme-ligand adducts. The first two panels (previous page) depict how each intermediate or product is trapped on the active site lysine by $NaBH_4$, which reduces the imine to a non-hydrolysable amine, which can be observed in a mass-spec assay. Fourier transform mass spectrometry spectrum for RA34.6, RA95.4, and RA110.4 at different reaction time points. Protein is first mixed with substrate for the stated time (in minutes), then $NaBH_4$ was added at each time point before injection into the mass spectrometer. Peak 1 is the protein peak. Peak 2 is protein MW+42 Da, which correspond to an acetone adduct on a Lys side chain or acetylation. Peak 3 is protein MW+170 Da, which corresponds to the product 6-methoxy-2-naphthaldehyde bound to a lys side chain. Peak 4 is protein MW+226 Da, which corresponds to the substrate on a Lys side chain.

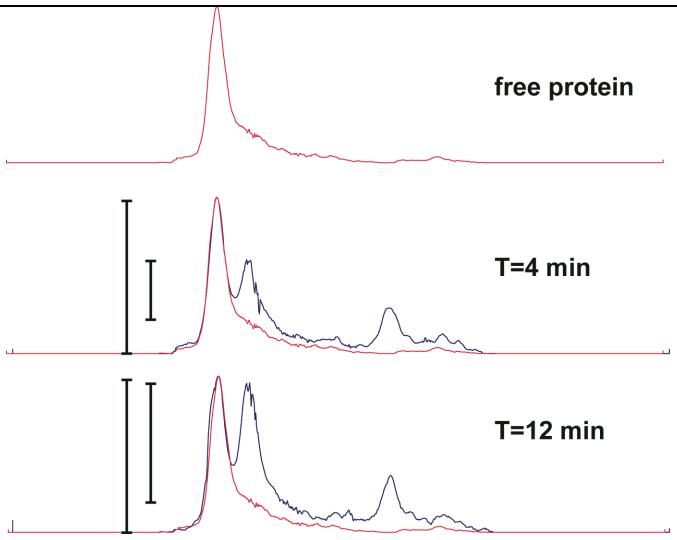


Figure S11: Overlap of free protein peak and enzyme-acetone species. The shoulder seen in the protein peak in the absence of added substrate, acetone, or product was consistently observed with multiple protein preps and assay conditions. This peak shape for the free protein overlaps with the enzyme-acetone species peak, as shown in the overlay of the free protein peak (red) with the observed traces at 4 minutes and 12 minutes (blue). When overlap in peak shape between the free protein peak and the enzyme-acetone peak is considered, the relative proportion of enzyme-acetone species does not appear to be greater than 40%. This suggests no more than a two-fold reduction in rate due to buildup-of enzyme-acetone species, consistent with the lack of observable burst kinetics.