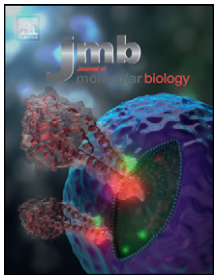

# Exploration of Alternate Catalytic Mechanisms and Optimization Strategies for Retroaldolase Design

**Sinisa Bjelic[1,†], Yakov Kipnis[1,†], Ling Wang[1,†], Zbigniew Pianowski[2], Sergey Vorobiev[3], Min Su[3], Jayaraman Seetharaman[3], Rong Xiao[4], Gregory Kornhaber[4,5,6], John F. Hunt[3], Liang Tong[3], Donald Hilvert[2] and David Baker[1,7]**

*1 - **Department of Biochemistry,** University of Washington, Seattle, WA 98195, USA*
*2 - **Laboratory of Organic Chemistry,** ETH Zurich, 8093 Zurich, Switzerland*
*3 - **Department of Biological Sciences,** Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA*
*4 - **Center for Advanced Biotechnology and Medicine,** Rutgers, The State University of New Jersey, NJ 08854, USA*
*5 - **Robert Wood Johnson Medical School,** University of Medicine and Dentistry of New Jersey, NJ 08854, USA*
*6 - **Northeast Structural Genomics Consortium,** 679 Hoes Lane, Piscataway, NJ 08854, USA*
*7 - **Howard Hughes Medical Institute,** University of Washington, Seattle, WA 98195, USA*

***Correspondence to David Baker:** Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.*
*dabaker@u.washington.edu*
http://dx.doi.org/10.1016/j.jmb.2013.10.012
***Edited by G. Schulz***

## Abstract

Designed retroaldolases have utilized a nucleophilic lysine to promote carbon–carbon bond cleavage of β-hydroxy-ketones via a covalent Schiff base intermediate. Previous computational designs have incorporated a water molecule to facilitate formation and breakdown of the carbinolamine intermediate to give the Schiff base and to function as a general acid/base. Here we investigate an alternative active-site design in which the catalytic water molecule was replaced by the side chain of a glutamic acid. Five out of seven designs expressed solubly and exhibited catalytic efficiencies similar to previously designed retroaldolases for the conversion of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone to 6-methoxy-2-naphthaldehyde and acetone. After one round of site-directed saturation mutagenesis, improved variants of the two best designs, RA114 and RA117, exhibited among the highest $k_{cat}$ ($>10^{-3}$ $s^{-1}$) and $k_{cat}/K_M$ (11–25 $M^{-1}$ $s^{-1}$) values observed for retroaldolase designs prior to comprehensive directed evolution. In both cases, the $>10^5$-fold rate accelerations that were achieved are within 1–3 orders of magnitude of the rate enhancements reported for the best catalysts for related reactions, including catalytic antibodies ($k_{cat}/k_{uncat} = 10^6$ to $10^8$) and an extensively evolved computational design ($k_{cat}/k_{uncat} > 10^7$). The catalytic sites, revealed by X-ray structures of optimized versions of the two active designs, are in close agreement with the design models except for the catalytic lysine in RA114. We further improved the variants by computational remodeling of the loops and yeast display selection for reactivity of the catalytic lysine with a diketone probe, obtaining an additional order of magnitude enhancement in activity with both approaches.


## Introduction

Computational enzyme design has been used to create catalysts for a wide range of reactions [1–5], but considerable challenges remain in both increasing the fraction of designs that have activity and increasing the overall level of activity. The *in silico* design process proceeds in three stages: (i) construction of an idealized active-site description, or theozyme; (ii) placement of the theozyme in a suitable protein scaffold; and (iii) optimization of the surrounding sequence for transition-state binding. Choosing an appropriate theozyme is critical as the catalytic mechanism and the chemical composition of the catalytic residues and their interactions with the transition state must be decided upon. Each

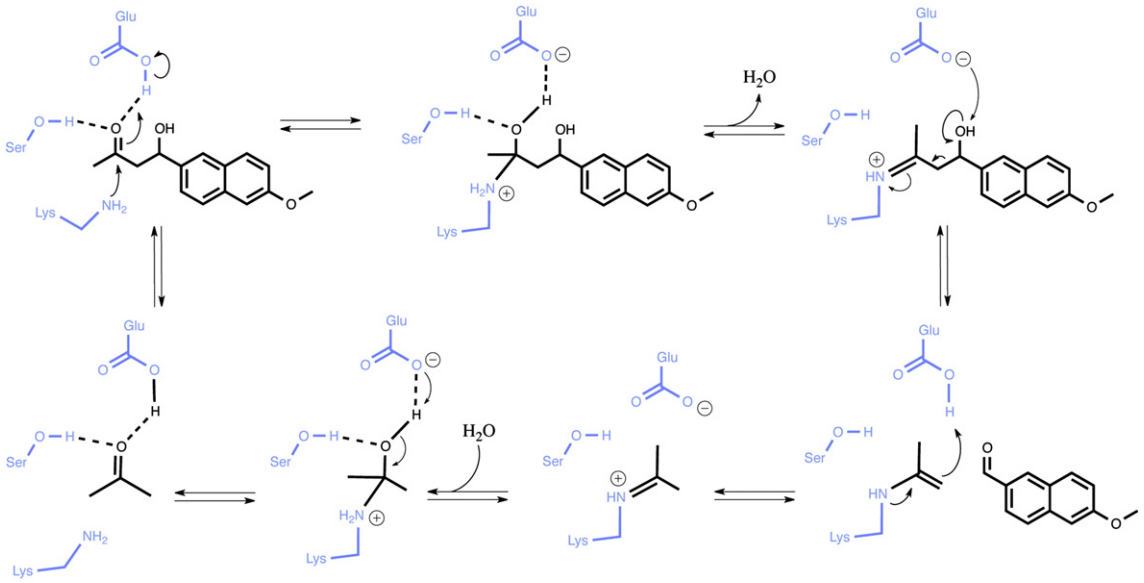


**Fig. 1.** Catalysis of the retroaldol reaction. A reactive lysine accelerates the cleavage of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (methodol) by forming protonated Schiff base intermediates. 6-Methoxy-2-naphthaldehyde and acetone are the final products. A glutamic acid and a serine are examples of potential catalytic residues that could activate the substrate and stabilize transition states and intermediates along the multistep reaction pathway through hydrogen-bonding interactions and acid/base chemistry.

different theozyme represents a hypothesis about how catalysis can be achieved, which can be evaluated using quantum mechanical calculations [6] and ultimately by the experimentally observed activity of the designed enzymes.

Multistep retroaldol reactions, which are subject to amine catalysis, were among the first transformations tackled by computational design [2,7]. Catalysis is initiated by attack of a reactive lysine on the carbonyl group of the β-hydroxy-ketone substrate to form a tetrahedral carbinolamine intermediate that subsequently breaks down to give a protonated Schiff base. The latter serves as an electron sink, facilitating cleavage of the adjacent carbon–carbon bond to generate an aldehyde and an enamine. Protonation and hydrolysis of the enamine leads finally to release of acetone and regeneration of the enzyme. This mechanism, which is exploited by natural type I aldolases [8], has been successfully mimicked by lysine-rich helical peptides [9–11] and proteins [12], as well as catalytic antibodies selected against 1,3-diketones [13,14] and β-keto sulfones [15].

The first computationally designed retroaldolases were obtained by explicitly modeling the structure of the carbinolamine intermediate and flanking transition states, the most sterically demanding species along the reaction coordinate. These designs also included an ordered water molecule, bound by two hydrogen-bonding side chains, to promote carbinol-amine formation and breakdown. It was envisaged that the water would additionally assist proton transfer from the β-alcohol in the cleavage step. The designed catalysts exhibited significant retroaldolase activity, with rate accelerations of up to 4 orders of magnitude over background [2,7]. Detailed mutagenesis and structural studies of representative designs have confirmed the importance of the reactive lysine, but a significant catalytic role for the explicit water has not been observed [16]. Although naturally occurring class I aldolases such as D-2-deoxyribose-5-phosphate aldolase often use a water molecule for acid/base catalysis, this water is typically oriented and activated by an extensive network of polar side chains that is difficult to emulate with current computational protein design methodologies [17,18]. We speculated that, in the absence of such a network, amino acid side chains interacting directly with bound ligands at the designed active sites might provide better control over the reaction coordinate than a loosely bound water molecule and thus afford higher activity.

Here we describe the results of design calculations in which the explicit water in the earlier theozymes is replaced by the carboxylic acid side chain of glutamic or aspartic acid, to function as a general acid/base, plus a serine or threonine residue, to provide additional hydrogen-bonding interactions. We also describe approaches to increase the activity of the designed

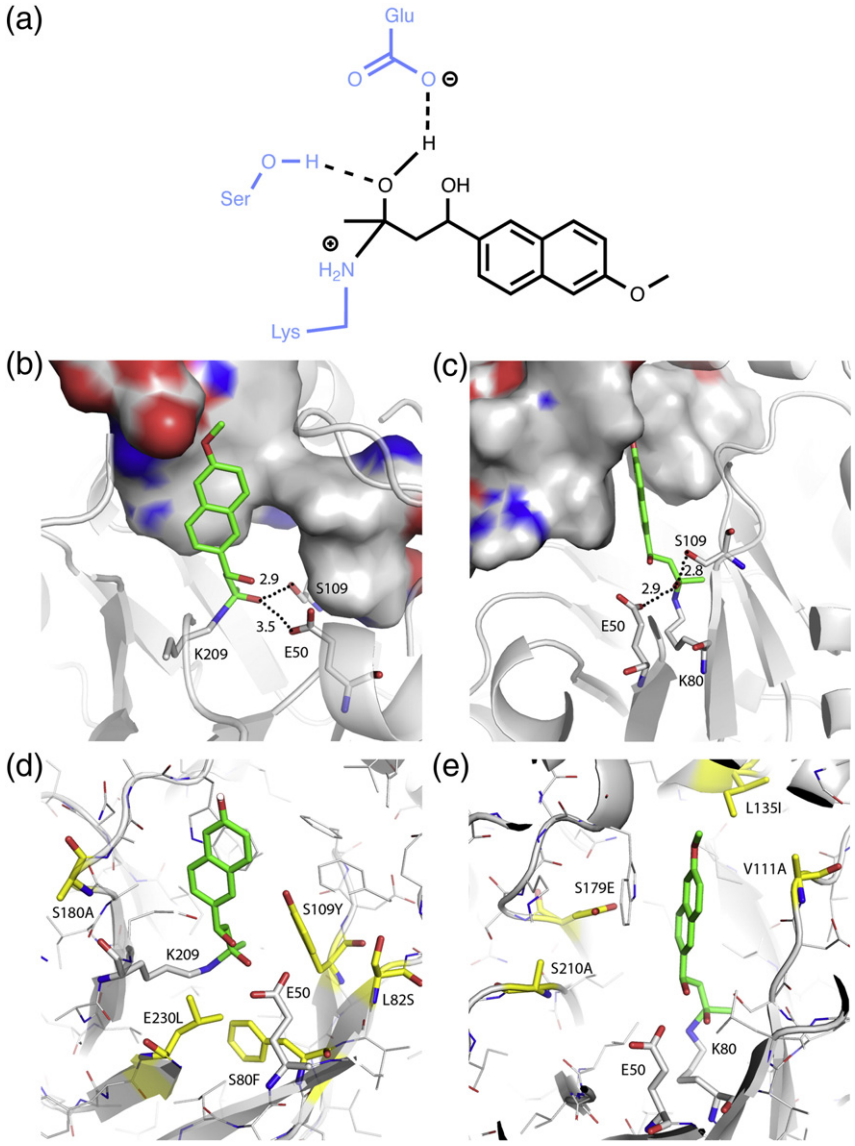


**Fig. 2.** Computational design of retroaldolases and experimental optimization. (a) Schematic of the theozyme used in the design calculations. Structure of the RA114 (b) and RA117 (c) computational designs. Broken lines show hydrogen-bonding interactions with the hydroxyl group of the carbinolamine intermediate. The amino acid substitutions identified during site-saturation mutagenesis of the active site are shown in yellow in RA114.3 (d) and RA117.1 (e).

catalysts by computational loop remodeling and by *in vitro* protein evolution using yeast display with a mechanism-based inhibitor.

## Results

### Computational design strategy

As in our previous work [2,7], we focused on amine catalysis of the retroaldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone [19] to give 6-methoxy-2-naphthaldehyde and acetone (Fig. 1). However, the water molecule in the original theozyme was replaced with the side chains of two amino acids, an aspartic or glutamic acid plus a serine or threonine, which can make hydrogen-bonding interactions directly with the carbinolamine. We hypothesized that such residues would be better suited for acid/base chemistry than a loosely bound water molecule. For example, carboxylic acids are effectively utilized for acid/base catalysis in aspartyl proteases and glycosidases. The carboxylic acid side chain of Asp/Glu could promote several steps in the retroaldolase reaction (Fig. 1), including (i) formation of the carbinolamine intermediate

**Table 1.** Steady-state parameters of the designed retroaldolases

| Design | $k_{cat}$ ($s^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) | $k_{cat}/k_{uncat}$ |
|---|---|---|---|---|
| RA114 | $3.22 \times 10^{-6}$ | 34 | 0.095 | 494 |
| RA115 | $1.51 \times 10^{-5}$ | 1740 | 0.0087 | 2320 |
| RA116 | $1.38 \times 10^{-5}$ | 1570 | 0.0088 | 2110 |
| RA117 | $2.76 \times 10^{-5}$ | 473 | 0.058 | 4270 |
| RA120 | $3.63 \times 10^{-6}$ | 880 | 0.0041 | 559 |

Kinetic assays were performed in 25 mM HEPES buffer (100 mM NaCl; pH 7.4) at room temperature using 5 μM enzyme, up to 640 μM substrate and 3% acetonitrile for substrate solubility. The $k_{cat}$ and $K_M$ parameters were obtained by fitting initial rates to the Michaelis–Menten equation. Values for catalysts with $K_M > 640$ μM are extrapolations and hence only approximate. $k_{cat}$ and $K_M$ are based on initial rates and a single measurement at each substrate concentration.

For comparison, $k_{uncat} = 6.5 \times 10^{-9}$ $s^{-1}$; the $k_{cat}$ and $K_M$ values of our previously designed retroaldolases vary from 8 to 150 $\mu s^{-1}$ and from 210 to 1600 μM, respectively [1,7].

by protonation of the oxyanion, (ii) breakdown of the carbinolamine with release of water to generate a protonated Schiff base, (iii) proton abstraction from the β-hydroxyl group of the intermediate to initiate C–C bond cleavage, (iv) protonation of the resulting enamine, and (v) general-base-assisted hydrolysis of the protonated Schiff base to release acetone and regenerate the catalyst. This theozyme contrasts with earlier designs in which the hydrogen-bonding amino acids tyrosine and histidine–aspartate were built in to interact with the β-hydroxyl of the retroaldol substrate rather than with the carbinolamine hydroxyl.

Theozymes consisting of the carbinolamine covalently bound to the catalytic lysine and hydrogen bonded to two amino acids via the hydroxyl group (Fig. 2a) were constructed as described in Materials and Methods. In our previous work, we used a composite transition state consisting of superimposed reaction intermediates and transition states. Here we focus on the carbinolamine intermediate as a transition-state surrogate because the hydroxyl group approximates the position of the water molecule that participates in the reaction (Fig. 1). RosettaMatch was then used to search for sets of backbone positions in six protein scaffolds that had successfully hosted retroaldolase theozymes based on the ordered water motif [20].

Circa 40,000 Asp/Glu–Ser/Thr matches were designed. After initial minimization, 2423 matches had excellent geometry without clashes with the scaffold backbone. For each of these matches, we redesigned a sphere of residues close to the ligand and sampled side-chain rotamers of residues outside this shell using the Rosetta enzyme design methodology [21]. Twenty-two designs were then selected from this set based on the active-site geometry, protein–carbinolamine interaction energy, and the total system energy according to the Rosetta energy function. They were further optimized using Foldit [22] by reverting unnecessary mutations to the amino acids present in the starting scaffold and by introducing new mutations to minimize the overall system energy. Seven designs were subsequently selected for experimental characterization based on (i) hydrophobic packing of the naphthyl group, (ii) supporting interactions for the Asp/Glu residue (salt bridges or interactions to

**Table 2.** Kinetics of representative RA114 variants

| Variant | $k_{cat}$ ($s^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) | $k_{cat}/k_{uncat}$ | $(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{mut}$ |
|---|---|---|---|---|---|
| RA114.1 | $1.35 \times 10^{-3}$ | 582 | 2.32 | $2.08 \times 10^{5}$ | 1 |
| RA114.1-E50A | $2.05 \times 10^{-4}$ | 179 | 1.15 | $3.16 \times 10^{4}$ | 2 |
| RA114.1-K209R | $3.35 \times 10^{-6}$ | 1260 | 0.00 | $5.16 \times 10^{2}$ | 869 |
| RA114.1-K209A | $9.43 \times 10^{-7}$ | 458 | 0.00 | $1.45 \times 10^{2}$ | 1126 |
| RA114.1-E50A/K209A | $1.38 \times 10^{-6}$ | 252 | 0.01 | $2.13 \times 10^{2}$ | 5021 |
| RA114.2 | $6.83 \times 10^{-4}$ | 30 | 22.6 | $1.04 \times 10^{5}$ | 1 |
| RA114.2-E50A | $1.08 \times 10^{-3}$ | 71 | 15.2 | $1.66 \times 10^{5}$ | 1.5 |
| RA114.2-K209R | $4.83 \times 10^{-7}$ | 155 | 0.0032 | $7.54 \times 10^{1}$ | 7140 |
| RA114.2-K209A | $3.83 \times 10^{-6}$ | 109 | 0.035 | $5.93 \times 10^{2}$ | 639 |
| RA114.2- E50A/K209A | $5.83 \times 10^{-6}$ | 68 | 0.086 | $8.99 \times 10^{2}$ | 263 |
| RA114.2-E50S | $1.72 \times 10^{-3}$ | 54 | 31.8 | $2.65 \times 10^{5}$ | 0.7 |
| RA114.3[a] | $(2.63 \pm 0.12) \times 10^{-3}$ | 107 ± 13 | 24.6 | $4.04 \times 10^{5}$ | — |
| RA114.4[b] | $(2.4 \pm 0.13) \times 10^{-3}$ | 18 ± 3 | 133 | $3.69 \times 10^{5}$ | — |

Kinetic assays were performed in 25 mM HEPES buffer (100 mM NaCl; pH 7.4) at room temperature using 5 μM enzyme and up to 320 μM substrate. We used 3% acetonitrile for substrate solubility; $k_{cat}$ and $K_M$ are based on initial rates and a single measurement at each substrate concentration. RA114.3 and RA114.4 assay conditions: 0.1 μM enzyme, 2% acetonitrile, and triplicate measurement at each point.

[a] RA114.3 is a RA114.1 variant with the S80F and S180A mutations.

[b] RA114.4 was generated by loop library screening with RA114.3.

**Table 3.** Kinetics of representative RA117 variants

| Variant | $k_{cat}$ ($s^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) | $k_{cat}/k_{uncat}$ | $(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{mut}$ |
|---|---|---|---|---|---|
| RA117.1 | $(1.18 \pm 0.13) \times 10^{-3}$ | 113 ± 33 | 11 | $1.82 \times 10^5$ | — |
| RA117.1[a] | $1.55 \times 10^{-3}$ | 390 | 3.96 | $2.37 \times 10^5$ | 1 |
| RA117.1-E50A[a] | $6.50 \times 10^{-5}$ | 341 | 0.19 | $1.00 \times 10^4$ | 21 |
| RA117.1-K80R[a] | $1.38 \times 10^{-5}$ | 1140 | 0.012 | $2.12 \times 10^3$ | 329 |
| RA117.1-K80A[a] | $6.17 \times 10^{-7}$ | 209 | 0.0029 | $9.38 \times 10^1$ | 1356 |
| RA117.2[b] | $(1.18 \pm 0.04) \times 10^{-3}$ | 39 ± 6 | 30 | $1.82 \times 10^5$ | — |
| RA117.3[c] | $(3.00 \pm 0.14) \times 10^{-3}$ | 92 ± 13 | 32 | $4.62 \times 10^5$ | — |
| RA117.4[d] | $(21.0 \pm 0.72) \times 10^{-3}$ | 373 ± 23 | 56 | $3.23 \times 10^6$ | — |

Kinetic assays were performed in 25 mM HEPES buffer (100 mM NaCl; pH 7.4) at room temperature using 0.1 μM enzyme, up to 360–500 μM substrate and 2–4% acetonitrile for substrate solubility. Measured in triplicate at each concentration.

[a] Kinetics assay at 5 μM enzyme and up to 320 μM substrate. $k_{cat}$ and $K_M$ are based on initial rates and a single measurement at each substrate concentration.

[b] RA117.2 was derived from RA117.1 following loop remodeling.

[c] RA117.3 contains a three-residue WET duplication at residues 183–185 in RA117.1.

[d] RA117.4 was obtained after yeast display.

stabilize the negative charge with good packing of the side chain), and (iii) packing of the lysine side chain to restrict its movement. Their sequences are shown in Scheme S1.

## Experimental characterization of the designs

The seven selected designs were cloned into a T7 expression vector and produced in *Escherichia coli*. Of these, five expressed as soluble protein at high levels and all exhibited detectable retroaldolase activity. This 100% success rate for the soluble proteins is significantly higher than in previous enzyme design efforts and likely reflects the use of scaffolds and lysine positions that had previously yielded functional retroaldolases. Nevertheless, the $k_{cat}$ and $K_M$ values of the new enzymes (Table 1) are, at best, only comparable to the steady-state parameters of the earlier designs. In fact, judging from their $k_{cat}/K_M$ values, the chemical efficiency of these catalysts is 1 order to 2 orders of magnitude lower than that of the best previously described *in silico* designs [2,7]. The catalytic benefit of the new hydrogen-bonding motif over an explicit water molecule, if any, thus appears to be small. This motif could be inherently unproductive or placement of the additional functional groups may have been insufficiently accurate to achieve effective catalysis.

For further characterization, we focused on the designs that had the lowest $K_M$ (RA114), the highest $k_{cat}$ (RA117), and the least product inhibition (RA115). All three designs are based on the TIM barrel protein indole-3-glycerolphosphate synthase (PDB code 1A53) [2,7]. Active-site scanning [7] was used to optimize the catalytic efficiency of these designs. For RA115, the best single mutation yielded only a 2-fold improvement; thus, this variant was not further pursued. In contrast, for RA114 and RA117, which share the same glutamic acid but differ in the position of the catalytic lysine (Fig. 2b and c), several sites were readily optimizable. Favorable mutations were combined to give variants RA114.1, RA114.2, and RA117.1 that have $k_{cat}/K_M$ values that are 25, 250, and 66 times larger than that of their corresponding starting design (Tables 2 and 3). RA114.1 contains the substitutions L82S, S109Y, and E230L (Fig. 2d), whereas RA114.2 combines L82S, W183F, and E230F; the four mutations in RA117.1 are V111A, L135I, S179E, and S210A (Fig. 2e).

To determine whether the designed lysine and glutamic acids are important for catalysis, we carried out knockout studies with the optimized variants (Tables 2 and 3). Mutation of the catalytic lysine to alanine or arginine decreased activity more than 600-fold for RA114.1 and RA114.2 and 300-fold for RA117.1. The E50A mutation in RA117.1 decreased the catalytic efficiency 20-fold. In contrast, mutation of glutamic acid to alanine in RA114.1 and RA114.2 only slightly affected activity, suggesting that this residue is not designed optimally for its intended role in catalysis; in this case, another nearby glutamic acid at position 210 may be able to provide the desired acid/base functionality in the absence of Glu50 as both residues are within hydrogen-binding distance of the catalytic lysine in the experimentally determined structure.

RA114.1 and RA117.1 were subjected to a second round of active-site optimization. Further improved versions of RA117.1 were not found, but the activity of RA114.1 was augmented an additional 12-fold by the S80F/S180A mutations giving rise to the RA114.3 variant (Table 2).

## Comparison of the designed models to experimental structures

The structures of RA114.3 and RA117.1 were experimentally determined by X-ray crystallography

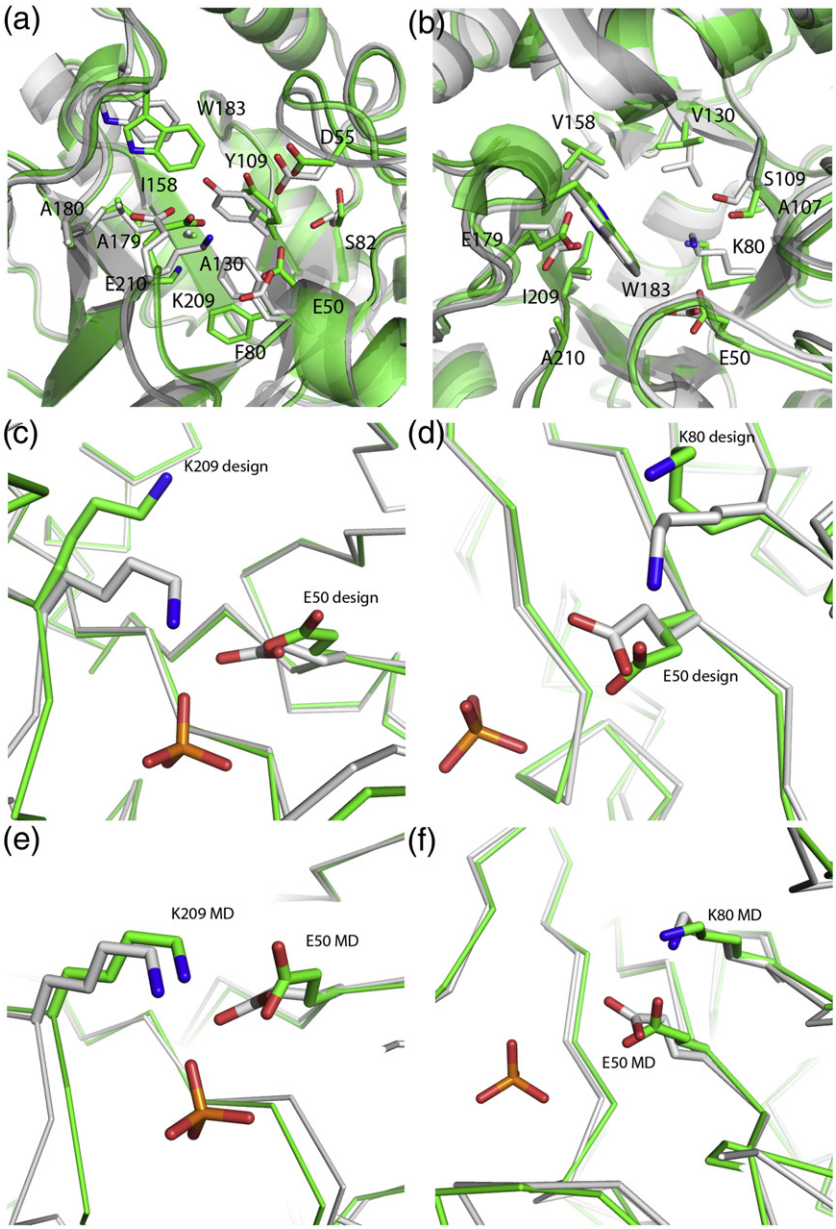


**Fig. 3.** Structural comparison of design models (green) with their crystal structures (gray). (a) RA114.3 and (b) RA117.1 are compared to the parent 1A53 scaffold, highlighting the amino acid changes in the active site (the catalytic glutamate, Glu50, has the same identity as in 1A53). (c) RA114.3 and (d) RA117.1 are depicted with the nucleophilic lysine and the glutamate that interacts with the carbinolamine. In RA114, the lysine is pulled away from the designed position by a bound phosphate. In RA117, the catalytic residues are close to the designed positions. MD simulations of the RA114.3 (e) and RA117.1 (f) variants more closely recapitulate the position of the catalytic lysine seen in the crystal structures.

at high resolution (1.6 Å and 1.5 Å, respectively; Table S1) and compared to the design models (with the additional substitutions modeled using Rosetta; Fig. 3a and b). The overall $C^{\alpha}$ RMSD for RA114.3 compared to the Rosetta model is 0.7 Å with an RMSD of 1.1 Å for the three largest loops bracketing the active site (structures were superimposed over all possible $C^{\alpha}$ atoms). For RA117.1, the corresponding values are 0.7 Å for the overall $C^{\alpha}$ RMSD and 0.9 Å for the loop RMSD. A phosphate molecule from the crystallization buffer is bound at both active pockets in the native phosphate-binding site of the parent scaffold (Fig. 3c and d).

In the case of RA114.3, there is a 5-Å movement of the ε-amino group of the catalytic lysine, Lys209, compared to the design. The altered rotameric preference likely results from favorable electrostatic interactions with the glutamic acid that was

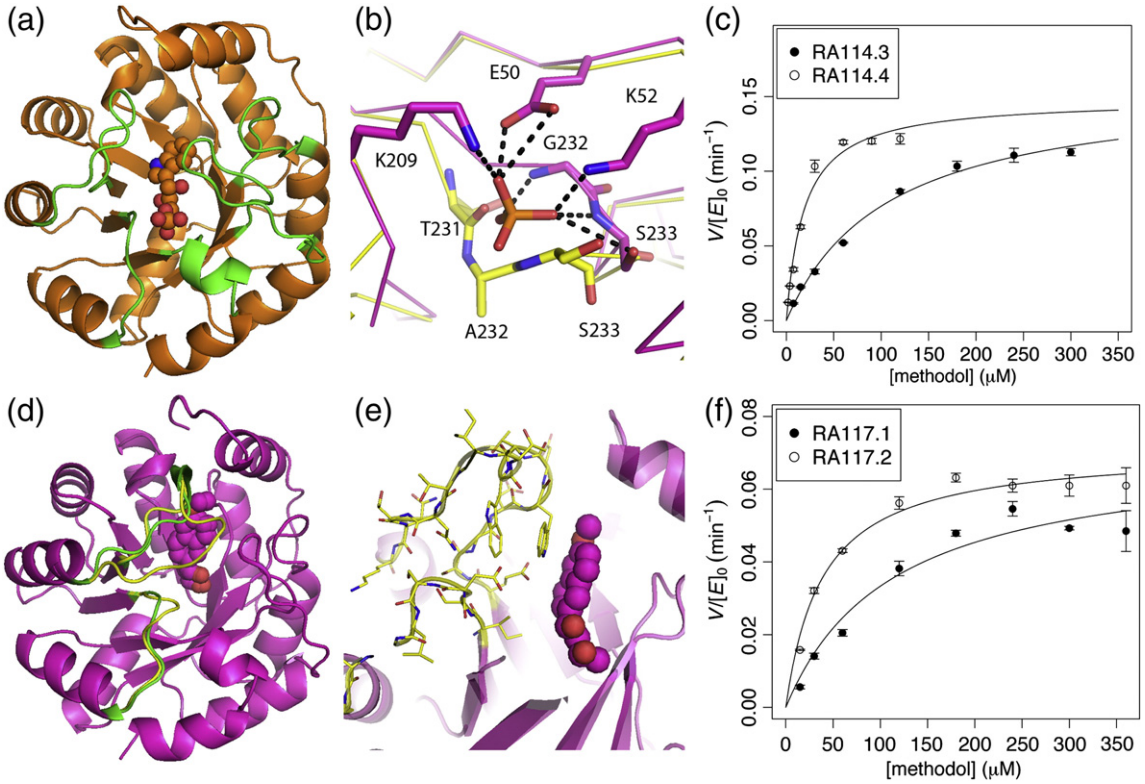


**Fig. 4.** Remodeling of active-site loops in RA114.3 and RA117.1. (a) 1A53 scaffold indicating in green the active-site loops subjected to computational redesign by insertions. (b) X-ray structure of the redesigned phosphate-binding loop in RA114.4 (yellow; loop **20** in Table 4) compared to the X-ray structure of the RA114.3 parent variant. The new loop completely abolishes the phosphate interactions with G232 and S233 in RA114.3 (T231–S233 residues in the RA114.4 structure are 1.3–2.0 Å from the RA114.3 phosphate). (c) Retroaldolase velocity *versus* substrate concentration profiles for RA114.3 and the improved loop variant RA114.4. (d) Cartoon representation of the modified loops in RA117.2 (yellow; loops **5** and **7** in Table 4) in comparison to the parent variant RA117.1 (green). (e) Representation of the designed loops in RA117.2. (f) Comparison of retroaldolase activity of RA117.1 and the loop variant RA117.2.

supposed to function as a general acid/base (Glu50), another nearby glutamic acid (Glu210), and the bound phosphate (Fig. 3c). The $C^{\delta}$ of Glu50 likewise shifts by 2 Å due to the electrostatic interactions with the bound phosphate ($d_{OE–phosphate}$ = 2.4 Å) and Lys52 ($d_{OE–NZ}$ = 2.9 Å). In RA117, the side chain of Lys80 is more deeply buried in the core and better held in place by surrounding residues, but there still is a 3-Å movement of the ε-amine toward the surface of the barrel. The designed glutamic acid, also at position 50, is close to the designed orientation with a 1-Å deviation of the $C^{\delta}$ carbon (distance between Glu50 from the phosphate and Lys52 is $d_{OE–phosphate}$ = 4.2 Å and $d_{OE–NZ}$ = 3.5 Å, respectively).

The significant displacements from the models clearly reflect problems in the design strategy. They highlight the importance of holding catalytic residues fixed through extensive interactions with second-shell residues. To determine the extent to which these rearrangements could have been anticipated, we carried out Rosetta side-chain optimization calculations with a simplified Coulomb potential and molecular dynamics (MD) simulations. In MD trajectories starting from RA114.3 (Rosetta model with a bound phosphate), the amino group of the catalytic lysine moved toward its position in the experimental structure with an average distance of 1.6 ± 0.3 Å (calculated over eight MD models; each model was averaged over the last 200 ps of the MD trajectory; Fig. 3e). This is clearly a better prediction than the initial designed position, which was 5 Å away. In RA117.1, the MD simulations also recapitulated the position of the catalytic lysine in the crystal structure (0.6 ± 0.3 Å; Fig. 3f). The position of Glu50 did not change substantially in either MD simulation.

The mutations found during optimization of RA114 and RA117 are distributed throughout the respective active sites (Fig. 2d and e). In RA114.3, four of the five mutations (L82S, S109Y, E230L, and S80F) may influence the conformation of the catalytic residues and restrict their flexibility (S109 was designed to interact with the carbinolamine). The fifth targeted residue, S180A, points away from the active site and

**Table 4.** RA114 and RA117 computational loop modifications tested experimentally

| RA117.1 | | | | RA114.3 | | | |
|---|---|---|---|---|---|---|---|
| | Sequence | I[a] | LOOP/WT[b] | | Sequence | I[a] | LOOP/WT[b] |
| **WT** | **LEINK (186–190)** | | | **WT** | **LIGSS (230–234)** | | |
| **1** | LTDNKK | 1 | + | **13** | SHASSA | 1 | + |
| **2** | LTSKKK | 1 | + | **14** | SISSHK | 1 | + |
| **3** | YNGVYNT | 2 | + | **15** | AADASA | 1 | + |
| **4** | LKIGKTR | 2 | + | **16** | AISGSSK | 2 | + |
| **5** | WTITNTNK | 3 | ++ | **17** | AISSSLKG | 3 | + |
| **6** | AWKNKYKVTK | 5 | + | **WT** | **GISE (211–214)** | | |
| **WT** | **IAGI (209–212)** | | | **18** | FKSETS | 2 | ++ |
| **7** | ISDSL | 1 | +++ | | | | |
| **8** | ISQSAL | 2 | + | | | | |
| **9** | FAVDSL | 2 | + | | | | |
| **10** | IAVDGL | 2 | ++ | | | | |
| **11** | IAYSAA | 2 | ++ | | | | |
| **12** | IGEGSI | 2 | ++ | | | | |
| **WT** | | | | | **D(S[c])PSG (55–58)** | | |
| **19[d]** | | 1 | | | (NDGS)–(ANDT)–(GS)–(ARDQEGHLPV)–(G) | | |
| **WT** | | | | | **LIGSS (230–234)** | | |
| **20[d]** | | 1 | | | (DHLV)–(AGISTV)–(AGST)–(AGST)–(GS)–(S) | | |

[a] Number of insertions.
[b] Activity of the variant compared to the wild type (WT).
[c] In RA117.1.
[d] Loop libraries.

is situated in the transition region between a core β-strand and the loop involved in binding of the substrate naphthyl ring. Two of the four substitutions in RA117.1, V111A and L135I, are in close proximity to the designed position of the naphthyl ring. The S179E mutation introduces a hydrogen bond to the indole ring of Trp183 and may stabilize the loop in a favorable conformation for substrate binding. The S210A mutation is probably involved in stabilizing the binding loop by providing second-shell packing interactions for Arg181 and possibly Glu179.

## Remodeling of the active-site loops

The active sites in both designs are situated under three major loops and adjacent to two phosphate-binding loops (Fig. 4a). These loops were evolutionarily optimized for the original reaction and are not likely to be optimal for the retroaldolase reaction. Since these protein segments are essential for substrate binding and the presence of nearby phosphate may be deleterious for the designed reaction, we remodeled the loops to remove native scaffold functionality and improve binding of the retroaldol substrate.

Starting with the Rosetta models, we rebuilt the loops to increase contacts with the substrate using fragments from the PDB with RosettaRemodel [23]. Loops of specified length were built by fragment insertion following by closure in the middle of the loop using the cyclic coordinate descent algorithm [24]. Following closure, we optimized the sequence of the loop using RosettaDesign. Loops with low energies, consisting of other residues than only glycine and alanine, and favorable interactions with the substrate that do not introduce cavities in the structures were selected and tested experimentally. Six to twelve individual loop redesigns were tested for RA114 and RA117 (Table 4), but none yielded increases in activity.

Because of the difficulty in accurately controlling the conformation of loops not well anchored to the rest of the scaffold (as is the case for TIM barrel proteins), we computationally built families of loops using RosettaRemodel that could be evaluated by experimental screening (Table 4). We investigated two approaches for coupling experimental optimization with the design process. In the first strategy, applied to RA114, we computed families of sequences for loop regions, then constructed libraries that recapitulated the sequence variation in these families using doped oligonucleotides and screened for more active clones. In the second strategy, used for RA117, we combined the individual loop redesigns **1–6** with **7–12** (Table 4).

The loop library screening procedure with RA114.3 yielded a variant, RA114.4, with 4-fold higher catalytic efficiency (Fig. 4b and c and Table 2). The sequence of the 234–238 segment was changed from LIGSS in the original design to VTASGS in the improved variant. The remodeled loop disrupts the phosphate-binding site as intended (Fig. 4b compares the X-ray structure of the segment before and after remodeling). Loop recombination screening with RA117.1 identified another clone, RA117.2, with 3-fold increased activity in which loops **5** and **7** were introduced (Tables 3 and 4 and Fig. 4f). These two loops are in direct contact with each other in the designed structure (Fig. 4d and e),

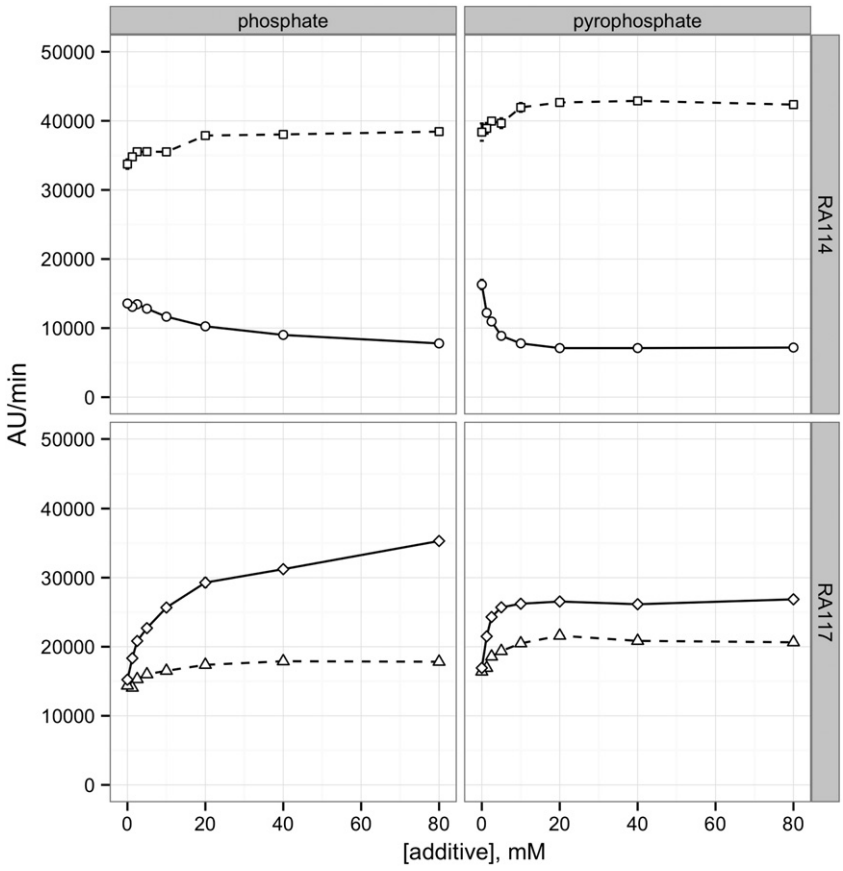


**Fig. 5.** Effect of phosphate and pyrophosphate on retroaldolase activity. Continuous and broken lines, without (RA114.3, circles; RA117.3, diamonds) and with (RA114.4, squares; RA117.2, triangles) phosphate-binding loop remodeling, respectively. Each point is an average of three measurements. Error bars are small and overlap with symbols; they are omitted for clarity. Lines connecting the points are for illustration purposes only. The assay was performed in 25 mM HEPES buffer (100 mM NaCl; pH 7.4) at room temperature with 0.25 μM enzyme and 100 μM substrate.

and the synergistic activity of the loop replacements (neither loop increases activity on its own) is likely due to interactions between them. A third variant with activity similar to RA117.2, RA117.3 (Table 3), was obtained by introducing a three-residue WET duplication at residues 183–185 in RA117.1.

## Effect of phosphate on retroaldolase activity

The phosphate anion bound to RA114.3 and RA117.1 is located within the original substrate-binding site of indole-3-glycerolphosphate synthase but does not overlap with the designed binding site for the retroaldol substrate. To investigate possible phosphate dependence of the designs, we assayed their activity in the presence of phosphate and pyrophosphate.

The activity of RA114.3 decreases upon addition of increasing amounts of phosphate or pyrophosphate (Fig. 5, top). The close proximity of the phosphate-binding site to the catalytic lysine may perturb the residue from its designed position via electrostatic interactions (as observed in the crystal structure) and/or affect its p$K_a$. Since loop remodeling is likely to disrupt the phosphate-binding sites, we expected that the loop redesigns would have altered sensitivity to the phosphate compounds. Indeed, the RA114.4 variant containing the remodeled 234–238 segment does not bind phosphate (there is no phosphate bound in the X-ray structure; Fig. 4b) in contrast to the crystallographically characterized RA114.3 variant (Fig. 3c) and is not inhibited by phosphate or pyrophosphate; it is instead slightly activated by these compounds (Fig. 5, top).

In contrast to RA114.3, the activity of RA117.3 increases markedly upon addition of phosphate (Fig. 5, bottom). The catalytic lysine is further away from the phosphate-binding site, but Glu50 is quite close and the phosphate ion may help to fix it in a catalytically more favorable orientation. These

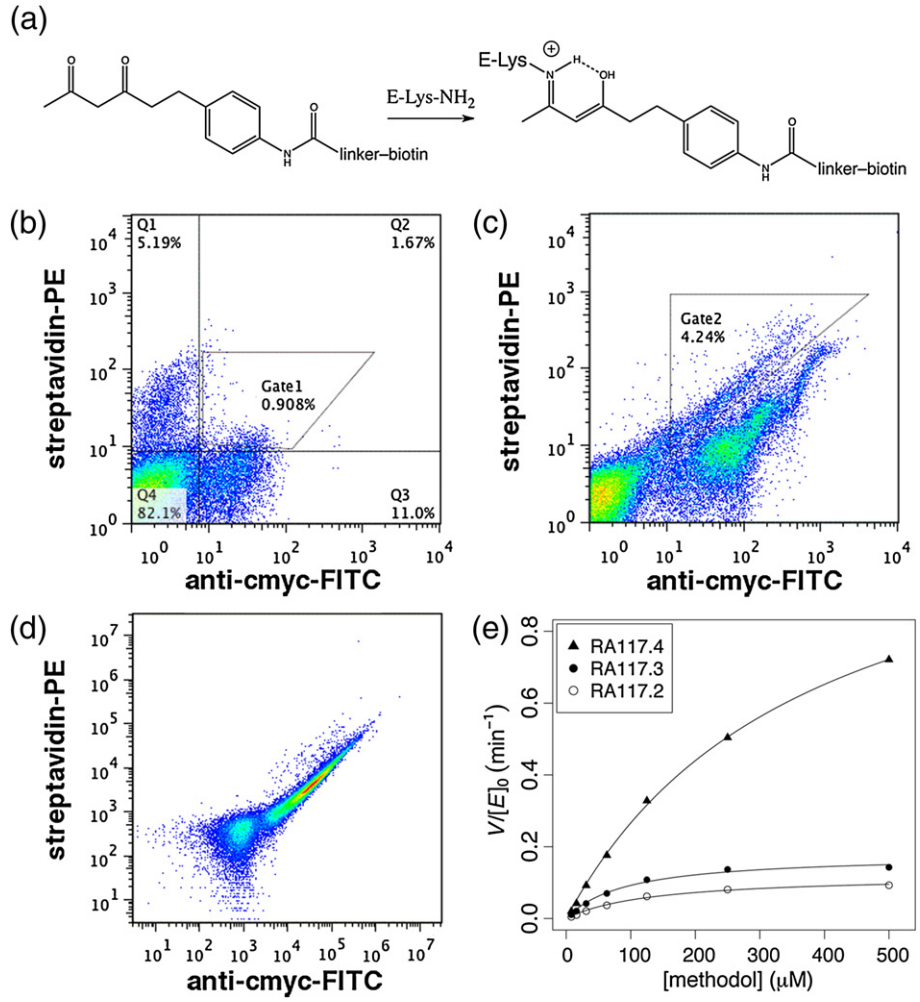


**Fig. 6.** Optimization of RA117.1 by yeast display and FACS. (a) A biotinylated diketone was used as a mechanism-based inhibitor. It reacts with the catalytic lysine of the retroaldolase to form a covalent adduct. (b) Yeast display of the naïve RA117.1 retroaldolase library. The displayed catalysts were visualized by immuno-labeling of the c-myc epitope tag; enzymes capable of reacting with the biotinylated diketone were detected by binding to fluorescently labeled streptavidin. (c) Third sort of the RA117.1 library. (d) Labeling of one clone obtained from the sort shown in (c). An Influx cell sorter was used for (b) and (c), while (d) was measured on an Accuri C6 flow cytometer. (e) Michaelis–Menten plot of representative RA117 variants before (RA117.2 and RA117.3) and after yeast display (RA117.4).

effects are substantially dampened following loop remodeling in the RA117.2 variant (Fig. 5, bottom).

The modulation of activity by phosphate ion in different directions in the designs before and after loop remodeling illustrates how small-molecule binding can perturb reactivity even in the absence of evolutionary selection for such contributions. The refinement of such effects by natural selection may have given rise to the complex allosteric regulation of many modern day enzymes.

## Retroaldolase activity of yeast display identified variants

Direct screening for increased enzymatic activity assay in crude lysates is typically limited to ~1000 variants. The throughput for binding selections, in contrast, can be 4 orders of magnitude higher. Taking advantage of the ability of 1,3-diketones to trap the catalytic lysine covalently [14], we explored yeast display [25] as a means of optimizing the retroaldolases. We used fluorescence-activated cell sorting (FACS) to search large libraries of displayed retroaldolases for variants that react rapidly with a biotinylated diketone probe (Fig. 6a). A similar diketone-based probe has been successfully used previously with phage-displayed libraries to optimize the retroaldolase activity of small peptides [10] and catalytic antibodies [26].

Libraries were generated from RA114.3 and RA117.1 containing on average 3–6 mutations per gene. The variants were generated with mutagenic

**Table 5.** Representative peptides, antibodies, and previously evolved computational designs that catalyze the retroaldol cleavage of racemic 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone

| Catalyst | $k_{cat}$ ($s^{-1}$) | $K_M$ (μM) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) | $k_{cat}/k_{uncat}$ | Reference |
|---|---|---|---|---|---|
| FT-YLK3[a] | $9.3 \times 10^{-6}$ | 1800 | $5.2 \times 10^{-3}$ | $1.4 \times 10^{3}$ | 10 |
| FluS303-FTYLK3[a] | $1.2 \times 10^{-5}$ | 1100 | $1.1 \times 10^{-2}$ | $1.9 \times 10^{3}$ | 10 |
| ScLS-2[b] | $3.3 \times 10^{-5}$ | 1000 | $3.3 \times 10^{-2}$ | $3.0 \times 10^{3}$ | 12 |
| β-Peptide[c] | $2.2 \times 10^{-3}$ | 5000 | $4.3 \times 10^{-1}$ | $3.0 \times 10^{3}$ | 9 |
| Ab 33F12[d] | $1.8 \times 10^{-3}$ | 43 | $4.2 \times 10^{1}$ | $2.4 \times 10^{5}$ | 13 |
| Ab 38C2[d] | $1.2 \times 10^{-2}$ | 25 | $4.8 \times 10^{2}$ | $1.8 \times 10^{6}$ | 14 |
| Ab 84G3[d] | $5.8 \times 10^{-2}$ | 23 | $2.5 \times 10^{3}$ | $3.6 \times 10^{6}$ | 15 |
| Ab 93F3[d] | $4.4 \times 10^{-2}$ | 15 | $2.9 \times 10^{3}$ | $2.7 \times 10^{6}$ | 15 |
| RA45.2-10[e] | $3.8 \times 10^{-3}$ | 80 | $4.7 \times 10^{1}$ | $5.8 \times 10^{5}$ | 7 |
| RA110.4-6[e] | $3.8 \times 10^{-3}$ | 69 | $5.5 \times 10^{1}$ | $5.9 \times 10^{5}$ | 7 |
| RA95.5-8[e] | $1.7 \times 10^{-1}$ | 200 | $8.5 \times 10^{2}$ | $2.6 \times 10^{7}$ | 27 |

[a] Lysine-rich peptides selected from phage display libraries by covalent capture with 1,3-diketones.
[b] An engineered variant of *Saccharomyces cerevisiae* lumazine synthase containing a lysine-rich protein tunnel.
[c] An amphiphilic β-peptide with an array of $\beta^3$-homolysine residues. This catalyst was assayed with 4-phenyl-4-hydroxy-2-oxobutyrate, which is somewhat more reactive than 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone.
[d] Catalytic antibodies elicited either with 1,3-diketone or 1,3-keto sulfone haptens.
[e] Evolutionarily optimized computational designs.

oligonucleotides that targeted residues pointing into the active site from the eight β-strands of the barrel. The total library size was on the order of ~$10^7$. After labeling with the biotinylated diketone, we analyzed the samples by FACS. A representative example of successive rounds of FACS selection with the diketone probe is shown in Fig. 6. Although the starting libraries contained relatively few expressible variants (occupancy of fluorescein-isothiocyanate-positive quadrants in Fig. 6b was about 10%), the subpopulation of functional clones increased after repeated rounds of selection for expression and diketone binding, as shown for RA117 in Fig. 6c and d. Selection pressure was steadily increased by decreasing the probe labeling time in each round.

Rapidly reacting RA114.3 and RA117.1 variants exhibited a similar pattern of 5–7 mutations situated on β-strands 2, 3, 4, and 5 (RA114.3) and β-strands 3, 4, 5, and 6 (RA117.1). Unfortunately, when tested individually, selected variants were found to cleave the retroaldol substrate less efficiently than the starting enzymes, mainly due to lower turnover and product inhibition. Indeed, a limitation of this approach is the lack of selective pressure on steps in the reaction cycle beyond irreversible formation of the vinylogous amide from the catalytic lysine and the diketone probe. Nevertheless, we hypothesized that subsets of the selected substitutions might activate the lysine and hence increase overall retroaldolase activity. We therefore prepared a small library in which individual yeast display mutations were introduced into the optimized RA114.4, RA117.2, and RA117.3 variants in different combinations. Screening 150 individual colonies from each library for activity in crude cell lysates did not yield more active RA114.4 variants, but we identified several RA117 variants that were 2- to 4-fold more active than the parent clones (Fig. 6e). The best variant RA117.4 had five mutations in the active site compared to RA117.3: L82I, A107L, S109F, L233I, and G235L. In contrast, screening of 200 random clones from each unselected library did not identify active variants. Thus FACS selection can help identify useful sets of mutations that optimize one of the reaction steps and also have the potential to increase the overall catalytic efficiency.

## Discussion

Natural enzymes are characterized by complex arrays of functional groups that enable high catalytic efficiency. Computationally designed enzymes, in contrast, are considerably less sophisticated. For example, the absence of groups other than the catalytic lysine likely accounts for the comparatively low activities of previously described retroaldolases [2]. Recent directed evolution experiments with the RA95 retroaldolase have shown that adding supporting catalytic functionality in the vicinity of the catalytic lysine can substantially boost catalytic efficiency to levels approaching those of natural aldolases [27]. An asparagine residue and a tyrosine residue that were introduced into the RA95 active site during optimization are likely to promote critical proton transfers during catalysis, either directly or via bound water molecules. As a consequence, the evolved RA95.5-8 variant achieves a catalytic efficiency ($k_{cat}/K_M$) of 850 $M^{-1}$ $s^{-1}$ and a rate enhancement over the

background reaction ($k_{cat}/k_{uncat}$) of $2.6 \times 10^7$. For comparison, the best catalytic antibodies cleave the same substrate with rate accelerations in the range $2.4 \times 10^5$ to $3.6 \times 10^6$ (Table 5) [26,27].

In the current study, we incorporated additional catalytic functionality into the retroaldolase active site directly by design. We succeeded in generating active enzymes using a scheme that exploited a carboxylic acid and an alcohol side chain in addition to the Schiff-base-forming lysine. Although the starting catalysts (Table 1) were not more active than previous designs or simple peptide-based retroaldolases, which accelerate the reaction by more than a factor of $10^3$ (Table 5) [9–11], the extra functionality of the general acid/base residues imbedded in the TIM scaffold may make them more evolvable. In fact, one round of site-directed saturation mutagenesis afforded catalysts with the best $k_{cat}$ and $k_{cat}/K_M$ values observed for artificial retroaldolases prior to extensive directed evolution (Tables 2 and 3). The resulting $>10^5$-fold rate accelerations are within 1 order to 2 orders of magnitude of the rate enhancements reported for the best catalysts developed for this reaction (Table 5), including catalytic antibodies ($k_{cat}/k_{uncat} = 10^4$ to $>10^6$) [15] and highly optimized computational designs such as RA95.5-8 ($k_{cat}/k_{uncat} > 10^7$) [27]. Moreover, the activity of RA117 could be boosted by an additional order of magnitude by yeast display (Table 2). In the optimized RA117 variant knockout studies showed that the designed carboxylic acid side chain of Glu50 contributes to catalysis. That said, there are many steps in the retroaldolase reaction cycle, and we do not yet have direct evidence that it functions as intended.

Proficient catalysis requires accurate positioning of the key catalytic functionalities (Fig. 2a–c). As evident from the crystal structures in Fig. 3c and d, the charged catalytic lysine and glutamate residues are not sufficiently locked into the designed target conformations. A clear challenge for enzyme design is to introduce sufficient packing and hydrogen-bonding interactions to more precisely control the positions of catalytic groups. The importance of fixing functional residues in the desired conformation was highlighted in a recent study of small-molecule binding-site design, where substitutions which reduced the entropy of binding site side chains increased ligand binding affinity [28]. Computational methods such as MD simulations (Fig. 3e and f) and explicit computation of side-chain rotamer probabilities [29] show promise for identifying regions of designs where greater precision is required; it will be important both to further explore these and other methods for evaluating the precision of active-site design and to develop approaches for better constraining catalytic functionalities in the design process.

Given the complexity of enzyme catalysis, experimental optimization techniques will continue to be important for generating practical catalysts in the years to come. The designs presented here are attractive candidates for further optimization because of the greater diversity of nearby functional groups; for example, tuning of the orientations of the carboxylate containing residues by multiple rounds of directed evolution could significantly increase the rate of catalysis. In this context, the loop remodeling computational techniques have shown promise for increasing catalytic efficiency. The combination of high-throughput yeast display selection for covalent intermediate formation followed by lower throughput screening of the identified mutations for increase in activity could be generally useful for directed evolution of reactions proceeding through a covalent enzyme–substrate intermediate. Additionally, using better mimics of high-energy retroaldolase intermediates, such as the β-keto sulfone haptens used to generate highly efficient antibodies [15], or optimizing the substrate structure for the active site may increase the activity of our designs even further.

## Materials and Methods

### Matching and enzyme design

The class I aldolase mechanism [8] considered here involves attack of a reactive lysine residue on the carbonyl group of the substrate to form a tetrahedral carbinolamine intermediate. The design calculations focused on protein scaffolds 1A53, 1IGS, 1LBF, 1LBL, 1OHO, and 1THF. Lysine residues were placed at positions that had given rise to active designs in previous work. Glu/Asp–Thr/Ser pairs were arranged to interact productively with the carbinolamine [2,7]. We also incorporated amino acid substitutions found in previous studies to increase protein expression and solubility [30].

RosettaMatch was used to place the theozyme (Appendix A Supplementary Information) and the catalytic residues onto the protein backbones [2,20]. For each successful placement of the theozyme, all residues within 7 Å of the ligand were redesigned, and side-chain rotamers of residues 10 Å from the ligand were repacked using the Rosetta enzyme design methodology (in the design calculations, the Monte Carlo search covers all rotamer conformations for all possible amino acid residues, whereas in repacking calculations, the search covers all rotamer conformations for the native amino acid) [21]. The designed structures were sorted based on overall Rosetta energy, ligand binding energy, and accuracy of theozyme recapitulation (distances, angles, and torsions between the small-molecule and the catalytic residues).

### Gene synthesis, protein expression and purification, mutagenesis, and reaction assay

In total, seven designs (RA114–RA120) were produced. The genes encoding the designs were synthesized by GenScript and cloned into the NdeI/XhoI sites of a

pET-29b(+) vector containing a C-terminal $His_6$ tag. Proteins were expressed and purified as described previously [2,7,16]. Retroaldolase product formation was assayed by fluorescence spectroscopy ($\lambda_{ex}$ = 330 nm and $\lambda_{em}$ = 452 nm) in black, flat-bottom 96-well microtiter plates (Corning). Protein activity was measured in a 25 mM HEPES buffer (100 mM NaCl; pH 7.4), at room temperature using 5 μM enzyme, up to 640 μM substrate and 3% acetonitrile for substrate solubility. Product concentration was determined from a standard product fluorescence curve, and Michaelis–Menten kinetics was derived accordingly.

### Enzyme optimization by site-directed mutagenesis

Designs RA114, RA115, and RA117—all based on the 1A53 [31] scaffold—were optimized by randomizing selected active-site positions and combining favorable mutations. The genes for the variants were generated according to the Kunkel protocol [32] using degenerate oligos synthesized by Integrated DNA Technologies (IDT). The first round of site-directed mutagenesis included 20 positions (NNK: S57, S80, L82, L107, S109, and G186; L156A/L + G177L/A, V207L/V + E210V/E, E230MLFQE + W183W/Y + E50Q/E, I132VLIFA + A130VLI, F111VILFY + F88LIFY, and D55NYEQH + I158VLIF) in RA114, 15 positions (NNK: P56, S88, R183, G211, and S233; W186 FYLMW + W7PTSVLIWF, W87YF + E57QLE + F111WFY, N179LIFMN + E209QMIFLE, and T52MLKT + S50AS + S180LVITS) in RA115, and 14 positions (NNK: L32, L82, F88, S109, D110, S179, and S210; V111FAYLI + L135FAYLI and V130F + V158G + I132A + A107VILSTA + I209ELVI) in RA117 (all residue numbering is according to the 1A53 structure without the leading Met). The second round of optimization included six positions (E50, D55, S80, S109, S180, and R181) in RA114 and ten positions (W7, A8, V11, V12, L82, S109, L156, W183, I209, and L230) in RA117. The activity of the variants was assessed in a 96-well plate format. Starter cultures of *E. coli* BL21(DE3)* were obtained by inoculating clones in 1 mL Terrific Broth with kanamycin and growing overnight. After transfer of 50 μL of the starters to 1 mL of Terrific Broth, growth for 3 h, and induction for 4 h with 0.1 mM IPTG, the expanded cultures were harvested by centrifugation at 4000 rpm for 15 min. After five freeze/thaw cycles, the pellets were resuspended in 0.25–0.3 mL of 25 mM HEPES buffer (100 mM NaCl; pH 7.4). We transferred 0.15 mL to a 96-well assay plate in which 4 μL of 6 mM retroaldol substrate in acetonitrile was added (final concentration, 0.16 mM). The reaction was monitored as mentioned above.

### Structure solution and refinement

Crystallization screening was performed using a microbatch-under-oil crystallization method at 4 °C [33]. After optimization, crystals of RA114.3 (Northeast Structural Genomics Consortium, NESG, target OR62), RA114.4 (OR422), and RA117.1 (OR63) were grown for structure determination in drops composed of 1.0 μL of protein and 1.0 μL of precipitant solution [100 mM sodium acetate, 100 mM sodium phosphate monobasic, and 40% (w/v) polyethylene glycol 400; pH 5.0] under paraffin oil (Hampton Research). Data sets were collected at beamline X4C at the National Synchrotron Light Source at 100 K. Diffraction data from single crystals were processed with the HKL2000 package [34]. The structures were solved by molecular replacement using the program BALBES [35] and models 1IGS, 3TC7 (RA114.3), and 1JUL for RA114.3, RA114.4, and RA117.1, respectively. The models were completed using iterative cycles of manual rebuilding in Coot [36] and refined against 1.50 Å (RA114.3), 1.93 Å (RA114.4), or 1.60 Å (RA117.1) data with the program PHENIX [37]. The quality of the model was inspected by the program PROCHECK [38].

### MD simulations

MD simulations were carried out on eight different Rosetta models of RA117.1 or RA114.3 retroaldolase variants with the MD package Q [39]. A sphere $d$ = 36 Å was placed on the crystallographic phosphate P atom coordinates and equilibrated for 100 ps releasing the positional harmonic restraints and increasing the temperature to 300 K. The 400-ps production phase was performed after an equilibration trajectory of 50 ps [40]. From the last 200 ps of the trajectory, 400 frames were collected and averaged for comparison with corresponding experimental structures.

### Diversification and assay of loop remodeled variants

RA114.3 and RA117.1 loops were designed with the Rosetta loop modeling protocol as implemented in the Rosetta enzyme design application [3,23]. The loop ensembles were clustered based on backbone RMSD, and the loop clusters were evaluated based on the interactions with the ligand and surrounding active site. Additional loop insertions were generated experimentally by duplicating three-residue segments in the active-site loops in the RA117.1 variant.

Designed loops were introduced individually or as libraries by Kunkel mutagenesis [32]. Primers were ordered from IDT, and libraries were synthesized with degenerate oligos. Loops were assayed as described above but for the assay volume that consisted of 0.1 mL of resuspended lysed pallet with additional 0.1 mL of 0.2 mM substrate. For each designed loop, activity was measured for eight different clones. In the case of libraries, 200–400 variants were screened.

The effect of phosphate and pyrophosphate on retroaldolase activity was investigated by performing the activity assay at varying concentrations of the additive. The assay was performed in 25 mM HEPES buffer (100 mM NaCl; pH 7.4) at room temperature with 0.25 μM enzyme. The retroaldolase substrate concentration was kept fixed at 100 μM.

### Yeast display library construction

Plasmids pET-29b(+)–RA114.3 and pET-29b(+)–RA117.1 were used to prepare single-stranded template for site-directed mutagenesis or library construction by a modified Kunkel mutagenesis protocol. Mutagenic and wild-type oligos were obtained from IDT. One to three mutagenic oligos were synthesized containing various combinations of degenerate codons. The diversified

residues point toward the substrate and were situated on eight β-strands of the scaffold (Tables S2 and S3).

The library was assembled by annealing the single-stranded DNA template with a mixture of 5′-phosphorylated mutagenic and wild-type oligos. The frequency of mutations was controlled by changing the ratio between oligos annealing to the same fragment of the gene [41]. Two sets of libraries were made for both RA114 and RA117. In the first set with lower mutagenesis level (libRA114-L, libRA117-L), the ratio of wild-type to mutagenic oligos was 1:1, and in the second set with a higher mutagenesis level (libRA114-H, libRA117-H), the ratio was 1:2. Ten to twelve colonies were sequenced to estimate diversity, and the rest of the cultures were grown overnight, shaking at 37 °C. The plasmid libraries were used as templates for PCR amplification with flanking pETCON-up/pETCON-down specific sequences to mediate homologous recombination into the yeast display vector. We mixed 0.5 μg of pETCON vector, linearized with NdeI/XhoI and purified on agarose gel, with 1 μg of PCR amplified library of RA114 or RA117, electroporated into competent EBY100 yeast cells, prepared, and treated following the protocol described previously [42]. The transformed cells were plated on selective C-UT agar plates [42] and incubated at 30 °C until colonies appeared, 48–72 h. The remainder of the transformed cells was transferred to selective SDCAA liquid medium [25] and grown at 30 °C with shaking until the induction step (see below). The diversity of the libraries was confirmed by randomly sequencing an additional 10–12 colonies.

### FACS screening

Stationary-phase liquid yeast cultures harboring RA114 or RA117 libraries in SDCAA medium were harvested by centrifugation at 2000*g* for 10 min and resuspended in an appropriate volume of inducing SGCAA medium to obtain an $OD_{600}$ of about 0.5. To promote protein production and display on the surface of yeast cells, we incubated cultures at 18 °C or 30 °C with vigorous shaking for 24 h. After induction, culture volume containing $10^7$ cells ($OD_{600}$ = 1 corresponds to $10^7$ cells/mL) was collected by centrifugation and washed three times by 25 mM HEPES (pH 7.4) containing 100 mM NaCl (reaction buffer, RB). Labeling with biotinylated diketone inhibitor (BDK; Fig. 6a, synthesis Scheme S2) was achieved by resuspending cells in 100 μL of RB containing ~30 μM BDK [from a 20-mM stock in 50:50 (v/v) acetonitrile:methanol] and incubating for 10–120 min at room temperature with gentle shaking. After BDK labeling, cells were washed three times with RB supplemented with 0.1% bovine serum albumin (BSA) and then resuspended in 50 μL of RB + BSA containing streptavidin-phycoerythrin and fluorescein isothiocyanate conjugated to an anti-myc-antibody. The cell suspension was incubated for 60 min at 4 °C with gentle shaking. Before sorting, cells were washed three times with RB + BSA and finally resuspended in RB + BSA. Cells positive for expression and labeling with BDK were selected by appropriate gating on an Influx cell sorter (BD Biosciences), collected in 1 mL of SDCAA medium containing 1 μL of a penicillin-streptomycin solution (15140-122; Invitrogen) to suppress bacterial contamination, and grown overnight at 30 °C. We determined the recovery efficiency to be ~50% by plating 10–100 μL of recovery culture on C-UT plates and counting the number of colonies. The selection cycle was repeated three to four times decreasing BDK labeling time to enrich the population of faster-reacting retroaldolase variants. Efficiency of cell labeling was assessed by flow cytometry on Accuri C6 (BD Biosciences).

### Accession numbers

Coordinates for RA114.3, RA114.4, and RA117.1 have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank with accession codes 3TC7, 4LNY, and 3TC6, respectively.

## Acknowledgments

We would like to thank Melissa Maglaqui, Seema Sahdev, and Colleen Ciccosanti for their technical assistance. We thank John Everett for project coordination support. S.B. acknowledges support from The Swedish Research Council (623-2008-497) and Foundation BLANCEFLOR Boncompagni-Ludovisi, née Bildt. D.B. thanks Defense Advanced Research Projects Agency/Defense Threat Reduction Agency for financial support. D.H. is grateful for support from the Schweizer Nationalfonds and the ETH Zurich. This work was supported by a grant from the Protein Structure Initiative of the National Institutes of Health (U54-GM094597).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.10.012.

*Received 27 July 2013;*
*Received in revised form 8 October 2013;*
*Accepted 9 October 2013*
Available online 23 October 2013

***Keywords:***
retroaldolase;
enzyme design;
protein engineering;
enzyme optimization

† S.B., Y.K., and L.W. These authors contributed equally to this work.

***Abbreviations used:***
MD, molecular dynamics; FACS, fluorescence-activated cell sorting; IDT, Integrated DNA Technologies.

## References


[1] Bjelic S, Nivon LG, Celebi-Olcum N, Kiss G, Rosewall CF, Lovick HM, et al. Computational design of enone-binding proteins with catalytic activity for the Morita-Baylis-Hillman reaction. ACS Chem Biol 2013;8:749–57.

[2] Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, et al. *De novo* computational design of retro-aldol enzymes. Science 2008;319:1387–91.

[3] Richter F, Blomberg R, Khare SD, Kiss G, Kuzin AP, Smith AJT, et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. J Am Chem Soc 2012;134:16197–206.

[4] Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, et al. Kemp elimination catalysts by computational enzyme design. Nature 2008;453:190–5.

[5] Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, Clair JLS, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 2010;329:309–13.

[6] Zhang X, DeChancie J, Gunaydin H, Chowdry AB, Clemente FTR, Smith AJT, et al. Quantum mechanical design of enzyme active sites. J Org Chem 2008;73:889–99.

[7] Althoff EA, Wang L, Jiang L, Giger L, Lassila JK, Wang ZZ, et al. Robust design and optimization of retroaldol enzymes. Protein Sci 2012;21:717–26.

[8] Gefflaut T, Blonski C, Perie J, Willson M. Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. Prog Biophys Mol Biol 1995;63:301–40.

[9] Müller MM, Windsor MA, Pomerantz WC, Gellman SH, Hilvert D. A rationally designed aldolase foldamer. Angew Chem Int Ed 2009;48:922–5.

[10] Tanaka F, Fuller R, Barbas CF. Development of small designer aldolase enzymes: catalytic activity, folding, and substrate specificity. Biochemistry 2005;44:7583–92.

[11] Tanaka F, Barbas CF. A modular assembly strategy for improving the substrate specificity of small catalytic peptides. J Am Chem Soc 2002;124:3510–1.

[12] Wörsdörfer B, Henning LM, Obexer R, Hilvert D. Harnessing protein symmetry for enzyme design. ACS Catal 2012;2:982–5.

[13] Tanaka F, Lerner RA, Barbas CF. Reconstructing aldolase antibodies to alter their substrate specificity and turnover. J Am Chem Soc 2000;122:4835–6.

[14] Wagner J, Lerner RA, Barbas CF. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. Science 1995;270:1797–800.

[15] Zhong GF, Lerner RA, Barbas CF. Broadening the aldolase catalytic antibody repertoire by combining reactive immunization and transition state theory: new enantio- and diastereoselectivities. Angew Chem Int Ed 1999;38:3738–41.

[16] Wang L, Althoff EA, Bolduc J, Jiang L, Moody J, Lassila JK, et al. Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of *de novo* enzyme design. J Mol Biol 2012;415:615–25.

[17] Heine A, DeSantis G, Luz JG, Mitchell M, Wong CH, Wilson IA. Observation of covalent intermediates in an enzyme mechanism at atomic resolution. Science 2001;294:369–74.

[18] Heine A, Luz JG, Wong CH, Wilson IA. Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 angstrom resolution. J Mol Biol 2004;343:1019–34.

[19] List B, Barbas CF, Lerner RA. Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis. Proc Natl Acad Sci USA 1998;95:15351–5.

[20] Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, et al. New algorithms and an *in silico* benchmark for computational enzyme design. Protein Sci 2006;15:2785–94.

[21] Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D. *De novo* enzyme design using Rosetta3. PLoS One 2011;6:e19230.

[22] Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature 2010;466:756–60.

[23] Huang PS, Ban YEA, Richter F, Andre I, Vernon R, Schief WR, et al. RosettaRemodel: a generalized framework for flexible backbone protein design. PLoS One 2011;6:e24109.

[24] Canutescu AA, Dunbrack RL. Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 2003;12:963–72.

[25] Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc 2006;1:755–68.

[26] Tanaka F, Fuller R, Shim H, Lerner RA, Barbas CF. Evolution of aldolase antibodies *in vitro*: correlation of catalytic activity and reaction-based selection. J Mol Biol 2004;335:1007–18.

[27] Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, et al. Evolution of a designed retroaldolase leads to compleate active site remodeling. Nat Chem Biol 2013;9:494–8.

[28] Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, et al. Computational design of ligand-binding proteins with high affinity and selectivity. Nature 2013;501:212–6.

[29] Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. Protein Sci 2011;20:753–7.

[30] Khersonsky O, Kiss G, Röthlisberger D, Dym O, Albeck S, Houk KN, et al. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. Proc Natl Acad Sci USA 2012;109:10358–63.

[31] Hennig M, Darimont BD, Jansonius JN, Kirschner K. The catalytic mechanism of indole-3-glycerol phosphate syntase: crystal structures of complexes of the enzyme from *Sulfolobus solfataricus* with substrate analogue, substrate, and product. J Mol Biol 2002;319:757–66.

[32] Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 1985;82:488–92.

[33] Chayen NE, Stewart PDS, Maeder DL, Blow DM. An automated-system for microbatch protein crystallization and screening. J Appl Crystallogr 1990;23:297–302.

[34] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–26.

[35] Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr Sect D Biol Crystallogr 2002;58:1772–9.

[36] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2126–32.

[37] Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr Sect D Biol Crystallogr 2002;58:1948–54.

[38] Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.

[39] Marelius J, Kolmodin K, Feierberg I, Åqvist J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. J Mol Graphics Modell 1998;16:213–25.
[40] Bjelic S, Brandsdal BO, Åqvist J. Cold adaptation of enzyme reaction rates. Biochemistry 2008;47:10049–57.
[41] Kipnis Y, Baker D. Comparison of designed and randomly generated catalysts for simple chemical reactions. Protein Sci 2012;21:1388–95.
[42] Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel 2010;23:155–9.