doi:10.1016/j.jmb.2011.10.043 J. Mol. Biol. (2012) 415, 615–625

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Structural Analyses of Covalent Enzyme–Substrate Analog Complexes Reveal Strengths and Limitations of *De Novo* Enzyme Design

Ling Wang[1]†, Eric A. Althoff[1,2]†, Jill Bolduc[3], Lin Jiang[1,4], James Moody[1], Jonathan K. Lassila[5], Lars Giger[6], Donald Hilvert[6], Barry Stoddard[3] and David Baker[1]*

[1]*Department of Biochemistry, University of Washington, Seattle, WA 98195, USA*
[2]*Arzeda Corp., Seattle, WA 98102, USA*
[3]*Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA*
[4]*Department of Biological Chemistry, UCLA, Los Angeles, CA 90095, USA*
[5]*Department of Biochemistry, Stanford University, Stanford, CA 94305, USA*
[6]*Laboratory of Organic Chemistry, ETH Zürich, 8093 Zurich, Switzerland*

Received 1 July 2011;
received in revised form
21 October 2011;
accepted 26 October 2011
Available online
3 November 2011

Edited by M. Guss

Keywords:
retro-aldolase;
de novo;
enzyme design;
complex structure;
directed evolution

We report the cocrystal structures of a computationally designed and experimentally optimized retro-aldol enzyme with covalently bound substrate analogs. The structure with a covalently bound mechanism-based inhibitor is similar to, but not identical with, the design model, with an RMSD of 1.4 Å over active-site residues and equivalent substrate atoms. As in the design model, the binding pocket orients the substrate through hydrophobic interactions with the naphthyl moiety such that the oxygen atoms analogous to the carbinolamine and β-hydroxyl oxygens are positioned near a network of bound waters. However, there are differences between the design model and the structure: the orientation of the naphthyl group and the conformation of the catalytic lysine are slightly different; the bound water network appears to be more extensive; and the bound substrate analog exhibits more conformational heterogeneity than typical native enzyme–inhibitor complexes. Alanine scanning of the active-site residues shows that both the catalytic lysine and the residues around the binding pocket for the substrate naphthyl group make critical contributions to catalysis. Mutating the set of water-coordinating residues also significantly reduces catalytic activity. The crystal structure of the enzyme with a smaller substrate analog that lacks naphthyl ring shows the catalytic lysine to be more flexible than in the naphthyl–substrate complex; increased preorganization of the active site would likely improve catalysis. The covalently bound complex structures and mutagenesis data highlight the strengths and weaknesses of the *de novo* enzyme design strategy.

**Corresponding author*. Biomolecular Structure and Design, Department of Biochemistry, University of Washington and Howard Hughes Medical Institute, HSB Room J-555, 1959 NE Pacific Street, Seattle, WA 98195, USA. E-mail address: dabaker@u.washington.edu.

† L.W. and E.A.A. contributed equally to this work.

Abbreviations used: PDB, Protein Data Bank; PEG, polyethylene glycol.

Introduction

Computational enzyme design approaches have the potential to produce new enzymatic catalysts for many chemical reactions. However, computational design is still in its infancy. While structures of apoenzymes have been determined for several designed enzymes,[1,2] to date, there are no structures of designed enzymes with covalently bound substrate analogs.

As described previously,[1] a two-stage computational approach was used to design enzymes catalyzing the retro-aldol reaction shown in Fig. 1. In the first step, we generated ideal active-site "theozymes" that consist of superimposed carbinolamine intermediate and bond-breaking transition-state models surrounded by a catalytic lysine residue, which forms a Schiff base with the substrate and hydrogen-bonding residues to position a water molecule near the carbinolamine oxygen. Locations in a set of scaffold proteins where one of these ideal active sites could be created were identified using the RosettaMatch algorithm.[4] In the second step, the residues within 8 Å of the model ligand were optimized to maximize binding of the intermediate/transition-state models, packing around the catalytic lysine residue, and nonpolar packing around the substrate naphthyl group.

In this study, we focus on one of the most active designs identified in our previous study, RA34,[1] which has a k_{cat}/K_m of 0.11 M^{-1} s^{-1}, comparable to computationally designed catalysts[2,5,6] for other reactions and in the range of previous peptide and catalytic antibody aldol catalysts.[7–13] To identify shortcomings in the design calculations and to guide the improvement of the design methodology, we first used mutagenesis and screening to optimize the active-site residues and to identify positions that were suboptimal in the original design. Next, we solved the crystal structure of the optimized enzyme covalently bound to substrate analogs both with and without the naphthyl group modeled in the design process. The contributions of the active-site residues to catalysis were probed by alanine scanning mutagenesis, and the contributions of the sequence changes that arose during optimization were investigated by reversion mutations. The structures and mutational analyses illustrate how the designed and optimized residues influence catalysis, as well as reveal areas for improvement in computational design methodology.

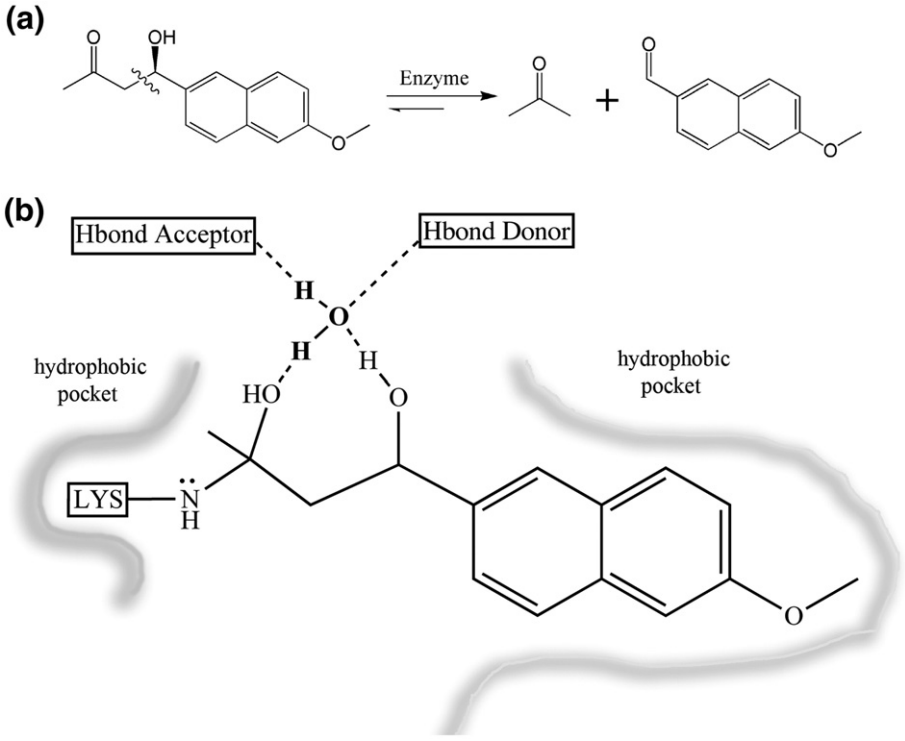

Fig. 1. Retro-aldol reaction and active-site description. (a) Schematic of the retro-aldolase reaction. The product is fluorescent (λ_{ex} = 330 nm, λ_{em} = 452 nm).[3] (b) The minimal active site used in the design calculations. The Schiff base serves as an electron sink to promote bond cleavage. The bridging water is positioned by hydrogen-bonding residues to allow proton shuttling on and off the substrate and the product.

Results

Design and optimization of the retro-aldol enzyme RA34

The RA34 design was created by using RosettaMatch[1] to identify a location on the TIM (*t*riosephosphate *isom*erase) barrel indole-3-glycerolphosphate synthase (1lbf) scaffold[14] where the ideal active site schematically illustrated in Figure 1b could be recapitulated. Rosetta design[15] calculations were subsequently carried out to optimize substrate/transition-state binding. In total, the design calculations introduced 13 mutations from the original 1lbf scaffold. The final design model contains (i) a catalytic lysine residue; (ii) surrounding hydrophobic residues to hold the lysine in place and to lower its pK_a;[9] (iii) several neighboring polar residues to stabilize the carbinolamine reaction intermediate via water-mediated interactions;[16] and (iv) a hydrophobic pocket designed to bind and orient the substrate naphthyl group. As reported previously,[1] the activity of the purified RA34 protein for the retro-aldol cleavage of (±)-methodol was well above background, but the catalytic parameters were quite low: $k_{cat} = 0.7 \times 10^{-4}$ s^{-1} and $k_{cat}/K_m = 0.11$ M^{-1} s^{-1}.

To investigate the extent to which the active site could be optimized and identify positions in which the computational design was suboptimal, we screened amino acid sequence variants at 19 positions around the active site (Fig. S1). Twelve residues were mutated into all other amino acids, and seven residues participating primarily in packing interactions were mutated to other hydrophobic residues. We started from a variant (Y51T) found serendipitously to increase RA34 activity by about 2-fold. Sequence changes were introduced via Kunkel mutagenesis using degenerate oligonucleotides. After transformation into BL21(DE3) cells,

Table 1. Steady-state enzyme activity

	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	$\frac{k_{cat}}{k_{uncat}}$	$\frac{(k_{cat}/K_m)^{mutant}}{(k_{cat}/K_m)^{RA34.6}}$
RA34 optimization					
RA34 original design	620	0.7E−04	0.11 ± 0.01	1.1E+04	0.009
RA34 Y51T	626	1.22E−04	0.19 ± 0.01	1.9E+04	0.016
RA34.6	30	3.60E−04	12 ± 1	5.54E+04	1
Ala mutation scan					
K159A	1696	6.63E−07	(3.9 ± 0.4)E−04	1.02E+02	3.3E−05
C83A	102	3.29E−04	3.2 ± 1.5	5.06E+04	0.3
S210A	41	1.53E−04	3.7 ± 1.2	2.35E+04	0.32
W8A	34	6.65E−05	2.0 ± 0.7	1.02E+04	0.18
P57A	136	1.37E−04	1.0 ± 0.2	2.11E+04	0.09
F112A	25	5.82E−05	2.3 ± 0.3	8.95E+03	0.19
I157A	39	1.72E−05	0.4 ± 0.2	2.65E+03	0.04
T51A	55	2.77E−04	5.0 ± 1.1	4.26E+04	0.44
C180A	31	2.62E−04	8.5 ± 0.4	4.03E+04	0.72
T211A	45	4.19E−04	9.3 ± 2.4	6.45E+04	0.83
M53A	98	1.54E−04	1.6 ± 0.5	2.37E+04	0.14
W58A	143	1.19E−04	0.8 ± 0.04	1.83E+04	0.07
W184A	102	1.18E−04	1.2 ± 0.2	1.82E+04	0.1
S81A	32	2.69E−04	8.4 ± 0.6	4.14E+04	0.7
S181A	42	3.67E−04	8.7 ± 1.6	5.65E+04	0.75
S231A	27	1.21E−04	4.5 ± 0.8	1.86E+04	0.39
L108A	40	1.53E−04	3.8 ± 0.6	2.35E+04	0.33
I133A	37	1.75E−04	4.7 ± 0.4	2.69E+04	0.39
I233A	62	1.93E−04	3.1 ± 0.6	2.97E+04	0.27
Reversion mutants					
C83T	107	2.34E−04	2.2 ± 1.3	3.60E+04	0.21
I233G	110	8.54E−05	0.8 ± 0.6	1.31E+04	0.08
T211Y	82	2.33E−04	2.8 ± 1.8	3.58E+04	0.27
C180V	58	2.58E−04	4.4 ± 1.7	3.97E+04	0.4
I157L	70	3.08E−04	4.4 ± 2.6	4.74E+04	0.41
P131A	88	2.98E−04	3.4 ± 1.9	4.58E+04	0.31
Water network mutants					
T51L	69	6.05E−05	0.88 ± 0.02	9.31E+03	0.073
T51I	65	4.54E−05	0.70 ± 001	6.98E+03	0.058
S81V	47	7.24E−05	1.55 ± 0.21	1.11E+04	0.13
T51V/S81A	50	2.91E−05	0.59 ± 0.08	4.47E+03	0.049
S210A/S231A	66	1.41E−05	0.21 ± 0.01	2.17E+03	0.018
T51V/S81A/S210A/S231A	96	1.30E−06	0.014 ± 0.003	2.00E+02	0.0012

individual constructs were expressed, and retro-aldolase activity was measured in clarified lysates. We began by screening single amino acid substitutions at each position (Table S1). Subsequently, variants that displayed at least 2-fold improvements in activity over the original design were combined to select the best possible combination, termed RA34.2 (Fig. S1). Additional mutations, found previously to increase the solubility of the scaffold (Olga Kheronsky, personal communication), were made outside of the active site, producing the variant RA34.3. An additional round of screening for single mutations was performed, and the top three active variants (RA34.4, RA34.5, and RA34.6) were identified and purified. Of these, RA34.6 had the highest activity, with a k_{cat} of 3.6×10^{-4} s^{-1}, a K_m of 30 μM, and a k_{cat}/K_m of ~12 M^{-1} s^{-1} (Table 1). Further single mutations of RA34.6 yielded less than 2-fold improvement.

In total, the seven mutations around the active site in RA34.6 relative to the original designed enzyme RA34 (Y51T, T83C, A131P, L157I, V180C, Y211T, and G233I) resulted in a 100-fold increase in k_{cat}/K_m: from $k_{cat}/K_m=0.11$ M^{-1} s^{-1} to $k_{cat}/K_m=12$ M^{-1} s^{-1}. Although the k_{cat}/K_m value of RA34.6 is still low compared to the values of most native aldolases,[17,18] which are in the range of 10^5 M^{-1} s^{-1}, it represents a 5×10^6-fold acceleration over the second-order rate constant for the same reaction catalyzed by the lysine analog butylamine at pH 7.5.[19] The rate acceleration of RA34.6 relative to the first-order nonenzymatic reaction in solution ($k_{cat}/k_{uncat}=5.5\times10^4$) exceeds those of all but the best of the aldolase catalytic antibodies and peptides that have been developed[7–13] based on the enamine mechanism of natural enzymes.[20,21] The rate acceleration of RA34.6 exceeds that of most computationally designed enzymes[2,5,6] and is comparable to that of designs after improvement by directed evolution.[22,23]

The positions and identities of the substitutions found during optimization in the original Rosetta design model suggest that they could increase activity by improving the packing around the catalytic lysine and the substrate naphthyl group,

Table 2. Crystallization, data collection, structure determination, and refinement

	Cross-linked inhibitor	Cross-linked 4h4m2p	Apo
Data collection			
Space group	$P3_121$	$P3_121$	$P3_121$
Cell dimensions			
a, *b*, *c* (Å)	62.68, 62.68, 123.68	61.92, 61.92, 121.24	62.59, 62.59, 123.70
α, β, γ (°)	90, 90, 120	90, 90 , 120	90, 90, 120
Wavelength (Å)	1.54	1.54	1.54
Resolution (Å)	50–2.09 (2.16–2.09)	50–2.4 (2.49–2.4)	50–2.1 (2.16–2.1)
Observations	110,823	176,391	248,913
Unique reflections	17,318	11,053	17,270
Data coverage (%)	99.0 (95.2)	99.1 (99.7)	96.8 (87.8)
Redundancy	2.9	3.9	5.0
R_{lin}	3.7 (15.3)	3.9 (29.0)	7.8 (34.0)
$I/\sigma I$	22.0 (6.9)	21.1 (5.8)	16.3 (5.1)
Mean FOM	0.80	0.74	0.74
Refinement			
Resolution range (Å)	19.47–2.09	53.6–2.4	50–2.1
Reflections	16,229	10,320	15,688
Completeness (%)	99.35	98.22	96.28
Total atoms			
Protein	1979	2036	1994
Water	61	60	82
Ligand	17	7	
Ion (SO_4)	10	5	15
R_{work} (%)	21.1	22.4	22.3
R_{free} (%)	26.0	26.8	29.0
RMSD			
Bonds (Å)	0.008	0.012	0.022
Angles (°)	1.028	1.4	2.0
Ramachandran plot			
Most favored (%)	92.1	93.3	90.7
Additionally allowed (%)	6.2	6.7	8.4
Generously allowed (%)	0.4	0	2
Disallowed (%)	1.3	0	0

Crystals were grown via vapor diffusion. Equal amounts of protein at 5 mg/ml in 100 mM NaCl and 25 mM Tris-HCl (pH 7.5) were mixed with well solutions consisting of 2 M ammonium sulfate, 4% PEG 400, and 100 mM sodium acetate (pH 5.5). Data were collected from a cryoprotectant buffer (2.1 M ammonium sulfate, 5% PEG 400, 0.2 M sodium acetate, and 500 mM sodium chloride) by a Rigaku Micromax 7HF with Saturn 944+ at −180 °C. Data were indexed, integrated, and scaled using the HKL2000 package. Molecular replacement and refinement were performed with the Phaser and REFMAC modules of CCP4i. Coot was used for model building.

as well as by altering the hydrogen-bonding network with substrate oxygens. To evaluate these possibilities, we performed structural studies of RA34.6.

Crystal structures of RA34.6

Previous attempts at solving the structure of the original RA34 design were unsuccessful,[1] but we were able to crystallize and determine the structures at 2.1–2.4 Å resolution (Table 2) of the improved RA34.6 protein in the absence of added inhibitors or substrate analogs, and of complexes of RA34.6 both with a mechanism-based inhibitor (1-(6-methoxy-2-naphthalenyl)-1,3-butanedione) and with a substrate analog that lacks a naphthyl ring (4-hydroxy-4-methyl-2-pentanone). The initial phases were determined via molecular replacement, using the coordinates of indole-3-glycerol-phosphate synthase [Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) entry 1a53, whose protein sequence is identical with 1lbf] as search model. In the original search model and in the corresponding starting model used for rebuilding, all residues that were subjected to computational redesign and subsequent mutagenesis were converted into alanine and then rebuilt manually into initial electron density maps to avoid modeling bias that might arise from the original computational design produced by Rosetta.

The structure of RA34.6 apoprotein was determined to 2.1 Å resolution (Fig. 2 and Table 2). Overall, the crystal structure is very similar to the RA34 design model with an RMSD of 0.4 Å over the C^{α} atoms. The most pronounced structural discrepancy is found in a flexible loop (residues 210–216) (Fig. 3a), which may be associated with the Y211T mutation. Density around the catalytic lysine suggests that more than one rotamer may be populated. Significant additional poorly ordered density emanates from the amino group of the catalytic lysine and extends into the binding site (Fig. 2a), suggesting Schiff base formation with an unknown compound from the protein preparation. Mass spectrometry of small crystals of the apoprotein showed the molecular mass to be 128 Da larger than expected; however, the adduct likely formed during crystallization as freshly purified RA34.6 samples have the expected molecular mass.

To better understand the interactions of RA34.6 with substrate, we next determined the structure of

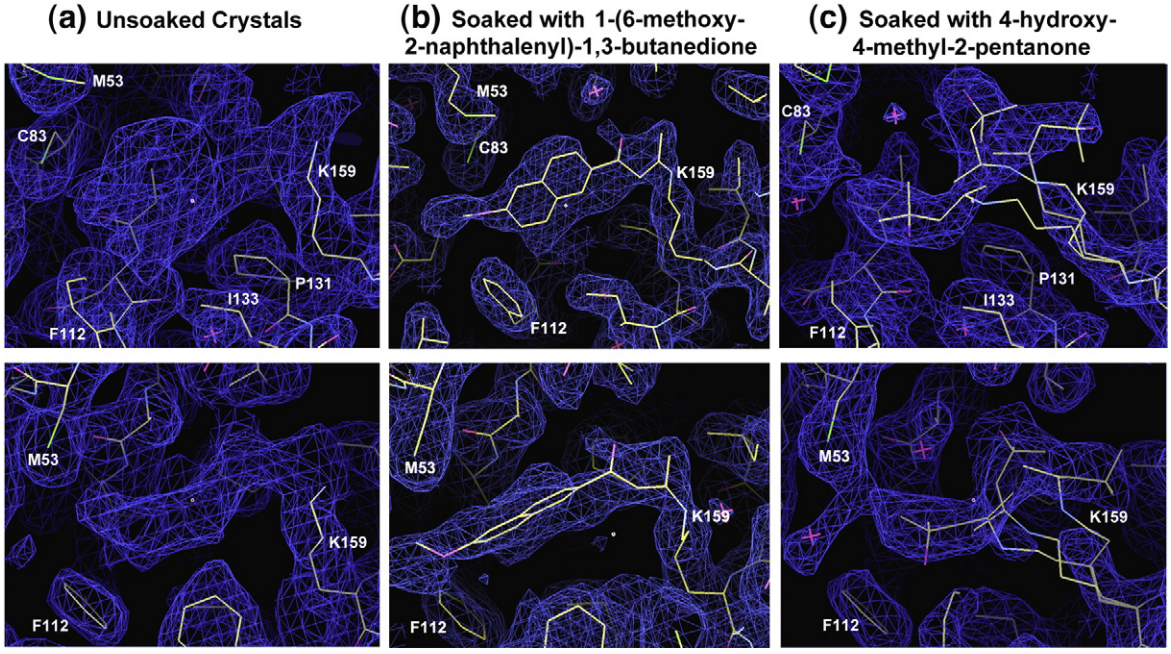

Fig. 2. Unbiased composite omit maps calculated after the building and refinement of protein side chains. The atoms and models corresponding to the bound substrate and product analogs were not modeled or used in any way for phase calculations prior to the generation of these maps. The three maps are all shown at two similar angles (viewed from the side of the binding pocket and from the top) for each crystal and data set. The contour level of (a) (corresponding to unsoaked crystals of a freshly purified enzyme) is 1.2σ; the contour levels of (b) and (c) (corresponding to soaks with substrate and product analogs) are 2σ. The ring system of the bound naphthalene group of the substrate analog (b) shows clear signs of mobility on the plane of the rings (i.e., up and down in the upper portion of (b)); the connectivity and rotameric modeling of chemical linkage to the active-site lysine (K159) look clear. Inverting the stereochemistry of this linkage results in a clear positive-difference density where the carbonyl oxygen is modeled.

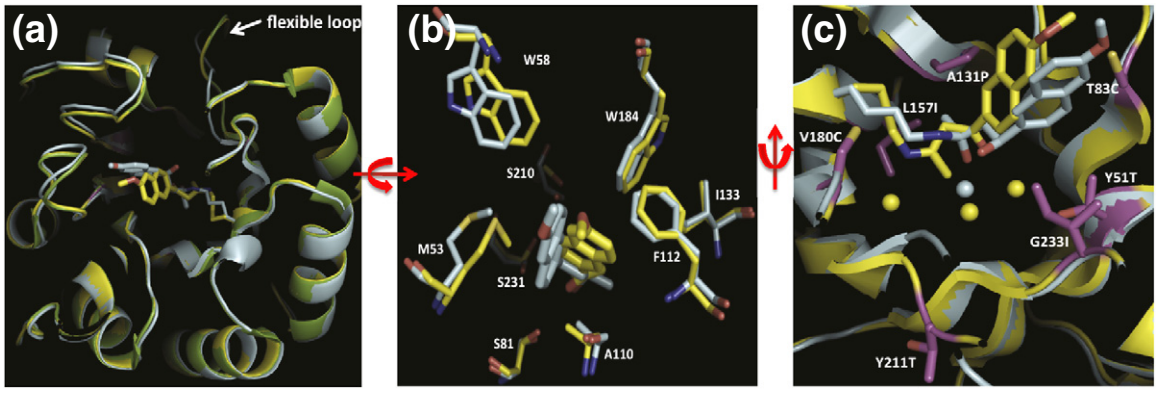

Fig. 3. Comparison of the structure of RA34.6 to the original design model. (a) Overlay of apo RA34.6 (green), bound RA34.6 (yellow), and the original design model RA34 (cyan).[1] The backbone conformations of apo and holo RA34.6 are nearly identical, with a C^{α} RMSD of 0.15 Å. The C^{α} RMSD of bound RA34.6 from the original design model of RA34 is 0.4 Å. The flexible loop in the highest discrepancy region (amino acids 210–216) is highlighted. (b) Active-site comparison of the cocrystal structures of RA34.6 (yellow) and of the design model (cyan) after superimposing all C^{α} atoms of the structures. The "superimposable" active site includes the carbinolamine intermediate, covalently bound Lys159, hydrogen-binding residues (Ser231, Ser210, and Ser81), and other hydrophobic residues contacting the substrate (Phe112, Ala110, Trp184, Met53, Trp58, and Ile133). The heavy-atom RMSD of this superimposable active site between the original design and the crystal structure of the evolved variant is 1.4 Å. (c) Comparison of water-mediated interactions in the active site in the RA34 model (cyan) with putative water positions modeled in the RA34.6 substrate analog complex structure (yellow). Only the side chains of the crystal structure are shown for simplicity. Optimized substitutions (from RA34 to RA34.6) in the active site are highlighted in purple. The modeled and crystallographic bound water molecules are shown as spheres. The PDB entry codes of apo and holo RA34.6 are 3O6Y and 3UD6, respectively. The RA34 design model had been reported previously.[1]

RA34.6 with covalently bound diketone substrate analog (1-(6-methoxy-2-naphthalenyl)-1,3-butanedione) at 2.1 Å resolution (Fig. 2b). The bound diketone was modeled in the final stage of model building and was fitted into a density that was clearly observable in completely unbiased composite omit electron density maps (Fig. 2). Electron density emanating from the catalytic lysine is consistent with the mechanism-based inhibitor and differs in shape, size, and strength from that observed in the "apo" structure (Fig. 2a). Although the density that we assign to the bound diketone analog exhibits considerable positional heterogeneity in the pocket (Fig. 2b), chemical linkage to the lysine is unambiguous, and the overall shape of the aromatic planar bicyclic naphthalene ring is well defined.

The RA34.6 structure with bound inhibitor was similar overall to the original RA34 design model, with an α-carbon RMSD of 1.4 Å over the active-site atoms, including the catalytic lysine, the three polar residues designed to position bound water molecules, and six surrounding apolar residues that pack on the naphthyl group and buttress the catalytic lysine (Fig. 3). A detailed comparison of the RA34.6 structure to the design model reveals several interesting differences. First, the catalytic lysine K159 appears to have more than one conformation in the structure but is also better packed by nonpolar groups than in the original RA34 design model. The two optimization-introduced mutations that contribute to better packing are L157I, which packs underneath the lysine, and A131P, which alters the lysine backbone conformation and may stabilize a slightly bent conformation of the lysine side chain (Fig. S3).

The second difference that we observed between the RA34.6 structure and the RA34 design model was that the naphthyl group of the substrate rotates and adopts an orientation similar to, but different from, that in the model. There is also a slight translational movement out of the hydrophobic pocket (Fig. 3a and b).

The third intriguing feature of the structure is a network of several putative water molecules near the oxygens of the substrate analog, within interaction distance of four polar residues at the bottom of the pocket (Fig. 4c). In the design model of RA34, we introduced hydrogen bonds from Tyr51 and Ser81 to a single water molecule to facilitate the reaction and to help shuffle the proton generated during the reaction. The RA34.6 structure, however, may indicate a more extensive water network with hydrogen bonds to at least three waters from four residues (Thr51, Ser81, Ser210, and Ser231).

A fourth notable feature of the structure is the tight packing of the naphthyl ring by the active-site residues (Fig. 4b). The substitutions G233I, T83C, and A131P introduced by optimization contribute to this tight packing. To determine the importance of the naphthyl ring packing, we solved the structure of RA34 complexed with a smaller substrate lacking

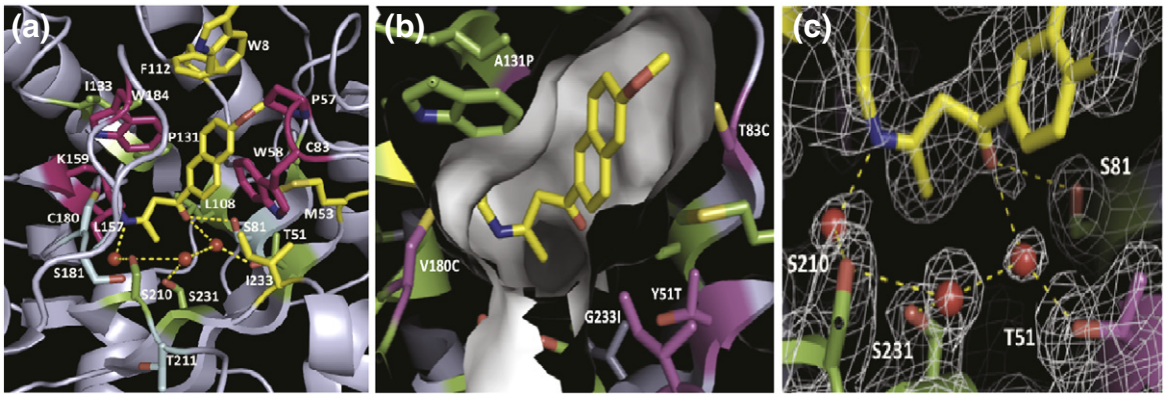

Fig. 4. Contributions of active-site residues to catalysis. (a) RA34.6 active-site alanine scanning results. Residues that decrease activity by more than 8-fold, 4-fold, 2-fold, and less than 2-fold when mutated to alanine are shown in red, yellow, green, and blue, respectively. The ligand is shown in yellow. (b) The RA34.6 binding pocket. Green side chains are derived from the original design, and purple side chains are those changed during the optimization procedure. (c) The electron density map of the active site of RA34.6 with bound substrate shows the three putative waters and the interacting residues around the active site.

the naphthyl ring (4-hydroxy-4-methyl-2-pentanone) to 2.4 Å resolution. In a prior study of designed retro-aldolase RA61, this compound showed substantially reduced activity as a substrate, consistent with the important role of the hydrophobic interactions of the enzyme with the naphthyl ring.[19] The electron density emanating from the catalytic lysine is again consistent with the size of the ligand, which is, in this case, considerably smaller than the unidentified compound present in the apostructure. The electron density shows multiple conformations of the smaller substrate analog bound to the catalytic lysine within the active site; the substrate without the naphthyl group is less well ordered than the naphthyl-containing substrate (Fig. 2c). This observation provides further indication that interactions between the designed active-site pocket and the naphthyl ring help position the substrate for catalysis.

Active-site mutagenesis

To probe the roles of individual side chains in these four active-site features, we carried out a series of mutagenesis studies around the active site (Table 1).

First, we mutated all the active-site residues in RA34.6 to alanine (Fig. 4a and Table 1, Ala mutation scan). Mutation of the catalytic lysine to alanine reduces activity by $>10^4$-fold, consistent with the critical role of the residue in Schiff base formation. Besides the crucial lysine, the largest decreases in activity upon alanine mutation were observed for large nonpolar side chains that pack against the naphthyl ring and the catalytic lysine in the structure. These interactions contribute to both substrate binding and positioning of the substrate and the catalytic lysine in orientations appropriate for catalysis, consistent with the crystal structure observations.

Second, we mutated residues that were changed during the optimization back to their identities in the original design to further understand how the active-site optimization increased activity (Table 1, reversion mutants). Reverting mutations P131A, G233I, and T83C, which pack around the naphthyl ring of the substrate (Fig. 4b), decreased k_{cat}/K_m by 3-fold, 13-fold, and 5-fold, respectively. Reversion of L157I and V180C, which surround the catalytic lysine (Fig. 4b), reduced activity by 2-fold to 3-fold. Although the active-site changes near the lysine could, in principle, affect activity by perturbing the lysine pK_a, RA34 and RA34.6 have similar pK_a values (7.2–7.3) under subsaturating conditions (Fig. S2), and given the Brønsted slope for the reaction, small changes in pK_a are not expected to have large impacts on reactivity.[19] Therefore, it is likely that the increased packing around the catalytic lysine increases activity by properly positioning the amine group for catalysis.

We also carried out single and combined mutations to evaluate the importance of the putative network of water molecules and interacting side chains (Fig. 4c and Table 1, water network mutants). Individual mutation of these side chains to alanine (T51A, S81A, S210A, and S231A) had small effects on activity (1-fold to 4-fold) (Table 1, Ala scan), but mutation of two side chains together (S210A/S231A) showed a larger effect (>50-fold) (Table 1, water network). The simultaneous mutations of the three residues to alanine and of Thr51 to valine led to a >800-fold loss in activity. Individual mutations to hydrophobic residues that can exclude water (T51L, T51I, and S81V) resulted in much larger reductions in k_{cat} (10-fold or greater) compared to Ala mutants (Table 1, water network). These results suggest that a

water network formed by these residues may make critical contributions to catalysis.

Overall, these results support the role of the designed lysine in forming the Schiff base intermediate during the reaction. They further suggest the importance of active-site side chains both in positioning the substrate within the active-site pocket and in surrounding active-site water molecules that can facilitate proton transfers in the catalytic cycle.

Discussion

The catalytic machinery of RA34.6 revealed by the crystal structures is similar in broad outline to the design model. The catalytic lysine is in a hydrophobic pocket and forms a Schiff base with the substrate, the naphthyl ring is packed by hydrophobic residues, and the hydroxyl group of the carbinolamine (not present in the inhibitor structure) is likely to interact with water molecules coordinated with surrounding residues. However, the structures also illustrate the limitations of the RA34.6 active site. For example, the naphthyl ring system is well ordered in one dimension (Fig. 2b, bottom) but appears to exhibit considerable motion in the plane of the ring (Fig. 2b, top). This structural variation suggests that the substrate is not bound as precisely as the corresponding group in native aldolase enzyme–substrate complex structures.[16,24–26] The observation of an unidentified endogenous compound in the apoprotein structure suggests that the binding pocket acts as a somewhat nonspecific hydrophobic binding chamber. The flexibility of the catalytic lysine suggested by the multiple conformers in the crystal structure is also likely to compromise catalytic activity. Finally, the water network that appears to be visible in electron density maps with the bound inhibitor differs from the modeled water molecule in the original design. In the design model, a water molecule was positioned to interact with the substrate oxygens and potentially to facilitate proton transfer chemistry; in the structure, there appears to be a more extensive group of waters that may serve a similar function. These results suggest that the identities of the side chains surrounding the water network are important but that the specific hydrogen-bonding solvent network in the design model is not recapitulated in the experimental structures. A future direction for improvement of activity would be the incorporation of direct side-chain polar interactions with the carbinolamine and the β-alcohol group.

The RA34.6 catalyst is the product of computational design followed by experimental active-site optimization. The starting computationally designed RA34 catalyst has a k_{cat}/K_m of 0.11 M^{-1} s^{-1} and a k_{cat} of 10^4-fold over the background k_{uncat}. Improvement of this activity by active-site mutagenesis reduced K_m by 20-fold and increased k_{cat} by 5-fold, for an overall 100-fold improvement of k_{cat}/K_m in RA34.6. Improved catalysts could be achieved both by improvement of methodology along the lines described in the previous paragraph and by a more extensive directed evolution. Continued efforts to develop catalysts with increased activities should provide insights into the features responsible for the remarkably high activities of native enzymes.

In summary, computational design is still at an early stage: key catalytic residues can be placed appropriately, and hydrophobic pockets that are complementary in shape to the substrate can be created, but the finer details of polar networks are likely very challenging to control precisely. Progress on this front will require more accurate models of the delicate tradeoff between favorable hydrogen-bonding interactions and cost of desolvation. A more precise recapitulation of the designed binding mode would also be facilitated by substrates with more binding "handles" (e.g., asymmetrically disposed hydrogen-bonding groups) compared to the relatively small nonpolar substrates used in our retro-aldolase and Kemp eliminase design efforts. The crystal structures and mutational analysis suggest several routes by which higher activities could potentially be achieved either by further computational design or by directed evolution—more extensive packing could better position the catalytic lysine and the naphthyl group, and replacement of some of the water-mediated interactions with general base interactions, as in native enzymes, could facilitate substrate positioning and proton transfers.

Materials and Methods

Saturation mutagenesis for active-site optimization and cell lysate preparation

Saturation mutagenesis was performed at each active-site position using Kunkel mutagenesis with degenerate oligonucleotides.[27] The resulting mutants were transformed into *Escherichia coli* BL21(DE3) cells, and individual clones were grown and screened for increased catalytic activity using a cell lysate assay. For the initial screening, we used 3-ml cultures of mutant enzymes because the activity signal was low, whereas 1-ml cultures could be used for later variants with higher activity. Cultures of the mutants were grown in 96-well plates at 37 °C in LB with 25 mg/ml kanamycin until an optical density of ~0.6 had been reached. Expression was then induced by the addition of IPTG to a concentration of 1.0 mM. The cells were then grown for 4 h at 37 °C before being pelleted by centrifugation and before we removed the supernatant media by aspiration. The pellets were frozen at −20 °C overnight. The cells were lysed with five cycles of dry ice/ethanol bath for 2 min and then with 10 °C water for 5 min. The pellet was then resuspended in 250 μl of 25 mM Hepes and 100 mM NaCl (pH 7.5) and incubated on ice for

15 min before centrifugation. One hundred eighty microliters of supernatant was transferred for activity assay.

Protein purification for steady-state kinetic assays

Proteins were expressed in BL21-star (DE3) cells using autoinduction media for 8 h at 37 °C, and then for 24 h at 18 °C. The cells were sonicated and purified over Qiagen Ni-NTA resin. After elution, the proteins were dialyzed at least three times into a 100-fold excess of buffer [25 mM Hepes and 100 mM NaCl (pH 7.5)] or by a desalting column (Sephadex G-25, HiPrep 26/10 desalting column from GE Healthcare) to remove imidazole. The proteins were then collected, flash frozen, and stored at −80 °C, or diluted for the activity assay.

Catalytic assays for cell lysates and purified proteins

For assay of cell lysates, 145 μl of cell lysate supernatant or diluted purified protein (10 μM) in buffer [25 mM Hepes and 100 mM NaCl (pH 7.5)] was pipetted into each well of 96-well black flat-bottom polystyrene nonbinding surface microplates (Corning, Lowell, MA). Then 5 μl of 10.0 mM 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone[28] in CH_3CN was added, and enzyme activity was monitored at room temperature by fluorescence in Spectramax M5^e (Molecular Devices, Sunnyvale, CA) with λ_{ex}=330 nm (9 nm bandwidth) and λ_{em}=452 nm (15 nm bandwidth), with an additional filter at 435 nm for increased noise reduction. Quartz cuvette measurements were used to verify plate measurements. The reactions were controlled for evaporation and generally stable up to 4–5 h with minimal evaporation, although most kinetic measurements were performed within the first hour.

Assays for the retro-aldol cleavage of (±)-methodol with purified protein[28] were performed at room temperature following product fluorescence for cell lysates. Measurements were performed in triplicate and averaged. The steady-state kinetic parameters k_{cat} and K_m were derived by fitting the initial rates to the Michaelis–Menten equation.

Crystallization

RA34.6 was overexpressed from *E. coli* strain BL21(RIL) as a noncleavable C-terminal 6× His-tagged protein construct using the pET-29b vector (Novagen) and was purified via affinity chromatography against Talon metal-chelate resin (Clontech).

The diketone inhibitor 1-(6-methoxy-2-naphthalenyl)-1,3-butanedione was prepared by the oxidation of methodol[3] with Dess–Martin periodinane[29] (see Supplementary Material). For the bound inhibitor structures, RA34.6 (42 μM, 1 ml) plus 50 μl of 10 mM diketone inhibitor (final concentration ~500 μM) was incubated for 30 min, then 10 μl of $Na(CN)BH_3$ (5 M in 1 M NaOH stock; final concentration ~50 mM) was added and incubated for another 30 min. The pH of the reaction was about 7.5. Afterwards, for mass spectrometry, an aliquot was taken to observe the extent of covalent modification. Finally, the rest of the sample was purified by a gel-filtration column and used to set up the crystallization tray.

For the 4-hydroxy-4-methyl-2-pentanone structure, RA34.6 (47.7 μM, 1 ml) plus 50 μl of 10 mM 4-hydroxy-4-methyl-2-pentanone (~500 μM) was incubated at room temperature for 30 min, after which 10 μl of $Na(CN)BH_3$ (5 M in 1 M NaOH stock; final concentration ~50 mM) was added and incubated for another 30 min. For mass spectrometry, an aliquot was taken to determine the extent of covalent modification (about 50% was derivatized, judging by mass spectrometry). Gel filtration was used to clean the protein sample, which was subsequently used to set up the crystal tray.

Purified proteins were each concentrated to approximately 5 mg/ml in 100 mM NaCl and 25 mM Tris (pH 7.5) by centrifugation against low-molecular-mass cutoff sieves (Centricon) and then screened under crystallization conditions using a commercial sparse matrix screen (Nextal Classic Suite; Qiagen). Crystals were grown by equilibration of the protein against a reservoir containing 2 M ammonium sulfate, 4% polyethylene glycol (PEG) 400, and 100 mM sodium acetate (pH 5.5). The crystals were frozen in a cryobuffer of 2.1 M ammonium sulfate, 5% PEG 400, 0.2 M sodium acetate, and 500 mM NaCl. For the structure with the inhibitor, the crystals were exposed to the inhibitor prior to freezing.

Molecular replacement was performed using PDB file 1a53 from the Research Collaboratory for Structural Bioinformatics database,[30] corresponding to the parental protein prior to computational redesign, with all side chains and extended peptide regions subjected to redesign deleted. The designed active site was rebuilt from the density (Fig. 1). All stages of molecular replacement, model building, and refinement were performed using programs from the CCP4 computational suite[31] (Phaser,[32] Coot,[33] and REFMAC[34]).

For RA34.6 apoprotein, crystals were grown via vapor diffusion. Equal amounts of protein at 5 mg/ml in 100 mM NaCl and 25 mM Tris-HCl (pH 7.5) were mixed with well solutions consisting of 2 M ammonium sulfate, 4% PEG 400, and 100 mM sodium acetate (pH 5.5). Data were collected from a cryoprotectant buffer (2.1 M ammonium sulfate, 5% PEG 400, 0.2 M sodium acetate, and 500 mM sodium chloride) by a Rigaku Micromax 7HF with Saturn 944+ at −180 °C. Data were indexed, integrated, and scaled using the HKL2000 package. Molecular replacement and refinement were performed with the Phaser and REFMAC modules of CCP4i. Coot was used for model building.

Subsequent incubation and cross-linking of the purified enzyme with the naphthyl–substrate analog or with the acetone product analog resulted in features of electron density that were consistent with each compound, with each being present at higher occupancy than the unidentified ligand in the unsoaked crystals.

Accession numbers

Coordinates and structure factors have been deposited in the PDB under RCSD PDB accession codes 3O6Y, 3UD6, and 3NXF.

Acknowledgements

We thank Dan Herschlag for helpful discussions and comments on an earlier version of the manu-

script, Arshiya Quadri for help with protein purification, and the Defense Advanced Research Projects Agency, the Howard Hughes Medical Institute, and the Schweizerischer Nationalfonds for funding.

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2011.10.043

References

1. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A. *et al.* (2008). *De novo* computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.
2. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J. *et al.* (2008). Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–195.
3. List, B., Barbas, C. F. & Lerner, R. A. (1998). Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis. *Proc. Natl Acad. Sci. USA*, **95**, 15351–15355.
4. Zanghellini, A., Jiang, L., Wollacott, A. M., Cheng, G., Meiler, J., Althoff, E. A. *et al.* (2006). New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794.
5. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA*, **98**, 14274–14279.
6. Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St Clair, J. L. *et al.* (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science*, **329**, 309–313.
7. Reymond, J. L. & Chen, Y. (1995). Catalytic, enantioselective aldol reaction using antibodies against a quaternary ammonium ion with a primary amine cofactor. *Tetrahedron Lett.* **36**, 2575–2578.
8. Wagner, J., Lerner, R. A. & Barbas, C. F. (1995). Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science*, **270**, 1797–1800.
9. Barbas, C. F., III, Heine, A., Zhong, G., Hoffmann, T., Gramatikova, S., Bjornestedt, R. *et al.* (1997). Immune *versus* natural selection: antibody aldolases with enzymic rates but broader scope. *Science*, **278**, 2085–2092.
10. Hilvert, D. (2000). Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751–793.
11. Hoffmann, T., Zhong, G., List, B., Shabat, D., Anderson, J., Gramatikova, S. *et al.* (1998). Aldolase antibodies of remarkable scope. *J. Am. Chem. Soc.* **120**, 2768–2779.
12. Müller, M. M., Windsor, M. A., Pomerantz, W. C., Gellman, S. H. & Hilvert, D. (2009). A rationally designed aldolase foldamer. *Angew. Chem. Int. Ed. Engl.* **48**, 922–925.
13. Johnsson, K., Allemann, R. K., Widmer, H. & Benner, S. A. (1993). Synthesis, structure and activity of artificial, rationally designed catalytic polypeptides. *Nature*, **365**, 530–532.
14. Hennig, M., Darimont, B. D., Jansonius, J. N. & Kirschner, K. (2002). The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from *Sulfolobus solfataricus* with substrate analogue, substrate, and product. *J. Mol. Biol.* **319**, 757–766.
15. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
16. Fullerton, S. W., Griffiths, J. S., Merkel, A. B., Cheriyan, M., Wymer, N. J., Hutchins, M. J. *et al.* (2006). Mechanism of the class I KDPG aldolase. *Bioorg. Med. Chem.* **14**, 3002–3010.
17. Wolfenden, R. & Snider, M. J. (2001). The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **34**, 938–945.
18. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, **50**, 4402–4410.
19. Lassila, J. K., Baker, D. & Herschlag, D. (2010). Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl Acad. Sci. USA*, **107**, 4937–4942.
20. Zhu, X., Tanaka, F., Lerner, R. A., Barbas, C. F. & Wilson, I. A. (2009). Direct observation of an enamine intermediate in amine catalysis. *J. Am. Chem. Soc.* **131**, 18206–18207.
21. Tanaka, F., Fuller, R. & Barbas, C. F. (2005). Development of small designer aldolase enzymes: catalytic activity, folding, and substrate specificity. *Biochemistry*, **44**, 7583–7592.
22. Khersonsky, O., Rothlisberger, D., Dym, O., Albeck, S., Jackson, C. J., Baker, D. & Tawfik, D. S. (2010). Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **396**, 1025–1042.
23. Khersonsky, O., Rothlisberger, D., Wollacott, A. M., Murphy, P., Dym, O., Albeck, S. *et al.* (2011). Optimization of the *in-silico*-designed Kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407**, 391–412.
24. Lorentzen, E., Siebers, B., Hensel, R. & Pohl, E. (2005). Mechanism of the Schiff base forming fructose-1,6-bisphosphate aldolase: structural analysis of reaction intermediates. *Biochemistry*, **44**, 4222–4229.
25. Heine, A., DeSantis, G., Luz, J. G., Mitchell, M., Wong, C. H. & Wilson, I. A. (2001). Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science*, **294**, 369–374.
26. Choi, K. H., Shi, J., Hopkins, C. E., Tolan, D. R. & Allen, K. N. (2001). Snapshots of catalysis: the structure of fructose-1,6-(bis)phosphate aldolase covalently bound to the substrate dihydroxyacetone phosphate. *Biochemistry*, **40**, 13868–13875.
27. Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.
28. Tanaka, F., Fuller, R., Shim, H., Lerner, R. A. & Barbas, C. F. (2004). Evolution of aldolase antibodies *in vitro*:

correlation of catalytic activity and reaction-based selection. *J. Mol. Biol.* **335**, 1007–1018.
29. Dess, D. B. & Martin, J. C. (1983). Readily accessible 12-I-5 oxidant for the conversion of primary and secondary alcohols to aldehydes and ketones. *J. Org. Chem.* **48**, 4155–4156.
30. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
31. Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 486–501.
34. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2184–2195.