



## De Novo Computational Design of Retro-Aldol Enzymes

Lin Jiang, *et al.*

*Science* **319**, 1387 (2008);

DOI: 10.1126/science.1152692

*This copy is for your personal, non-commercial use only.*

**If you wish to distribute this article to others**, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

**Permission to republish or repurpose articles or portions of articles** can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of January 7, 2011 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/319/5868/1387.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2008/03/06/319.5868.1387.DC1.html>

This article has been **cited by** 122 article(s) on the ISI Web of Science

This article has been **cited by** 39 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/319/5868/1387.full.html#related-urls>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

# De Novo Computational Design of Retro-Aldol Enzymes

Lin Jiang,<sup>1,2\*</sup> Eric A. Althoff,<sup>1\*</sup> Fernando R. Clemente,<sup>4</sup> Lindsey Doyle,<sup>5</sup> Daniela Röthlisberger,<sup>1</sup> Alexandre Zanghellini,<sup>1,2</sup> Jasmine L. Gallaher,<sup>1</sup> Jamie L. Betker,<sup>1</sup> Fujie Tanaka,<sup>6</sup> Carlos F. Barbas III,<sup>6</sup> Donald Hilvert,<sup>7</sup> Kendall N. Houk,<sup>4</sup> Barry L. Stoddard,<sup>5</sup> David Baker<sup>1,2,3,†</sup>

The creation of enzymes capable of catalyzing any desired chemical reaction is a grand challenge for computational protein design. Using new algorithms that rely on hashing techniques to construct active sites for multistep reactions, we designed retro-aldolases that use four different catalytic motifs to catalyze the breaking of a carbon-carbon bond in a nonnatural substrate. Of the 72 designs that were experimentally characterized, 32, spanning a range of protein folds, had detectable retro-aldolase activity. Designs that used an explicit water molecule to mediate proton shuffling were significantly more successful, with rate accelerations of up to four orders of magnitude and multiple turnovers, than those involving charged side-chain networks. The atomic accuracy of the design process was confirmed by the x-ray crystal structure of active designs embedded in two protein scaffolds, both of which were nearly superimposable on the design model.

Enzymes are excellent catalysts, and the ability to design new active enzymes could have applications in drug production (1), green chemistry (2), and bioremediation of xenobiotic pollutants (3). To date, most enzyme design efforts have used selection methodologies to retrieve very rare active catalysts from large libraries of candidate protein variants (4–7). Recent advances in computational protein design have made it possible to design new protein folds (8) and binding interactions (9) and have opened the door to the possibility of computationally designing enzymatic catalysts for any chemical reaction. Despite recent progress (10, 11), creating enzymes for chemical transformations not efficiently catalyzed by naturally occurring enzymes remains a major challenge. Here, we describe (i) general computational methods for constructing active sites for multistep reactions consisting of superimposed reaction intermediates and transition states (TS) surrounded by protein functional groups in orientations optimal for catalysis (Fig. 1) and (ii) the use of this methodology to design novel catalysts for a retro-aldol reaction in which a carbon-carbon bond is broken in a nonnatural (i.e., not found in biological systems) substrate: 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (Fig. 2A) (12).

The first step in the computational design of an enzyme is to define one or more potential

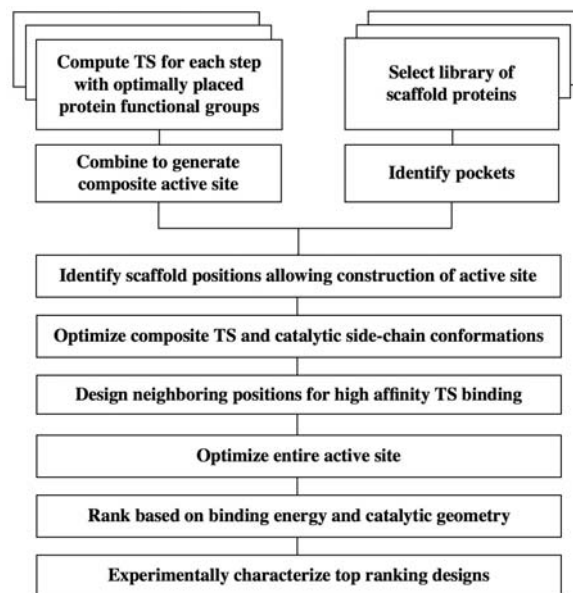
catalytic mechanisms for the desired reaction. For the retro-aldolase reaction, we focused on mechanisms involving enamine catalysis by lysine via a Schiff base or imine intermediate (13, 14). As shown in simplified form in Fig. 2B, the reaction proceeds in several distinct steps, involving acid-base catalysis by either amino acid side chains or water molecules. First, nucleophilic attack of lysine on the ketone of the sub-

strate forms a carbinolamine intermediate, which eliminates a water molecule to form the imine/iminium species. Next, carbon-carbon bond cleavage is triggered by the deprotonation of the  $\beta$ -alcohol, with the iminium acting as an electron sink. Finally, the enamine tautomerizes to an imine that is then hydrolyzed to release the covalently bound product and free the enzyme for another round of catalysis.

The second step of the design process is the identification of protein scaffolds that can accommodate the designed TS ensemble described above. To account for the multistep reaction pathway, we extended our enzyme design methodology (15) to allow the design of composite TS sites that are simultaneously compatible with multiple TS and reaction intermediates (16). Using this method, we generated design models using the four catalytic motifs shown schematically in Fig. 2C, which apply different constellations of catalytic residues to facilitate carbinolamine formation and water elimination, carbon-carbon bond cleavage, and release of bound product.

Because the probability of accurately reconstructing a given three-dimensional (3D) active site in an input protein scaffold is extremely small, it is essential to consider a very large set of active-

**Fig. 1.** Computational enzyme design protocol for a multistep reaction. The first step is to generate ensembles of models of each of the key intermediates and transition states in the reaction pathway in the context of a specific catalytic motif composed of protein functional groups. These models are then superimposed, based on the protein functional group positions, to create an initial composite active-site description. Large ensembles of distinct 3D realization of these composite active sites are then generated by simultaneously varying the degrees of freedom of the composite TS, the orientation of the catalytic side chains relative to the composite TS, and the internal conformation of the catalytic side chains. For each composite active-site description, candidate catalytic sites are generated in an input scaffold set by RosettaMatch (15). Briefly, each rotamer of each catalytic side chain is placed at each position in each scaffold, and the ensuing position of the composite TS is recorded in the hash. After the filling out of the hash table, which is linear in the numbers of scaffold positions and catalytic rotamers, the hash is searched for TS positions that are compatible with all catalytic constraints; such positions are termed “matches.” For each match, the rigid body orientation of the composite TS and the internal coordinates of the catalytic side chains are optimized to reduce steric clashes while maintaining the catalytic geometry within specified tolerances. The remaining positions (not included in the minimal catalytic site description) surrounding the docked composite TS model are redesigned to optimize TS binding affinity by means of the standard Rosetta design methodology (20, 21). The rigid body orientation of the composite TS, the side-chain torsion angles, and (in some cases) the backbone torsion angles in the active site are refined via quasi-Newton optimization (22). The resulting designs are ranked based on the total binding energy to the composite TS and the satisfaction of the specified catalytic geometry, and then the top-ranked designs are experimentally characterized. The SOM contains detailed descriptions of each step in the protocol.



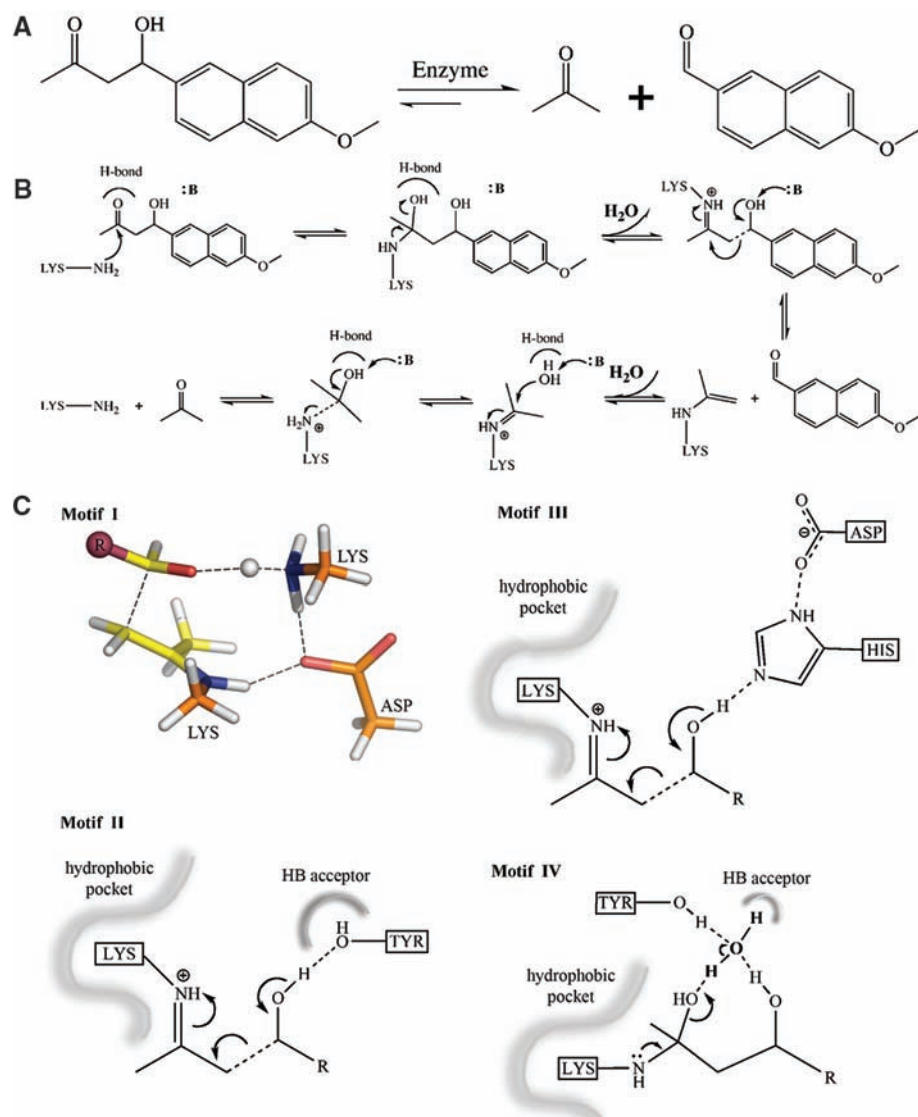
<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Biomolecular Structure and Design Program, University of Washington, Seattle, WA 98195, USA. <sup>3</sup>Howard Hughes Medical Institute (HHMI), University of Washington, Seattle, WA 98195, USA. <sup>4</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. <sup>5</sup>Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. <sup>6</sup>The Skaggs Institute for Chemical Biology and the Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. <sup>7</sup>Laboratory of Organic Chemistry, Eidgenössische Technische Hochschule (ETH) Zürich, 8093 Zürich, Switzerland.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu

site possibilities. We generated such a set by simultaneously varying (i) the internal degrees of freedom of the composite TS (fig. S1B), (ii) the orientation of the catalytic side chains with respect to the composite TS (fig. S3), within ranges that are consistent with catalysis, and

(iii) the conformations of the catalytic side chains (fig. S3). For example, in a representative calculation for motif III, we searched for placements of a total  $1.4 \times 10^{18}$  possible 3D active sites (table S3) at all triples or quadruples of backbone positions surrounding binding pockets in 71 different



**Fig. 2.** Retro-aldol reaction and active-site motifs. **(A)** The retro-aldol reaction. **(B)** General description of the aldol reaction pathway with a nucleophilic lysine and general acid-base chemistry. Several of the proton transfer steps are left out for brevity. **(C)** Active-site motifs with quantum mechanically optimized structures (23). (Top left) Motif I. Two lysines are positioned nearby one another to lower the  $pK_a$  of the nucleophilic lysine, and a Lys-Asp dyad acts as the base to deprotonate the hydroxyl group. (Bottom left) Motif II. The catalytic lysine is buried in a hydrophobic environment to lower its  $pK_a$  to make it a more potent nucleophile, and a tyrosine functions as a general acid or base. HB, hydrogen-bond. (Top right) Motif III. The catalytic lysine, analogous to motif II, is placed in a hydrophobic pocket to alter its  $pK_a$ , and a His-Asp dyad serves as a general base similar to the catalytic unit commonly observed in the serine proteases (24). (Bottom right) Motif IV. The catalytic lysine is again positioned in a hydrophobic environment. Additionally, an explicitly modeled bound water molecule is placed such that it forms a hydrogen bond with the carbinolamine hydroxyl during its formation, aids in the water elimination step, and deprotonates the  $\beta$ -alcohol at the carbon-carbon bond-breaking step. A hydrogen-bond donor/acceptor, such as Ser, Thr, or Tyr, is placed to position the water molecule in a tetrahedral geometry with the  $\beta$ -alcohol and the carbinolamine hydroxyl. The proton-abstracting ability of the water molecule is enhanced by a second hydrogen bond with a base residue. We incorporated, where possible, additional hydrogen-bonding interactions to stabilize the carbinolamine hydroxyl group and an aromatic side chain to optimally pack along the planar aromatic moiety of the substrate.

protein scaffolds (table S4). This combinatorial matching resulted in a total of 181,555 distinct solutions for the placement of the composite TS and the surrounding catalytic residues. Through extensive pruning at multiple levels, and by breaking the combinatoric explosion via hashing, the RosettaMatch algorithm (15) is able to rapidly eliminate most active-site possibilities in a given scaffold that are unfavorable as a result of poor catalytic geometry or significant steric clashes with very little computational cost. After optimization of the composite TS rigid body orientation and the identities and conformations of the surrounding residues, a total of 72 designs with 8 to 20 amino acid identity changes in 10 different scaffolds were selected for experimental characterization based on the predicted TS binding energy, the extent of satisfaction of the catalytic geometry, the packing around the active lysine, and the consistency of side-chain conformation after side-chain repacking in the presence and absence of the TS model (16). Genes encoding the designs were synthesized and the proteins were expressed and purified from *Escherichia coli*; soluble purified protein was obtained for 70 of the 72 expressed designs.

Retro-aldolase activity was monitored via a fluorescence-based assay of product formation (12) for each of the designs, and the results are summarized in Table 1. Our initial 12 designs used the first active site shown in Fig. 2C, which involves a charged side-chain (Lys-Asp-Lys)-mediated proton transfer scheme resembling that in D-2-deoxyribose-5-phosphate aldolase (13). Of these designs, two showed slow enaminone formation with 2,4-pentandione (17), which is indicative of a nucleophilic lysine, but none displayed retro-aldolase activity (16). Ten designs were made for the second, much simpler active site shown in Fig. 2C, which involves a single imine-forming lysine in a hydrophobic pocket similar to aldolase catalytic antibodies; of these designs, one formed the enaminone, but none were catalytically active. The third active site incorporates a His-Asp dyad as a general base to abstract a proton from the  $\beta$ -alcohol; of the 14 designs tested, 10 exhibited stable enaminone formation, and 8 had detectable retro-aldolase activity. In the final active site, we experimented with the explicit modeling of a water molecule, positioned via side-chain hydrogen-bonding groups, which shuttles between stabilizing the carbinolamine and abstracting the proton from the hydroxyl. Of the 36 designs tested, 20 formed the enaminone and 23 (with 11 distinct positions for the catalytic lysine) had significant retro-aldolase activity, with rate enhancements up to four orders of magnitude over the uncatalyzed reaction (18).

The active designs occur on five different protein scaffolds belonging to the triose phosphate isomerase (TIM)-barrel and jelly-roll folds. The most active designs exhibited multiple turnover kinetics; the linear progress curves for de-

signs RA60 and RA61, for example, continue unchanged for more than 20 turnovers. Progress curves [Fig. 3A and supporting online material (SOM)] show a range of kinetic behaviors: In some cases (RA45), there is a pronounced lag phase, likely associated with slow imine formation, whereas in others (RA61), there is little or no lag, and for a third set, there is an initial burst followed by a slower steady-state rate (RA22). Notably, simple linear kinetics are observed for the designs in the relatively open jelly-roll scaffold, whereas more complex kinetics are observed for the TIM-barrel designs, which have more enclosed active-site pockets that may restrict substrate access and product release. To obtain  $k_{\text{cat}}$  and  $K_{\text{M}}$  estimates for several of the best enzymes (Fig. 3B), we extracted reaction velocities from the steady-state portions of the

progress curves and assumed simple Michaelis-Menten kinetics. Given the simplifications, these values are best viewed as phenomenological; future characterization will be required to define rate constants in a particular kinetic model. The apparent  $k_{\text{cat}}$  and  $K_{\text{M}}$  values are given in Table 2;  $k_{\text{uncat}}$  was determined from measurements of the reaction progression in the absence of enzyme and is close to previously determined values (18).  $k_{\text{cat}}/k_{\text{uncat}}$  for the most active designs is  $2 \times 10^4$ . The catalytic proficiency of the designs is far from that of naturally occurring enzymes, which have a  $k_{\text{cat}}/K_{\text{M}}$  of about  $1 \text{ M}^{-1} \text{ s}^{-1}$  (Table 2); the very low  $k_{\text{cat}}$  value is probably associated with low reactivity of the imine-forming lysine. Rates for all active designs with 270  $\mu\text{M}$  substrate are reported in table S1. For each of the 11 catalytic lysine positions, a “knockout” mu-

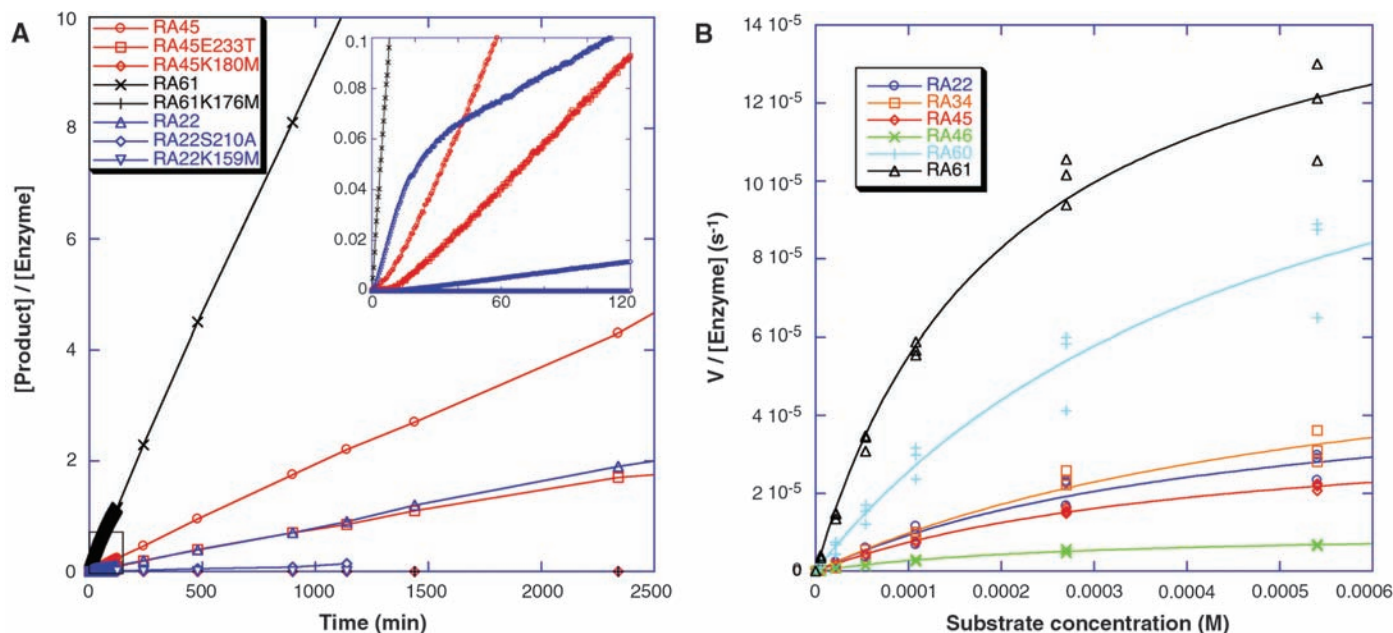
tation to methionine dramatically decreased the activity or, more commonly, abolished catalysis completely, verifying that the observed activity was due to the designed active site.

Design models for several of the most active designs with catalytic motif IV are shown in Fig. 4, A to C. Design RA60 (Fig. 4A) is on a jelly-roll scaffold, and RA45 (Fig. 4C) and RA46 (Fig. 4B) are on a TIM-barrel scaffold. The imine-forming lysine, the hydrogen-bonding residues coordinating the bridging water molecules, and the designed hydrophobic pocket (which binds the aromatic portion of the substrate) are clearly evident in all three designs.

To evaluate the accuracy of the design models, we solved the structures of two of the designs by x-ray crystallography (Fig. 4, D and E). The 2.2 Å resolution structure of the Ser<sup>210</sup>→Ala<sup>210</sup> (S210A) variant of RA22 (Fig. 4D) (19) shows that the designed catalytic residues Lys<sup>159</sup>, His<sup>233</sup>, and Asp<sup>53</sup> superimpose well on the original design model, and the remainder of the active site is nearly identical to the design. The 1.9 Å resolution structure of the M48K variant of RA61 likewise reveals an active site very close to that of the design model, with only His<sup>46</sup> and Trp<sup>178</sup> in alternative rotamer conformations, perhaps resulting from the absence of substrate in the crystal structure (Fig. 4E). Both crystal structures differ most significantly from the designs in the loops surrounding the active site; explicitly incorporating backbone flexibility in these regions

**Table 1.** Enaminone formation and enzyme activity for different active-site motifs. NC, not considered.

Motif	Catalytic lysine environment	Carbinolamine stabilization	Proton abstraction	Number tested	Number forming enaminone	Number of active designs	Rate enhancement
I	Polar	NC	Lys-Asp dyad	12	2	0	<4
II	Hydrophobic	NC	Tyr	9	1	0	<4
III	Hydrophobic	H-bond acceptor/donor	His-Asp dyad	13	10	10	$10^2$ – $10^3$
IV	Hydrophobic	Water, H-bond acceptor	Water	38	20	22	$10^3$ – $10^4$



**Fig. 3.** Experimental characterization of active enzyme designs. **(A)** Progress curves for RA61, RA61 K176M, RA22, RA22 S210A, RA22 K159M, RA45, RA45 E233T, and RA45 K180M. The enzymes were tested with 540  $\mu\text{M}$  of the racemic substrate; the reaction was followed by measuring the appearance of the fluorescent product (excitation wavelength, 330 nm; emission wavelength, 452 nm). The y axis is the concentration of product (determined from the fluorescence signal by a standard curve prepared with pure product solutions) divided by the enzyme concentration. In the design models, the serine-to-

alanine mutation in RA22 and the glutamate-to-threonine mutation in RA45 eliminate interactions that stabilize the carbinolamine intermediate and position the bound water molecule; both mutations reduce the reaction rate considerably. Mutation of the catalytic lysine residues to methionine completely eliminates enzyme activity. **(B)** Dependence of reaction velocity ( $V$ ) on substrate concentration. The rates are reported in Table 2. Reaction conditions for all experiments were 25 mM Hepes, 2.7%  $\text{CH}_3\text{CN}$ , 100 mM NaCl (pH 7.5), and substrate at the indicated concentration.

during the design process could yield improved enzymes in the future.

Each proposed catalytic mechanism can be treated as an experimentally testable hypothesis

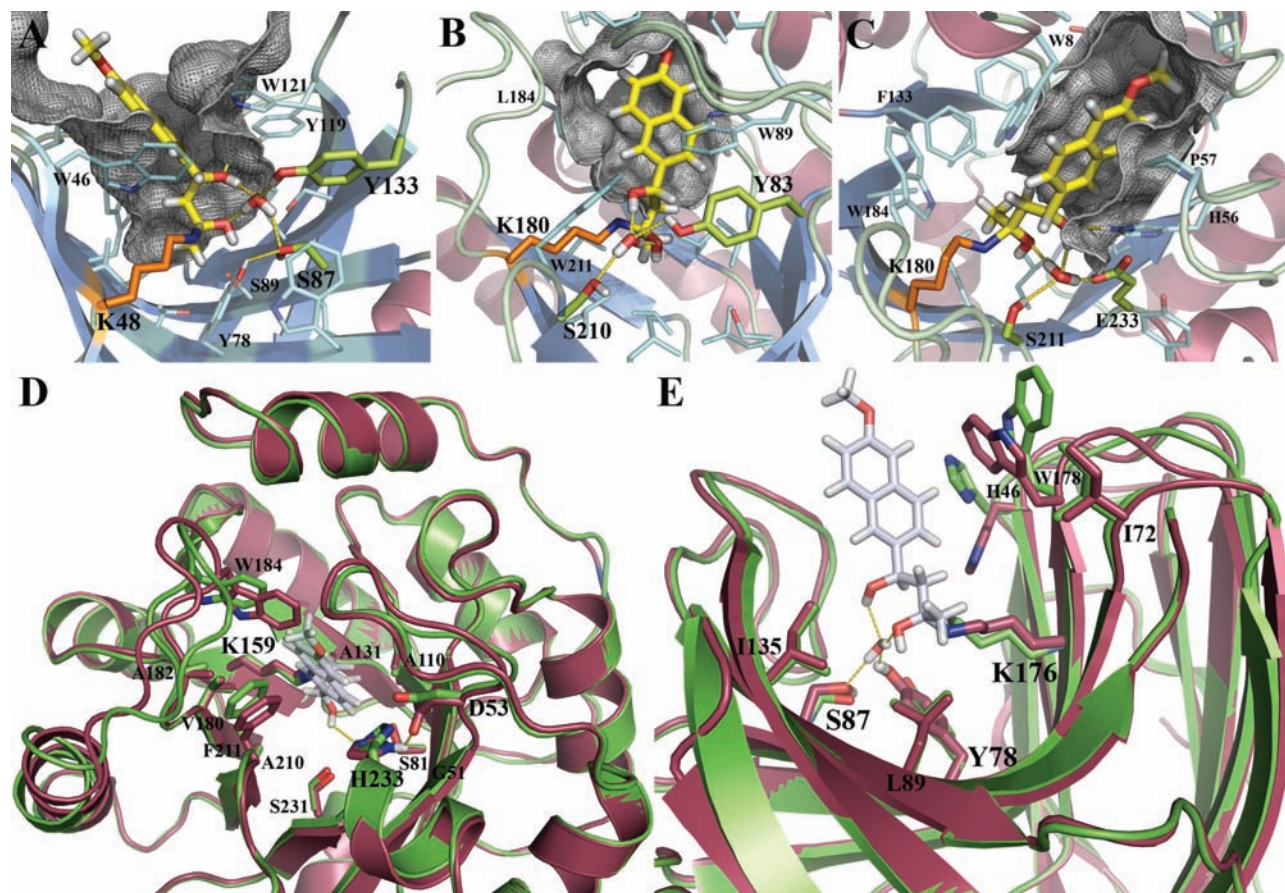
to be tested by multiple independent design experiments. Our lack of success with the first active sites that were tested contrasts markedly with our relatively high success rate with the active

site in which proton shuffling is carried out by a bound water molecule rather than by amino acid side chains acting as acid-base catalysts. The charged polar networks in highly optimized naturally occurring enzymes require exquisite control over functional group positioning and protonation states, as well as the satisfaction of the hydrogen-bonding potential of the buried polar residues, which leads to still more extended hydrogen-bond networks. Computational design of such extended polar networks is exceptionally challenging because of the difficulty of accurately computing the free energies of buried polar interactions, particularly the influence of polarizability on electrostatic free energies and the delicate balance between the cost of desolvation and the gain in favorable intraprotein electrostatic and hydrogen-bonding interactions. The sampling problem also becomes increasingly formidable

**Table 2.** Kinetic parameters of selected designs. b, burst phase; s, steady state.

Design	$k_{\text{cat}}$ ( $\times 10^{-3} \text{ min}^{-1}$ )	$K_M$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_M$ ( $\text{M}^{-1} \text{ s}^{-1}$ )	$k_{\text{cat}}/k_{\text{uncat}}^*$
RA22	$3.1 \pm 0.3$ (b)	$480 \pm 130$ (b)	$0.11 \pm 0.03$ (b)	$8.1 \times 10^3$ (b)
	$0.5 \pm 0.1$ (s)	$450 \pm 210$ (s)	$0.018 \pm 0.006$ (s)	$1.2 \times 10^3$ (s)
RA34	$4.2 \pm 1.1$ (b)	$620 \pm 180$ (b)	$0.11 \pm 0.01$ (b)	$1.1 \times 10^4$ (b)
	$0.6 \pm 0.1$ (s)	$600 \pm 140$ (s)	$0.016 \pm 0.004$ (s)	$1.5 \times 10^3$ (s)
RA45	$2.3 \pm 0.2$	$430 \pm 48$	$0.091 \pm 0.004$	$6.0 \times 10^3$
RA46	$0.62 \pm 0.5$	$290 \pm 60$	$0.037 \pm 0.006$	$1.6 \times 10^3$
RA60	$9.3 \pm 0.9$	$510 \pm 33$	$0.30 \pm 0.06$	$2.4 \times 10^4$
RA61	$9.0 \pm 1.0$	$210 \pm 50$	$0.74 \pm 0.11$	$2.3 \times 10^4$

\* $k_{\text{uncat}} = 3.9 \times 10^{-7} \text{ min}^{-1}$  (25).



**Fig. 4.** Structures of designed enzymes. (A to C) Examples of design models for active designs highlighting groups important for catalysis. The nucleophilic imine-forming lysine is in orange, the TS model is in yellow, the hydrogen-bonding groups are in light green, and the catalytic water is shown explicitly. The designed hydrophobic binding site for the aromatic portion of the TS model is indicated by the gray mesh. (A) RA60 (catalytic motif IV, jelly-roll scaffold). A designed hydrophobic pocket encloses the aromatic portion of the substrate and packs the aliphatic portion of the imine-forming Lys<sup>48</sup>. A designed hydrogen-bonding network positions the bridging water molecule and the composite TS. (B) RA46 (catalytic motif IV, TIM-barrel scaffold). Tyr<sup>83</sup> and Ser<sup>210</sup> position the bridging water molecule, which facilitates the proton shuffling required in active site IV. (C) RA45 (catalytic motif IV, TIM-barrel scaffold). The bridging water is hydrogen-bonded by Ser<sup>211</sup> and Glu<sup>233</sup>;

replacing the Glu<sup>233</sup> with Thr decreases catalytic activity threefold (Fig. 3A). (D and E) Overlay of design model (purple) on x-ray crystal structure (green). Designed amino acid side chains are shown in stick representation, and the TS model in the design is shown in gray. (D) The 2.2 Å crystal structure of the S210A variant of RA22 (catalytic motif III, TIM-barrel scaffold). The  $\alpha$  root mean square deviation (RMSD) between the design model and crystal structure is 0.62 Å, and the heavy-atom RMSD in the active site is 1.10 Å. (E) 1.8 Å crystal structure of M48K variant of RA61 (catalytic motif IV, jelly-roll scaffold). Design-crystal structure  $\alpha$  RMSD is 0.46 Å, and heavy-atom RMSD is 0.8 Å. The small differences in the high-resolution details of packing around the active site are due to slight movements in some of the loops above the binding pocket and two rotamer changes in RA61 that may reflect the absence of a TS analog in the crystal structure.

for more complex sites: The side-chain identity and conformation combinatorics dealt with by hashing in RosettaMatch become intractable for sites consisting of five or more long polar side chains, which for accurate representation may require as many as 1000 rotamer conformations each. At the other extreme, bound water molecules offer considerable versatility, because they can readily reorient to switch between acting as hydrogen-bond acceptors and donors and involve neither delicate free-energy tradeoffs nor intricate interaction networks.

It is tempting to speculate that our computationally designed enzymes resemble primordial enzymes more than they resemble highly refined modern-day enzymes. The ability to design simultaneously only three to four catalytic residues parallels the infinitesimal probability that, early in evolution, more than three to four residues would have happened to be positioned appropriately for catalysis; some of the functions played by exquisitely positioned side chains in modern enzymes may have been played by water molecules earlier in evolutionary history.

Although our results demonstrate that novel enzyme activities can be designed from scratch and indicate the catalytic strategies that are most accessible to nascent enzymes, there is still a significant gap between the activities of our designed catalysts and those of naturally occurring enzymes. Narrowing this gap presents an exciting prospect for future work: What additional features have to be incorporated into the design process to achieve catalytic activities approaching those of naturally occurring enzymes? The close agreement between the two crystal structures and the design models gives credence to our strategy of testing

hypotheses about catalytic mechanisms by generating and testing the corresponding designs; indeed, almost any idea about catalysis can be readily tested by incorporation into the computational design procedure. Determining what is missing from the current generation of designs and how it can be incorporated into a next generation of more active designed catalysts will be an exciting challenge that should unite the fields of enzymology and computational protein design in the years to come.

#### References and Notes

1. D. K. Ro *et al.*, *Nature* **440**, 940 (2006).
2. O. Kirk, T. V. Borchert, C. C. Fuglsang, *Curr. Opin. Biotechnol.* **13**, 345 (2002).
3. D. B. Janssen, I. J. Dinkla, G. J. Poelarends, P. Terpstra, *Environ. Microbiol.* **7**, 1868 (2005).
4. D. Hilvert, *Annu. Rev. Biochem.* **69**, 751 (2000).
5. B. Seelig, J. W. Szostak, *Nature* **448**, 828 (2007).
6. F. H. Arnold, A. A. Volkov, *Curr. Opin. Chem. Biol.* **3**, 54 (1999).
7. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498 (2006).
8. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
9. L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* **423**, 185 (2003).
10. D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14274 (2001).
11. J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11566 (2004).
12. F. Tanaka, R. Fuller, H. Shim, R. A. Lerner, C. F. Barbas III, *J. Mol. Biol.* **335**, 1007 (2004).
13. A. Heine *et al.*, *Science* **294**, 369 (2001).
14. S. W. Fullerton *et al.*, *Bioorg. Med. Chem.* **14**, 3002 (2006).
15. A. Zanghellini *et al.*, *Protein Sci.* **15**, 2785 (2006).
16. Materials and methods are available as supporting material on Science Online.
17. J. Wagner, R. A. Lerner, C. F. Barbas III, *Science* **270**, 1797 (1995).
18. F. Tanaka, C. F. Barbas III, *J. Am. Chem. Soc.* **124**, 3510 (2002).
19. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

20. G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, *J. Mol. Biol.* **332**, 449 (2003).
21. J. Meiler, D. Baker, *Proteins* **65**, 538 (2006).
22. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, ed. 2, 1992).
23. F. R. Clemente, K. N. Houk, *J. Am. Chem. Soc.* **127**, 11294 (2005).
24. C. T. Porter, G. J. Bartlett, J. M. Thornton, *Nucleic Acids Res.* **32**, D129 (2004).
25. G. Zhong *et al.*, *Angew. Chem. Int. Ed. Engl.* **37**, 2481 (1998).
26. Kinetic parameters of the designs reported here were determined at the University of Washington. For selected designs, the kinetic parameters were confirmed by independent experiments performed at the Scripps Research Institute. We thank R. Fuller for technical assistance. Thorough testing of the four catalytic motifs was made possible through gene synthesis by Codon Devices. We thank Rosetta@Home participants for their valuable contributions of computer time. E.A.A. is funded by a Ruth L. Kirschstein National Research Service Award. This work was supported by the Defense Advanced Research Projects Agency and HHMI. Coordinates and structure factors for the crystal structures of RA22 variant S210A and RA61 variant M48K were deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB) under the accession numbers 3B5V and 3B5L, respectively. The xyz coordinates of the designs RA22, RA34, RA45, RA46, RA60, and RA61 are included with the SOM as a zipped archive.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5868/1387/DC1

Materials and Methods

SOM Text

Figs. S1 to S8

Tables S1 to S8

References

Design Model Coordinates in PDB Format

7 November 2007; accepted 5 February 2008

10.1126/science.1152692

## A Cholesterol Biosynthesis Inhibitor Blocks *Staphylococcus aureus* Virulence

Chia-I Liu,<sup>1,2,3\*</sup> George Y. Liu,<sup>4\*</sup> Yongcheng Song,<sup>5\*</sup> Fenglin Yin,<sup>6</sup> Mary E. Hensler,<sup>7</sup> Wen-Yih Jeng,<sup>1,2</sup> Victor Nizet,<sup>7,8†</sup> Andrew H.-J. Wang,<sup>1,2,3†</sup> Eric Oldfield<sup>5,6†</sup>

*Staphylococcus aureus* produces hospital- and community-acquired infections, with methicillin-resistant *S. aureus* posing a serious public health threat. The golden carotenoid pigment of *S. aureus*, staphyloxanthin, promotes resistance to reactive oxygen species and host neutrophil-based killing, and early enzymatic steps in staphyloxanthin production resemble those for cholesterol biosynthesis. We determined the crystal structures of *S. aureus* dehydroqualene synthase (CrtM) at 1.58 angstrom resolution, finding structural similarity to human squalene synthase (SQS). We screened nine SQS inhibitors and determined the structures of three, bound to CrtM. One, previously tested for cholesterol-lowering activity in humans, blocked staphyloxanthin biosynthesis in vitro (median inhibitory concentration ~100 nM), resulting in colorless bacteria with increased susceptibility to killing by human blood and to innate immune clearance in a mouse infection model. This finding represents proof of principle for a virulence factor-based therapy against *S. aureus*.

Over the past 20 years, there has been an explosion in the prevalence of antibiotic-resistant bacterial infections, both in the

hospital and in the general community; in the United States, more deaths are now attributable to methicillin-resistant *Staphylococcus aureus*

(MRSA) infections than to HIV/AIDS (1, 2). Unfortunately, over the same time, there has been a decrease in the rate of discovery of new antibiotics, creating a pressing need for the development of novel infectious disease therapies. One approach, discussed at a recent National Research Council workshop (3), involves the specific neutralization of bacterial virulence factors to render pathogenic bacteria susceptible to innate immune system clearance. However,

<sup>1</sup>Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 11529, Taiwan. <sup>2</sup>National Core Facility of High-Throughput Protein Crystallography, Academia Sinica, Nankang, Taipei 11529, Taiwan. <sup>3</sup>Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei 10098, Taiwan. <sup>4</sup>Division of Pediatric Infectious Diseases and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA. <sup>5</sup>Department of Chemistry, University of Illinois, Urbana, IL 61801, USA. <sup>6</sup>Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801, USA. <sup>7</sup>Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA. <sup>8</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: vnizet@ucsd.edu (V.N.); ahjwang@gate.sinica.edu.tw (A.H.-J.W.); eo@chad.scs.uiuc.edu (E.O.)