



Optimization of the *In-Silico*-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution

Olga Khersonsky¹, Daniela Röthlisberger², Andrew M. Wollacott², Paul Murphy², Orly Dym³, Shira Albeck³, Gert Kiss⁴, K.N. Houk⁴, David Baker^{2,5,6*} and Dan S. Tawfik^{1*}

¹Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

²Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

³Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel

⁴Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

⁵Biomolecular Structure and Design, University of Washington, Seattle, WA 98195, USA

⁶Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

Received 30 June 2010;
accepted 19 January 2011
Available online
28 January 2011

Edited by I. Wilson

Keywords:
computational protein
design;
directed evolution;
enzymatic catalysis

Although *de novo* computational enzyme design has been shown to be feasible, the field is still in its infancy: the kinetic parameters of designed enzymes are still orders of magnitude lower than those of naturally occurring ones. Nonetheless, designed enzymes can be improved by directed evolution, as recently exemplified for the designed Kemp eliminase KE07. Random mutagenesis and screening resulted in variants with >200-fold higher catalytic efficiency and provided insights about features missing in the designed enzyme. Here we describe the optimization of KE70, another designed Kemp eliminase. Amino acid substitutions predicted to improve catalysis in design calculations involving extensive backbone sampling were individually tested. Those proven beneficial were combinatorially incorporated into the originally designed KE70 along with random mutations, and the resulting libraries were screened for improved eliminase activity. Nine rounds of mutation and selection resulted in >400-fold improvement in the catalytic efficiency of the original KE70 design, reflected in both higher k_{cat} values and lower K_{m} values, with the best variants exhibiting $k_{\text{cat}}/K_{\text{m}}$ values of $>5 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$. The optimized KE70 variants were characterized structurally and biochemically, providing insights into the origins of the improvements in catalysis. Three primary contributions were identified: first, the reshaping of the active-site cavity to achieve tighter substrate binding; second, the fine-tuning of electrostatics around the catalytic His-Asp dyad; and, third, the stabilization of the active-site dyad in a conformation optimal for catalysis.

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding authors. D. Baker, Department of Biochemistry, University of Washington, Seattle, WA 98195, USA; D. S. Tawfik, Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail addresses: dabaker@u.washington.edu; tawfik@weizmann.ac.il.

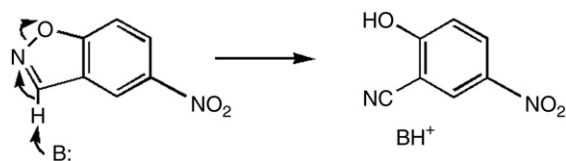
Abbreviations used: TS, transition state; PDB, Protein Data Bank; MD, molecular dynamics; Bistris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; PEG, polyethylene glycol; ESRF, European Synchrotron Radiation Facility.

Introduction

The advent of computational methods for predicting structure from sequence with atomic accuracy provides a new and powerful way of designing tailor-made active sites.^{1–4} The computational design of enzymes depends on two factors: (i) the ability to design an active-site configuration that confers efficient catalysis, and (ii) the ability to compute a sequence that confers the desired configuration. Both steps are currently performed with considerable success, but are far from optimal.⁵ Directed evolution, which requires no prior knowledge of structure–function, can be applied to improve computationally designed enzymes and can therefore complement our limited design skills. However, the bottleneck of directed evolution is the very limited sequence space that can be covered by library screening.

We have recently described a series of computationally designed enzymes that catalyze an unnatural reaction dubbed as Kemp elimination.⁶ Kemp elimination is a model reaction for proton transfer—a critical step in numerous enzymatic reactions. In this activated model system (in particular in the case of the 5-nitrobenzisoxazole substrate used here), a base-catalyzed proton transfer from carbon, concerted with the cleavage of a nitrogen–oxygen bond, leads to the cyanophenol product (Scheme 1). The mechanism of the Kemp elimination has been extensively studied, and this reaction has been used as a probe for studying medium effects on catalysis.^{7–9} Several enzyme-like systems that catalyze this reaction, including catalytic antibodies,^{10,11} synthetic polymers,¹² and promiscuous catalysis by nonenzymatic proteins,^{13,14} have been explored. In all these protein catalysts, the reaction is catalyzed by a side chain acting as a base within a hydrophobic active site. The alignment of the catalytic base relative to the substrate, medium effects that lead to the activation of base catalysts, and charge-dispersing interactions that stabilize the negatively charged transition state (TS) have all been shown to play a role.^{9,15}

The computationally designed Kemp eliminases were accordingly designed to have a base for proton abstraction aligned with the substrate C–H bond in an otherwise apolar active site. Two catalytic bases were explored—a carboxylic acid side chain (Glu, as



Scheme 1. Kemp elimination of 5-nitrobenzisoxazole.

for example, in the KE07 design,⁶ or Asp) or a histidine polarized and positioned by an adjacent aspartic or glutamic acid (His-Glu/Asp dyad), as in the KE70 design described here. The locations of these catalytic residues relative to the TS were optimized by quantum mechanical calculations. Other residues—such as a hydrogen-bond donor for the stabilization of the negative charge that develops in phenolic oxygen and an aromatic residue(s) for substrate binding and delocalization of the TS's negative charge—were included for maximal TS stabilization. The RosettaMatch algorithm was used to search for constellations of protein backbones that are capable of localizing these catalytic residues based on a large set of natural proteins with known structures. The Kemp eliminase active-site residues were then installed within these natural scaffolds by replacing 12–20 residues of the natural protein.⁶

Our previous report described the characterization and directed evolution of KE07, one of the early designs.¹⁶ The structural and mechanistic data shed light on the properties of KE07 and the routes that led to its optimization. Critical contributions from the refinement of electrostatics in the vicinity of the catalytic base, and of the active-site pocket in general, were identified. The mutations that accumulated in the directed evolution of KE07 led to a 200-fold increase in $k_{\text{cat}}/K_{\text{m}}$ and could be classified into the following groups: (a) mutations that improve the binding of the substrate and/or TS; (b) mutations that optimize the electrostatic environment and thus increase the $\text{p}K_{\text{a}}$ of the catalytic base and its catalytic efficiency; and (c) surface mutations that most probably increase protein solubility and stability.

This work describes the optimization of the KE70 design, which is one of the more advanced Kemp eliminase designs. KE70 is based on the TIM-barrel scaffold of the deoxyribose phosphate aldolase of *Escherichia coli* [Protein Data Bank (PDB) accession code 1JCL; Fig. 1a]. Unlike the glutamate side chain that acts as a base in the KE07 design, the catalytic base in KE70 is a His17-Asp45 dyad located at the bottom of the active site. Other key residues include Tyr48 (for π -stacking of the substrate and the TS) and Ser138 (designed to serve as a hydrogen-bond donor in the stabilization of the phenolic oxygen) (Fig. 1b). In addition, a number of hydrophobic residues (Ala19, Trp72, Ala103, Ile140, Val168, and Ile202) were introduced to create a tight hydrophobic pocket for substrate binding.

The catalytic proficiency of the KE70 design was nearly 10-fold higher than that of the KE07 design ($k_{\text{cat}}/K_{\text{m}}$: $\sim 80 \text{ s}^{-1} \text{ M}^{-1}$ versus $\sim 10 \text{ s}^{-1} \text{ M}^{-1}$), and the rate acceleration is relatively high for a designed enzyme or an enzyme mimic ($k_{\text{cat}}/k_{\text{uncat}} \sim 10^5$). However, the catalytic efficiency of KE70 is still far

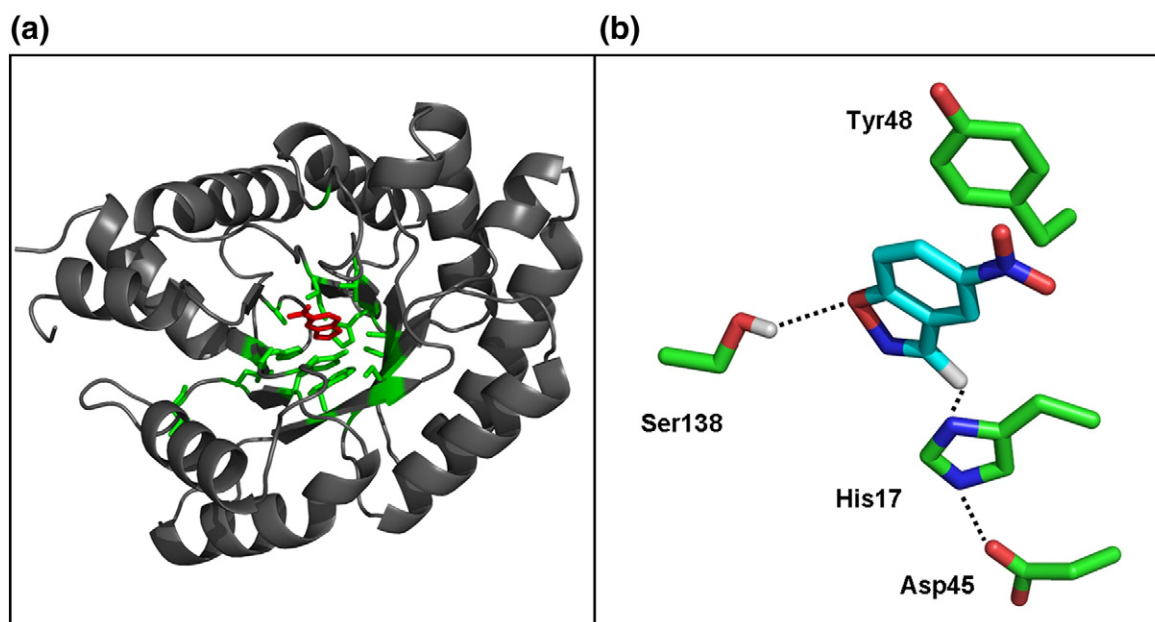


Fig. 1. (a) The KE70 design. The natural scaffold on which the design was based (PDB accession code 1JCL; gray), the modeled 5-nitrobenzisoazole substrate (red), and the side chains of 16 residues replaced to form the designed Kemp eliminase active site (green) are shown. (b) The key features of KE70's active site. The 5-nitrobenzisoazole substrate, the His17-Asp45 dyad, the H-bond donor (Ser138), and the stacking Tyr48 are shown. (Due to insertion of Ala after the initial Met to accommodate a restriction site for cloning the KE70 libraries, the numbering of the residues in this article is $n + 1$ relative to the numbering in the publication describing the original KE70 design.⁶)

below that of natural enzymes whose $k_{\text{cat}}/K_{\text{m}}$ values are, on average, in the range of $10^5 \text{ s}^{-1} \text{ M}^{-1}$ and can be as high as $10^8 \text{ s}^{-1} \text{ M}^{-1}$.¹⁷ We therefore opted for the optimization of KE70, with the ultimate goal of obtaining an enzyme with catalytic parameters comparable to the ones observed in natural enzymes. Whereas we used traditional random-mutation-based directed evolution to optimize KE07, for KE70, we included computationally guided optimization using several approaches that have not been applied in the original design. For example, to go beyond the fixed backbone approximation used in the original design calculations,⁶ we incorporated backbone flexibility into the computational optimization to search for mutations that mediate new backbone conformations conferring improved TS binding. However, rather than remaking new designs and testing them (a process that is costly and laborious), we individually tested the sequence changes suggested by the computational optimization. Those that improved activity were introduced into the evolving KE70 variants in a combinatorial manner to identify the most successful combinations of mutations. Overall, >400-fold improvement of KE70's catalytic efficiency was obtained via extensive changes in the designed active site, complemented by random mutations enriched throughout the directed evolution process.

Results

Computational optimization of the KE70 design

In the original Kemp eliminase design calculations, the backbone was kept fixed for higher computational efficiency.⁶ However, small backbone movements can considerably increase the range of possible side-chain conformations. With the goal of achieving tighter TS binding, we explored several approaches to incorporating backbone flexibility into the design calculations. We also explored alternative active-site arrangements and loop geometries, and varied the electrostatic environment of the catalytic dyad residues. As the active site of KE70 resides within the β -barrel of the protein, drastic reshaping of the backbone was not allowed, as it would disrupt the hydrogen-bond pattern between β -strands, which is critical for maintaining the TIM-barrel fold. The specific strategies that we explored are described in the text below.

Design category 1: Ensemble generation using backrub calculations

(a) A straightforward way to go beyond the fixed backbone approximation is to generate an ensemble

of slightly perturbed conformations within the neighborhood of a starting backbone structure. "Backrub" moves,¹⁸ in which the coordinates of a residue pivot around an axis defined by adjacent residues, provide an effective means to densely sample local conformational space.¹⁹ The advantage of backrub moves is that the perturbations are purely local. Since the C^α-C^β vector is perturbed, side chains emanate from the perturbed backbones in a range of orientations, so the variation in the positions of the terminal atoms can be greater than 1 Å. Indeed, backrub ensembles were found to allow improved recapitulation of protein side-chain conformations following mutations.¹⁹

We therefore subjected the original KE70 design model to Monte-Carlo-based backrub sampling, focusing on moves in the vicinity of the active site and thus generating a large ensemble of models with slightly perturbed active-site backbone geometries (Supplementary Fig. 1). For each of these, design calculations were carried out to identify amino acid sequences with particularly favorable TS binding energies. These calculations identified a number of mutations and combinations of mutations that have been predicted to improve TS stabilization: Ala19Thr/Ser, Tyr48Trp_Ser74Ala_Ser138Ala, Trp72Cys_Gly101Ala, Ser138Ala_His166Tyr, Arg70Ser, Tyr48Ala_Ser74Phe_Ala103Val, and Ser138Trp_His166Gly_Val168Ser. Since it is not certain that the backbone variation underlying each design could be realized by the predicted sequence changes, each mutant combination was tested before combining it with other sequence changes elsewhere in the protein. Experimental testing of the mutations indicated that the activities of the double mutants Trp72Cys_Gly101Ala and Ser138Ala_His166Tyr were 2.6-fold and 1.9-fold higher than that of the KE70 design, respectively. The Trp72Cys mutation is particularly interesting because it changes the packing environment of the catalytic His17 and may lead to a better positioning for proton abstraction of the substrate. Mutations of Trp72, Gly101, Ser138, and His166 were therefore introduced into the KE70 library in Round 3 of directed evolution (Table 1, Library 1).

(b) To search more broadly for active-site configurations that are related to, but differ from, the original KE70 design, we repeated the active-site search on the same scaffold (PDB accession code 1JCL) using RosettaMatch.²¹ In a subset of these calculations, we relaxed the requirement for a hydrogen-bonding group in the negatively charged phenolic oxygen in the TS, which resulted in a larger set of possible active-site designs than in the earlier calculations. This followed the observation that both in KE07 and in the initial KE70 design, this hydrogen-bond interaction did not promote higher rate accelerations. The catalytic geometry of each of the identified matches was optimized by gradient-

based minimization, and the remaining side chains of the active site were redesigned to maximize favorable interactions with the TS. To further sample in the vicinity of the best solutions obtained at this stage, we subjected the lowest-energy design variants to backrub Monte Carlo simulations and we redesigned the side chains at the active site in each of the resulting backbone conformations. Despite the additional sampling, the designs with the best catalytic dyad geometry and the most favorable TS binding geometry all retained the same catalytic dyad as KE70 (His17-Asp45), which appears to be the best solution in terms of the catalytic geometry and the TS binding energy for this scaffold. Eight of these optimized variants were selected for construction and expression (Supplementary Tables 1 and 2). Perhaps because the structural variation introduced by the backrub sampling was actually not realized in most cases, only the KE113 design was as active as the original KE70 design. Mutations of KE113 relative to KE70, and certain variations of these mutations (denoted with a slash) (Ser74Ala/Gly, Phe77Tyr, Leu136Trp, Lys173Thr/Asn, Ala178Ser, Ala231Ser, Ala238Ser, and Ser239Thr/Asn/His/Arg) were incorporated into the libraries in Rounds 5 and 6 of directed evolution (Table 1).

Design category 2: β -Strand perturbations

To fine-tune the positioning of the catalytic residues, we explored a remodeling of the conformation of the β -strands supporting these residues and the overall active-site configuration by systematically increasing and decreasing the size of the residues on the back side of the β -strands. The starting point for these computations was the double mutant Tyr48Phe_Phe77Tyr, suggested by the initial backrub calculations, which was ~ 1.7 -fold more active than the KE70 design (the Tyr48Phe mutation was also identified in the first round of directed evolution, and the adjacent Phe77 was changed to the more hydrophilic tyrosine, since the hydroxyl group is solvent exposed). The side chains adjacent to His17 and Tyr48 (which interact with the side chains from the surrounding α -helices) were either elongated or shortened to push the β -strand on which the side chain resides into the active site or to pull it into the protein core, respectively. All the tested second-shell mutations around Phe48 (Ile47-Val/Leu and Ile49Val/Leu) decreased the activity, indicating that the position of the aromatic side chain is likely to be optimal. In contrast, some of the mutations decreasing the side-chain size of residue 16 (Met16Ala/Val/Leu) or modifying the side-chain size of Leu18 (Leu18Val/Phe) were found to increase activity. Specifically, the mutations Met16Val and Leu18Phe increased the activity by ~ 2.9 -fold and 2.6-fold, respectively. Mutations of Met16 and Leu18 were therefore introduced into the libraries of

Table 1. Summary of the directed evolution of KE70

Round	Random mutagenesis	Recombination	Mutation spiking by ISOR ²⁰	Fold improvement measured with crude lysates ^a
1	2±1 random mutations per gene	—	—	≤6-fold relative to designed KE70; best variant is R1 8/9C
2	—	Shuffling of the 15 best variants from Round 1	—	≤1.5-fold relative to R1 8/9C; best variant is R2 7/12F (see Table 2)
3	—	—	Shuffling of the 15 best variants from Round 2, with incorporation of designed mutations: Library 1 (design categories 1a + 2)—Met16Ile/Leu/Val/Phe, Leu18Ile/Leu/Val/Phe, Trp72Ser/Cys/His/Leu, Gly101Glu/Gln/Ala/Ser, Ser138Ala, His166Tyr/Asp/Asn/Ala/Ser; Library 2 (design category 4a)—insertions after the residues Thr20 (Gly/Ser), Asn22 (Gly/Ser), Thr171 (Asn/Ala/Pro/Gly/Ser), Val204 (Ala/Pro/Gly/Ser), and Ser239 (Asn/Ala/Pro/Gly/Ser)	≤3-fold relative to R2 7/12F; best variants are R3 2/6D (Library 1) and R3 9/3B (Library 2)
4	—	Shuffling of the 18 best variants from Round 3, both from Libraries 1 and 2	—	≤3-fold relative to R3 9/3B; best variants are R4 4/1B and R4 4/5B (see Table 2)
5	—	—	Shuffling of the 12 best variants from Round 4, with incorporation of designed mutations: design category 1b—Ser74Ala/Gly, Phe77Tyr, Leu136Trp, Ala178Ser, Lys173Asn/Thr, Ala231Ser, Ala238Ser, Ser239Thr/Asn/His/Arg; design category 4a—Ala21Asn/Gln/Arg, Asn22Gln/Arg	≤3-fold relative to R4 4/1B; best variant is R5 7/4A (see Table 2)
6	—	Shuffling of the 10 best variants from Round 5	—	≤1.2-fold relative to R5 7/4A; best variants are R6 6/10A and R6 4/8B (see Table 2)
7	4 best variants from Round 5 and 4 variants from Round 6; 2±1 mutations per gene	—	—	≤1.2-fold relative to R6 6/10A; best variant is R7 7/1C
8	—	—	Shuffling of the 14 best variants of Round 7, with the incorporation of designed mutations: design category 2—Met16Ala/Val, Leu18Ile/Val/Leu/Phe; design category 4b—Ala238Met, Ser240Gly, Leu241Ala	≤1.2-fold relative to R7 4/2E and 3/2B; best variants are R8 12/12B and R8 15/11E (see Table 2)
9	22 best variants from Round 8; 2±1 mutations per gene	—	—	≤1.2-fold relative to R8 12/12B and R8 15/11E

^a The activity improvement measured in crude lysates is not corrected for protein expression and is therefore only a preliminary measure for an increase in protein activity.

Rounds 3 and 8 (Table 1, Library 1), and the mutation Phe77Tyr was included in Round 5 (Table 1).

Design category 3: pK_a modulation

To tune the pK_a of the catalytic His17-Asp45 dyad, we mutated the adjacent positively charged residues to either neutral, polar, or negatively charged amino acids, and we mutated the adjacent hydrophobic/polar residues to negatively charged amino acids. Mutations Lys14Glu, Arg70Leu, Leu136Asp, and Asn134Asp were tested; however, with the exception of mutation Lys14Glu, which increased the activity ~1.4-fold, mutations that either removed a

positive charge or introduced a negative charge in the vicinity of the catalytic dyad were found to slightly decrease the activity or had no effect.

Design category 4: Loop redesign

Changes in loop length (insertions or deletions) are often associated with the evolution of new enzymatic activities in nature.²² We experimented with redesigning active-site loops to increase the catalytic efficiency of KE70.

(a) Single amino acid insertions in four different loops were examined: loops formed by residues 20–27 (insertion of Gly/Ser after Thr20 and after

Table 2. Summary of sequence, kinetic, and structural data of representative KE70 variants

Variant	Kinetic parameters (5-nitrobenzisoxazole) ^a k_{cat} [s^{-1}] K_m [mM] k_{cat}/K_m [$s^{-1} M^{-1}$] Fold improvement in k_{cat}/K_m relative to the designed KE70	Mutations incorporated by error-prone PCR (designed and active active-site residues in boldface) ^b	β -Strand optimization by backrub protocol (design category 1a)	Insertions in loops (design category 4a)	Additional design optimizations (design categories 1b and 4b)	pK_a (k_{cat})	pK_a (k_{cat}/K_m)	Structure: resolution [\AA] (number of molecules in asymmetric unit)
Designed KE70	0.14 ± 0.01^c 1.11 ± 0.09 126 ± 4 1	—	—	—	—	6.3 ± 0.3 (basic), 6.6 ± 0.3 (acidic)	5.4 ± 0.1 (basic), 7.4 ± 0.1 (acidic)	2.25 (2)
R2 2/7E	0.23 ± 0.01 0.40 ± 0.04 570 ± 50 4.5	Tyr48Phe , Asp212Glu	—	—	—	—	—	2.15 (2)
R2 3/5G	0.229 ± 0.004 0.21 ± 0.03 1100 ± 140 8.7	Tyr48Phe , Val176Glu Ala204Val	—	—	—	—	—	2.02 (2)
R2 7/12F	0.316 ± 0.005 0.237 ± 0.006 1330 ± 13 10.5	Asp23Gly, Tyr48Phe Asp212Glu, His251Tyr	—	—	—	5.5 ± 0.1	6.2 ± 0.1	1.80 (2)
R4 4/1B	1.66 ± 0.02 0.18 ± 0.01 9240 ± 560 73	Lys29Asn, Thr43Asn Tyr48Phe , Ala204Val	Trp72Cys Ser138Ala	Ser20a Pro171a Ala239a	—	—	—	—
R4 4/5B	1.32 ± 0.05 0.16 ± 0.01 8000 ± 320 63	Lys29Asn, Thr43Asn Tyr48Phe , Ala204Val	Trp72Cys Gly101Ser His166Asn	Ser20a Ala204a Ser239a	—	—	—	—

R4 8/5A	0.79±0.02 0.18±0.01 4380±350 35	Lys29Asn, Tyr48Phe Lys197Asn, Ala204Val	Trp72Cys His166Tyr	Ser20a Gly239a	—	5.9±0.1	5.9±0.2	1.40 (1)
R5 7/4A	5.38±0.55 0.14±0.01 37,800±1000 300	Lys29Asn, Thr43Asn Tyr48Phe , Val176Glu Ala204Val	Trp72Cys Gly101Ser Ser138Ala His166Asn	Pro171a Ser239a	Ala231Ser			1.70 (1)
R6 4/8B	5.00±0.17 0.088±0.008	Lys29Asn, Thr43Asn Tyr48Phe , Lys197Asn, Thr198Ile, Ala204Val	Trp72Cys Gly101Ser	Ser20a Ala239a	Ser74Gly Ala178Ser			
R6 6/10A	57,300±6140 455 5.26±0.45 0.096±0.012 54,800±5370 435	Lys29Asn, Thr43Asn Tyr48Phe , Ala204Val	Ser138Ala His166Asn Trp72Cys Gly101Ser Ser138Ala His166Asn	Ser20a Ala239a	—	6.2±0.1	6.0±0.2	2.2 (2)
R8 12/12B	7.47±0.32 0.143±0.028 53,100±8180 421	Lys29Asn, Thr43Asn Tyr48Phe , Thr75Ala Lys197Asn, Thr198Ile Ala204Val	Trp72Cys Gly101Ser Ser138Ala His166Asn	Ser20a Ala239a	Ala178Ser			
R8 15/11E	5.30±0.18 0.150±0.015 34,900±2220 277	Lys29Asn, Gln36Lys Thr43Asn , Tyr48Phe Ala204Val	Trp72Cys Gly101Ser Ser138Ala His166Asn	Ser20a Ala239a	Ala178Ser Ala238Met			

^a The reported parameters are the average of at least three independent measurements, and error ranges reflect the differences between repeated determinations.

^b The numbering of the residues in this article is $n + 1$ relative to the numbering in the publication describing the original KE70 design⁶ due to insertion of Ala after the initial Met to accommodate a restriction site for cloning the KE70 libraries. Inserted loop residues are numbered as the residues after which they were inserted with the suffix "a."

^c The deviation (≤ 1.6 -fold) from the catalytic parameters previously reported for the KE70 design ($k_{\text{cat}} = 0.16 \pm 0.05 \text{ s}^{-1}$; $K_{\text{m}} = 2.1 \pm 0.8 \text{ mM}$; $k_{\text{cat}}/K_{\text{m}} = 78 \pm 14 \text{ s}^{-1} \text{ M}^{-1}$)⁶ may stem from the differences in protein production, purification, and assay temperature in assays performed in different laboratories and at different times.

Asn22); loops formed by residues 168–180 (insertion of Asn/Ala/Pro/Gly/Ser after Thr171); loops formed by residues 202–210 (insertion of Ala/Pro/Gly/Ser after Val204); and loops formed by residues 238–241 (insertion of Asn/Ala/Pro/Gly/Ser after Ser239). These positions and the inserted amino acids were chosen with the aim of increasing the flexibility of the loop, so that either the newly introduced amino acid or a polar residue within the existing loop sequence could interact with the nitro group of the substrate. These insertions were incorporated into Library 2 of Round 3 of directed evolution (Table 1). In addition, mutations Ala21Asn/Gln/Arg and Asn22Gln/Arg were incorporated into the libraries of Rounds 5 and 6 of directed evolution (Table 1) to form additional potential interactions with the substrate's nitro group.

(b) A recently developed loop design protocol,²³ which computationally introduced a predefined hydrogen-bond interaction to the nitro group of the substrate, was also applied. The length of the loop from residue Phe236 through Leu242 (seven amino acids with the sequence FGASSLL) was varied between five and eight amino acids while maintaining a hydrogen-bond interaction between either a Ser or an Asn and the nitro group of the substrate. Designs for each loop length were experimentally tested (Supplementary Table 3), and only the longest loop with the sequence FGMSAGAL was found to increase the activity. It increased the activity ~3.5-fold compared to the Tyr48Phe_Phe77Tyr mutant, and combining it with the β -strand mutation Met16Val increased the activity ~9-fold. The FGMSAGAL loop was subsequently incorporated at Round 8 of directed evolution (Table 1).

Directed evolution of KE70

Table 1 summarizes the results of nine rounds of directed evolution starting from the designed KE70. The first two rounds employed random mutagenesis and improved the activity of KE70 by ~10-fold. Several first-shell residues were mutated; in particular, the designed stacking residue Tyr48 was changed into Phe, and Ala204 was changed into Val (Table 2, Fig. 2a). These changes indicated that the KE70 design could be optimized by changes in the first-shell residues, which was not the case in the directed evolution of the KE07 design, where mutations in the second and third shells dominated, and only a single designed residue was mutated. The subsequent directed evolution rounds therefore included not only random mutagenesis but also designed mutations identified as described in [Computational optimization of the KE70 design](#). In Round 3, two areas of the protein were optimized (Table 1). Library 1 included mutations in β -strands at the bottom of the active site (design categories 1a

and 2; Fig. 2b). Library 2 included insertions in the loops at the top of the active site (design category 4a; Fig. 2c). The designed mutations were spiked in a combinatorial manner by the ISOR method,²⁰ so that each library variant contained, on average, two mutations out of the complete set of 22 mutations at six different positions for Library 1, and 18 insertions at five different positions for Library 2. Given the limited screening capacity (800–1600 variants per round), we aimed at libraries that carry only few designed mutations per gene. Therefore, designed mutations that could be readily introduced by random point mutations (e.g., Ala19Thr and Lys14-Glu) were not incorporated by synthetic oligonucleotides. In Round 4, the improved variants from Libraries 1 and 2 were shuffled to combine the mutations in the loops and β -strands. In Rounds 5 and 6, further design optimizations were explored based on design categories 1b and 4a (Table 1, Fig. 2d).

Following another round of random mutagenesis (Round 7), the improved variants were shuffled again (Round 8). In parallel, in Round 8, additional mutations, based on design optimizations around the catalytic His17 and around insertions in the active-site loops, were explored by the ISOR method (design categories 2 and 4b; Fig. 2d, Table 1). However, only mild improvements in catalytic activity were obtained in Rounds 7 and 8. Round 9 of directed evolution, in which random mutagenesis was applied on the variants from Round 8, also failed to yield any additional improvement (Table 1). The improvement in the catalysis rate of KE70 therefore appeared to have reached a plateau.

Selected mutations and activity improvements

Only a subset of the designed mutations found individually to increase activity was incorporated into the final KE70 variants with the highest activities (Table 2). This suggests that an accurate calculation of the coupling between multiple simultaneous changes in the active site is challenging for current computational methods, and our combinatorial incorporation strategy is a pragmatic solution to this yet unsolved problem. On the assumption that there was no bias in the incorporation method (as indicated by the sequencing of the unselected libraries), the mutations that were not found in the selected KE70 variants were probably not beneficial or even deleterious in the context other mutations that accumulated in the evolved variants. For example, mutations of the residues adjacent to the catalytic His17 (Met16 and Leu18, design category 2) were not incorporated, although they were beneficial on their own and were introduced twice in the directed evolution process (in Rounds 3 and 8). In contrast, all the designed modifications of the β -strands at the

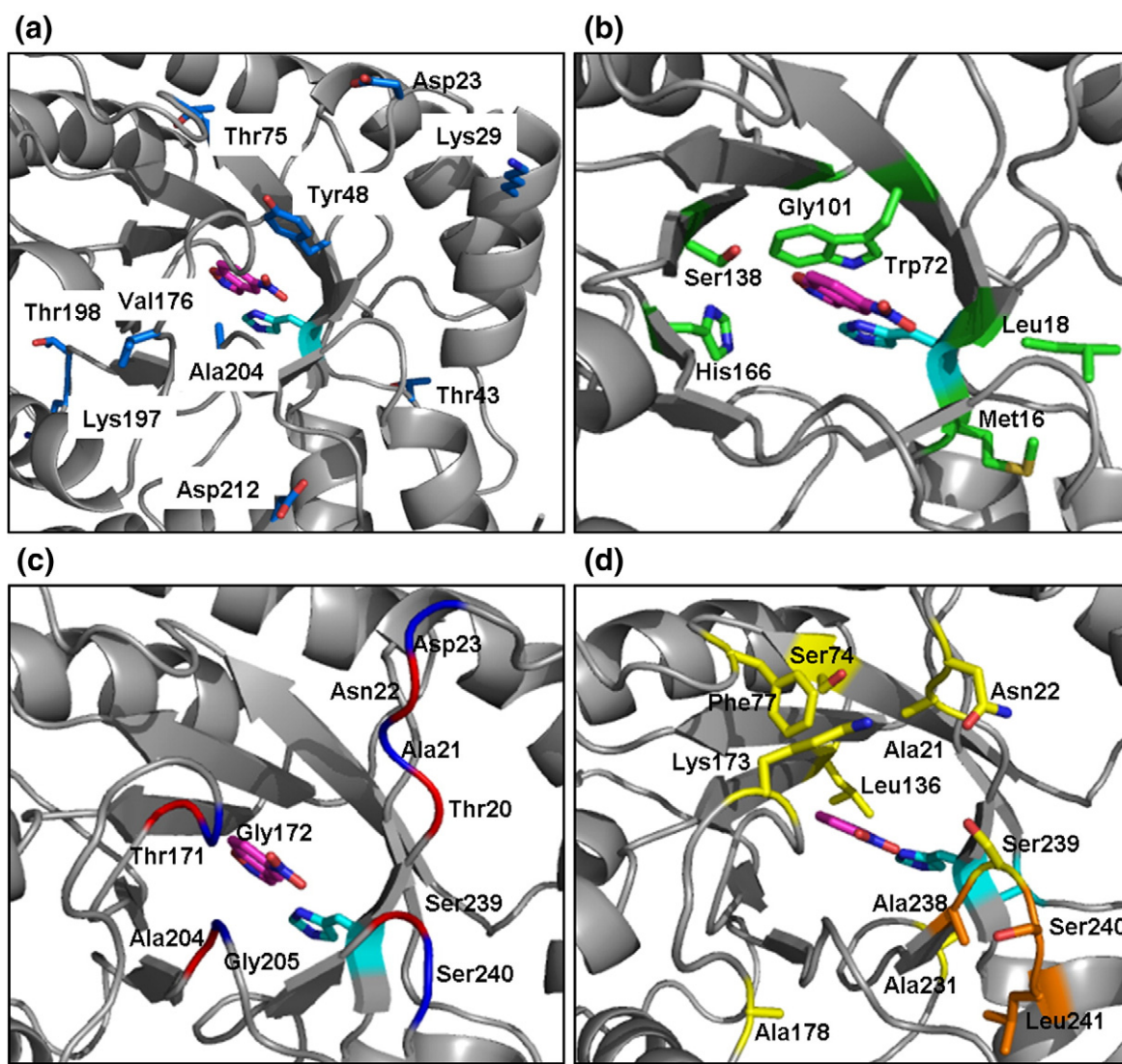


Fig. 2. Mutations in the evolved KE70 variants. The model of KE70 design, with the side chains mutated in the evolved variants, is shown. The catalytic His17 is shown in cyan, and the substrate is shown in pink. (a) Residues in which random mutations accumulated during nine rounds of directed evolution (blue). (b) Residues in β -strands at the bottom of the active site optimized by design categories 1a and 2 (green). (c) Designed insertions in the loops above the active site (design category 4; positions upstream of the insertions are shown in red, and positions downstream of the insertions are shown in blue). (d) Residues optimized in Round 5 (design categories 1b and 4a; yellow) and Round 8 (design category 4b; orange).

bottom of the active site that increased catalytic activity (Trp72Cys, Gly101Ser, Ser138Ala, and His166Asn) were ultimately incorporated. Insertions in the loops at the top of the active site were readily incorporated (Table 2, variants R4 4/1B and R4 4/5B); however, after further rounds of optimization, only two insertions were fixated (Ser20a and Ala239a). Further design optimizations (design categories 1b and 4b) resulted in only small improvements in the catalytic activity, with the key mutations being Ser74Gly, Ala178Ser, Ala231Ser, and Ala238Met (Table 2, Fig. 2d).

The random mutations in the evolved KE70 variants occurred mostly in the active site (Table 2,

Fig. 2a). Several of the designed residues and residues adjacent to the designed residues were mutated (Tyr48Phe, Thr43Asn, and Thr75Ala). Mutations in the loops adjacent to the active site, such as Asp23Gly, Val176Glu, Lys197Asn, and Thr198Ile, were as well associated with improvements in catalysis.

After six rounds of directed evolution, the catalytic proficiency of KE70 improved over 400-fold (Table 2). The turnover number (k_{cat}) increased >35-fold from 0.14 s^{-1} to $>5 \text{ s}^{-1}$, and K_m decreased >10-fold from $\sim 1 \text{ mM}$ to $<0.1 \text{ mM}$. The evolved variants carried 10–14 mutations, of which approximately half came from design optimizations

and half came from random mutagenesis (Table 2). Overall, the evolved variants differed from the natural enzyme that was used as template (PDB accession code 1JCL) by ≥ 19 positions. The catalytic machinery remained the same during the optimization process: mutating the catalytic His17 in the evolved variants from Rounds 4 and 6 decreased the activity >1000-fold, and mutating Asp45 caused an ~ 5 -fold decrease in activity (Supplementary Table 4).

Structural analysis of KE70 variants

Structures of the KE70 design and of six evolved mutants (R2 2/7E, R2 3/7G, R2 7/12F, R4 8/5A, R5 7/4A, and R6 6/10A) were solved at resolutions of 1.4–2.25 Å. These structures provide insights into the mechanism and evolution of KE70. However, despite extensive efforts, we could not obtain meaningful structures of KE70 variants with substrate or TS analogues. Benzimidazole ligands did not exhibit any inhibition. The product of the reaction (5-nitro-2-cyanophenol) was used for cocrystallization and soaking experiments, but no electron density corresponding to it was observed. Benzotriazoles did inhibit some of the evolved variants (with K_i values in the range of 0.05–60 μM), and a structure of the KE70 variant R6 6/10A with 5-nitrobenzotriazole ($K_i = 0.43 \pm 0.05 \mu\text{M}$) was obtained at a resolution of 2.2 Å. The ligand was found to be located in the active site and is aligned against the stacking Phe48. The key active-site residues, such as the catalytic His17-Asp45 dyad and the stacking Phe48, did not change their conformations (Supplementary Fig. 2a), nor did other active-site residues, with the exception of Ser101, which binds the nitro group of the benzotriazole (Supplementary Fig. 2b). The nitro group of the ligand, however, points towards His17, whereas the triazole nitrogens face the top of the active site (Supplementary Fig. 2a). Since both His17 and benzotriazole are protonated at the pH of crystallization (pH 4), this binding mode is somewhat expected and catalytically irrelevant. However, attempts to crystallize this complex at pH >8, namely with His17 in the catalytic deprotonated form, gave only poorly diffracting crystals.

The structure of the KE70 variant R4 8/5A contained additional electron density in the active site, which was modeled as benzamidine (Supplementary Fig. 3a and b). However, benzamidine showed no inhibition of eliminase activity ($\leq 5 \text{ mM}$ benzamidine), and no binding was observed in isothermal calorimetry measurements ($\leq 10 \text{ mM}$). No inhibition of eliminase activity was observed also with molecules similar to benzamidine, such as benzoic acid, nitrobenzene, and benzamide. In any case, as is the case with 5-nitrobenzotriazole, the presence of this unknown ligand did not cause any

significant conformational changes in the active site (Supplementary Fig. 3c).

Two structures of the original KE70 design were obtained: one with a resolution of 2.25 Å, and the other with a resolution of 1.4 Å. Both structures of KE70 design are similar overall to the computed model (Supplementary Table 5 and Supplementary Fig. 4). However, in the structure of the KE70 design with 1.4 Å resolution, there was a shift of one β -strand and a helix adjacent to it (residues 229–238) in one of the monomers of the asymmetric unit (monomer A). The side chains of the residues on the shifted β -strand that point into the active-site tunnel in the designed model were relocated into the protein core in monomer A, and vice versa. In particular, the side chain of Phe235 in the KE70 design structure points into the active site and may thus interfere with substrate binding (Fig. 3a; Supplementary Fig. 5). In the KE70 design structure with 2.25 Å resolution, both monomers had a shifted β -strand. Interestingly, in the structures of the evolved variants, all the β -strand residues align perfectly with those in the KE70 model, and the side chain of Phe235 points into the protein core as in the computed model (Fig. 3b). Since the original design is active, the conformation observed in its crystal structure could be an artifact of crystallization, even though the shifted β -strand is not on the protein surface. In any event, mutations incorporated in the early rounds of directed evolution seem to have eliminated the shifted and presumably nonactive conformation.

The largest deviations between the various structures are in the loops at the top of the barrel. This structural flexibility may explain why many insertions in the loops were tolerated. One of the loops (residues 21–27) is disordered, and electron density is observed only in one of the structures (variant R2 7/12F). Some loop movements appear to cause shifts in the positions of certain helices, but the β -strands overlap in all structures (with the exception of the strand noted in Fig. 3). The position of the catalytic His17-Asp45 dyad is essentially the same in all structures, and the distances between His17 and Asp45 (3–4 Å) vary by less than 1 Å. However, as indicated by the molecular dynamics (MD) simulations discussed below, these subtle changes may relate to the improved rates. Although the catalytic dyad remained largely unchanged, the shape of the active-site cavity changed significantly. Indeed, the active-site tunnel of variant R6 6/10A, bearing two insertions in the loops and a modified combination of amino acids at the bottom of the active site, is much deeper than in the original design. This is primarily caused by the mutation of Trp72 into the much smaller cysteine, which opened the central tunnel all the way to the other side of the protein (Fig. 4). As a result of the designed modifications at the bottom of the active site, the substrate is more

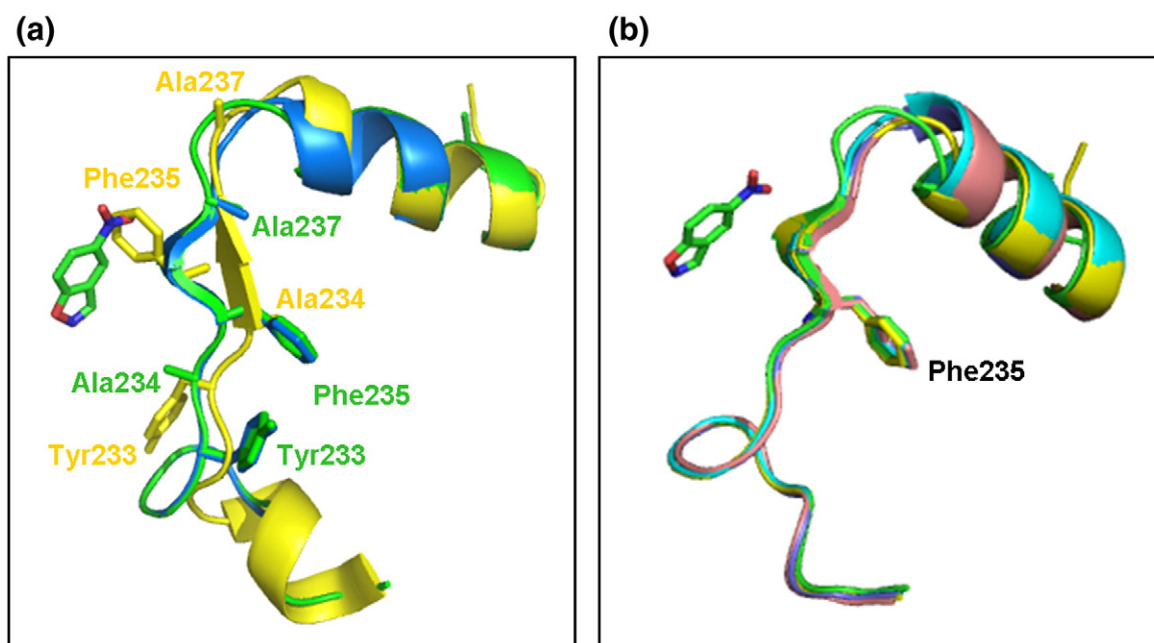


Fig. 3. (a) Strand register shift involving residues 229–238 in the crystal structure of KE70 design monomer A (yellow) relative to monomer B (blue) and to the KE70 design model (green). The region of residues 225–251 is shown with key residues as sticks and with the modeled substrate in green. (b) In the structures of the evolved KE70 variants, the region of residues 229–238 aligns well with the KE70 model (green, KE70 model; yellow, R2 7/12F; cyan, R4 8/5A; pink, R5 7/4A; blue, R6 6/10A).

tightly and firmly packed in the evolved variants, as indicated by the lower K_m values and by the MD simulations discussed below.

Although the designed changes related to the catalytic dyad were not incorporated (design categories 2 and 3), there are significant changes in the residues around the His17-Asp45 dyad and in the associated hydrogen-bond networks. In the design model, Arg69 points away from the catalytic dyad, and the distance between Arg69 and Asp45 is 7.1 Å (Fig. 5a). However, in the crystal structure of the KE70 design, Arg69 points towards the catalytic dyad, and the Arg69-Asp45 distance (2.8–3.1 Å) corresponds to a salt bridge (Fig. 5b). The Arg69-His17 distance was also reduced from 10.5 Å in the KE70 model to 3.1–4.0 Å in the crystal structure. The proximity of Arg69 may lower the basicity of the dyad and, therefore, the catalytic potential of both Asp45 and His17. In the Round 2 variants, Arg69 and Asp45 are still within salt-bridge formation, but Arg69 is detached from the catalytic His17 (>5.9 Å). Interestingly, in the Round 2 7/12F variant, Arg69 adopts different conformations in the two monomers of the asymmetric unit. One conformer is similar to the computed model, and the other is similar to the crystal structure of the designed KE70 (Fig. 5c). In the Round 6 variant 6/10A, the mutation Gly101Ser caused Arg69 to move away from the catalytic dyad, and the distance between

Arg69 and Asp45 increased to 6.2 Å, probably rendering the His17-Asp45 dyad more basic and therefore more active (Fig. 5d).

pH-rate profiles of KE70 variants

The environment of the catalytic dyad was also probed by determining the pH-rate profiles of KE70 design and its evolved variants. The pH-rate profile of the KE70 design is bell shape like, although the rates deviate from a simple bell-shaped fit at pH >7.5 (Fig. 6). The basic $pK_a(k_{cat})$ of KE70 design is 6.3, and the basic $pK_a(k_{cat}/K_m)$ is much lower at 5.3 (Fig. 6, Table 2). The decrease in activity at higher pH suggests that, in addition to His17, which is active in the basic form, there is another residue(s) with an acidic pK_a in the range of 6–7 [$pK_a(k_{cat}) \sim 6.6$; $pK_a(k_{cat}/K_m) \sim 7.4$] whose deprotonation reduces the activity of the KE70 design. The origins of the decrease at higher pH are unclear, since both catalytic residues His17 and Asp45 are supposed to be active in their deprotonated forms, but may be related to the strand register shift observed in the crystal structure of the original design (Fig. 3a). This is consistent with the absence of the acid shoulder in the pH-rate profiles of the evolved variants, in which the strand is correctly placed (Fig. 3b). The pH-rate profiles of the evolved variants R2 7/12F, R4 8/5A, and R6 6/10A

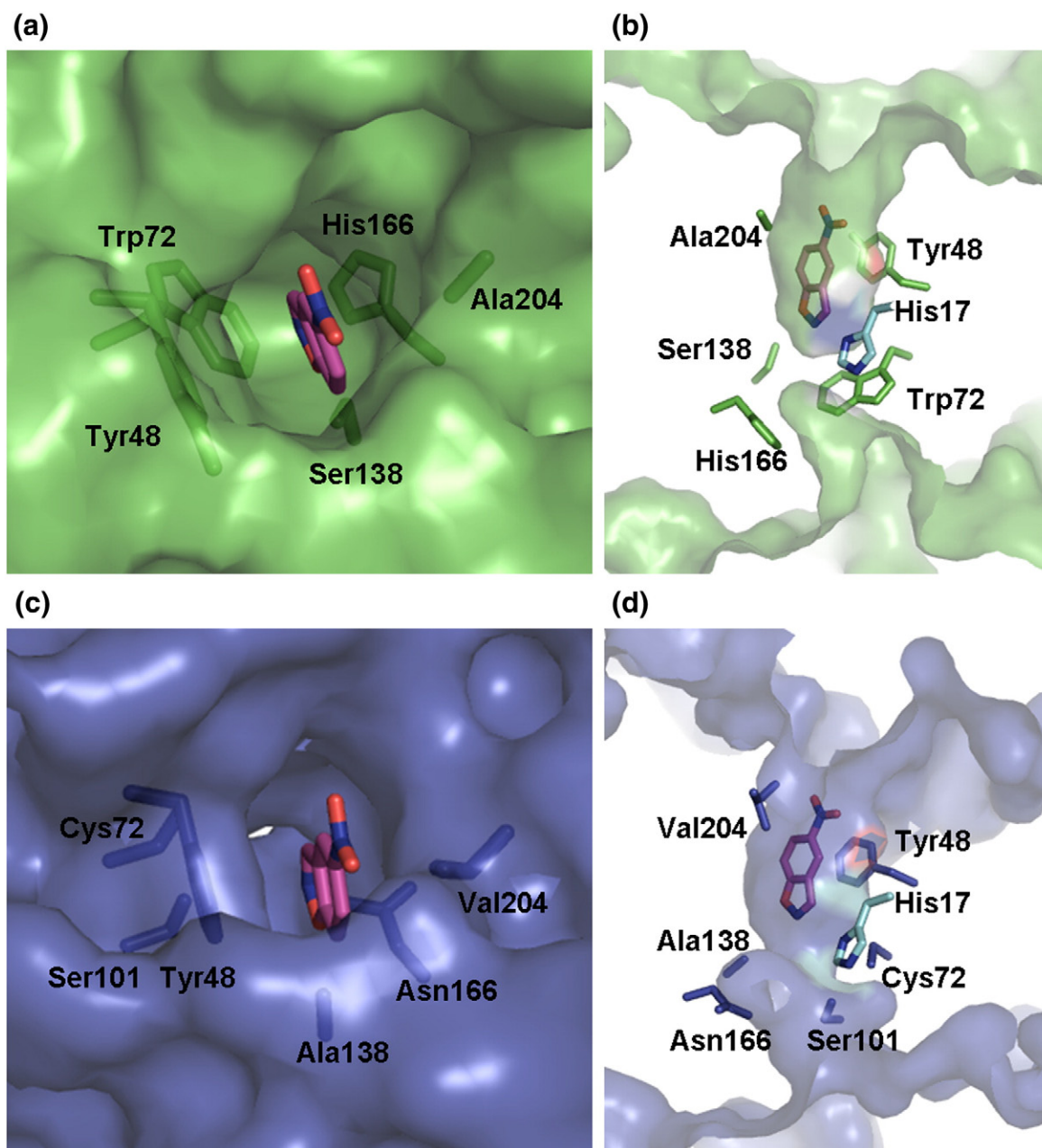


Fig. 4. The active-site cavity of the KE70 designed model (a, top view; b, cross-sectional view) and of the evolved variant R6 6/10A (c, top view; d, cross-sectional view). The substrate (magenta) was overlaid from the designed model. The active-site cavity of variant R6 6/10A is deeper than that of the original KE70 design, primarily due to the mutation Trp72Cys at the bottom of the active site. The central tunnel is also opened all the way to the other side of the protein. In addition, the entrance to the active site is narrower in the evolved variant, primarily due to mutation Ala204Val.

exhibit only one basic shoulder and could therefore be compared (Fig. 6). Consistently with their higher activities, the $pK_a(k_{cat})$ values of the evolved KE70 variants, indicating the pK_a of the enzyme–substrate complexes, increased with the progress of directed evolution from 5.5 to 6.2 (Fig. 6, Table 2). In contrast, the $pK_a(k_{cat}/K_m)$ values of the evolved KE70 variants, indicating the pK_a of the free enzyme, are similar and range between

5.9 and 6.2. The elevated $pK_a(k_{cat}/K_m)$ of variant R2 7/12F (6.2) is primarily due to a significant increase in K_m values with a decrease in pH below 6.5. This pattern suggests that changes in the environment of the catalytic dyad are primarily manifested at the level of the enzyme–substrate complex (changes in $pK_a(k_{cat})$ values) rather than at the level of the free enzyme ($pK_a(k_{cat}/K_m)$ values).

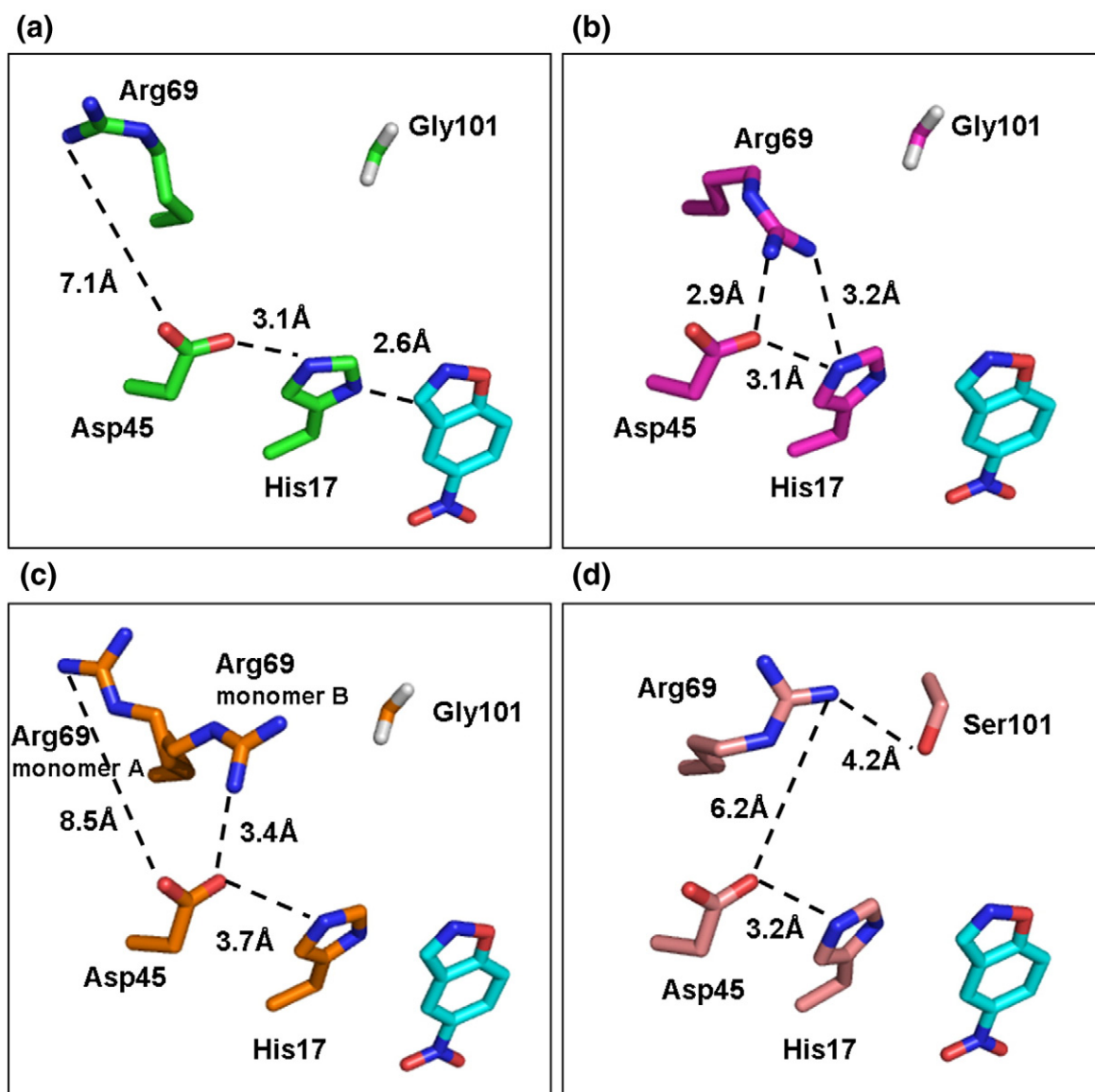


Fig. 5. Refinement of the environment of the catalytic dyad. (a) In the designed KE70 model, Arg69 points away from the catalytic His17-Asp45 dyad. (b) In the crystal structure of the KE70 design, Arg69 points towards the catalytic dyad and is within hydrogen-bond distance of both Asp45 and His17, thus reducing the basicity of His17. (c) In the structure of the evolved variant Round 2 7/12F, two different conformations of Arg69 are observed in the two molecules in the asymmetric unit. One rotamer resembles the computed model (as in (a)), and the other aligns with Arg69 from the crystal structure of the KE70 design (as in (b)). (d) In the evolved variant Round 6 6/10A, the mutation of Gly101 to serine appears to stabilize the rotamer of Arg69 that points away from the catalytic dyad. The substrate (cyan) was overlaid from the designed model.

The increase in the $pK_a(k_{cat})$ values of the catalytic His17 during evolution (~ 0.7 pKa units) is consistent with changes in the hydrogen bonding of adjacent residues, in particular with the repositioning of Arg69 (Fig. 5). These changes partly account for the increase in the k_{cat} value of the evolved KE70 variants (>30 -fold). However, as the catalytic efficiency (k_{cat}/K_m) has increased much more (>400 -fold), other factors, such as optimization of substrate binding in the active site and modification

of the active-site loops, also contributed to the increase in catalytic power.

MD simulations

MD simulations were performed on the structures of the KE70 design and its evolved variants. We utilized a protocol that had been previously established for the evaluation of general acid–base biocatalysts.²⁴ Polar contacts, distances between

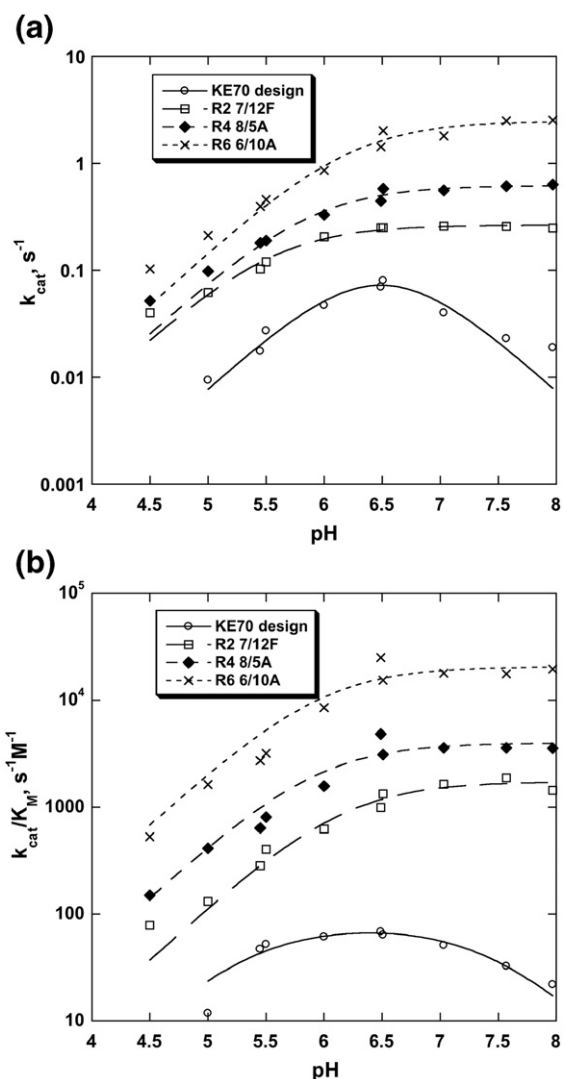


Fig. 6. pH-rate profiles of the KE70 variants. (a) k_{cat} versus pH. (b) k_{cat}/K_m versus pH. The resulting parameters are given in Table 2.

the residues of the catalytic dyad, distances between the catalytic residues and the corresponding atoms in the TS, root-mean-square deviations (RMSDs), and atomic fluctuations (equivalent to B -factors) of the active-site residues were monitored. For the purposes of MD simulations, the active site of each structure was defined to consist of the residues His17, Ala19, Asp45, Tyr/Phe48, Trp/Cys72, Ala103, Ser/Ala138, Ile140, Val168, Ser170, and Ile202.

The MD-derived backbone RMSDs decreased as KE70 was evolved (Fig. 7a). Compared to the original KE70 design, the average atomic fluctuations of the active site and also their spread (as denoted by standard deviation bars) appear lower in all the variants beyond the second round of

directed evolution (Fig. 7b). This trend implies that key active-site residues (His17, Ala19, and Asp45) that, in the original design, were more flexible than other active-site residues rigidified as the design evolved (Supplementary Fig. 6 and Supplementary Table 7). RMSD values quantify the degree to which the simulated structures deviate from a given reference (here the crystal structure), while the atomic fluctuations quantify the degree of flexibility (i.e., the variability between the runs of the same variant). As such, Fig. 7a and b and Supplementary Fig. 6 translate into the observation that each round of evolution rigidified the active site, giving rise to an increased level of preorganization.

Figure 7c and d plot $(\Delta r)^2$ versus $-\ln(k_{\text{cat}})$ and $-\ln(k_{\text{cat}}/k_m)$. Δr denotes the deviation of the measured His17-Asp45 distance (His- $\text{N}^{\text{eH}}-\text{O}^{\delta-}\text{-Asp}$) from the ideal catalytic hydrogen-bond distance (1.8 Å).²⁴ On the assumption of a simple harmonic model, the energetic cost of deviation from ideality is proportional to $(\Delta r)^2$, and on the assumption of a simple TS model, $\ln(k_{\text{cat}})$ is proportional to the activation free energy. The linear relationship between $(\Delta r)^2$ and $\ln(k_{\text{cat}})$ and $\ln(k_{\text{cat}}/k_m)$ suggests that the increased k_{cat} of evolved variants results, in part, from the tightening of the hydrogen bond between the catalytic dyad members as the active-site residues of the KE70 variants become more optimally placed and less mobile.

The distance between the catalytic base (His17) and the substrate proton that it abstracts remained unchanged during the process of KE70 optimization in the range of 2.6–2.7 Å (Supplementary Table 6). This suggests that the distance between His17 and the subtracted proton in the KE70 design need not be further optimized.

Discussion

Our previous report described the optimization of the computationally designed Kemp eliminase KE07 by directed evolution.¹⁶ There are several significant differences between the KE70 design and the KE07 design. Unlike the carboxylate catalytic base in the KE07 design and in other KE designs, in KE70, proton abstraction is mediated by a His-Asp dyad. The catalytic dyad feature is very abundant in enzymes, as it allows efficient positioning and activation of the His by the aspartic or glutamic acid residue.²⁵ However, the catalytic dyad is also computationally challenging, as both His and Asp residues, and the interaction between them have to be optimized. In addition, the $\text{p}K_a$ of histidine and its tautomeric state have to be regulated. The KE07 and KE70 designs also varied in their response to optimization by directed evolution. The KE07 design represented a deep energetic minimum. The design features and designed residues remained

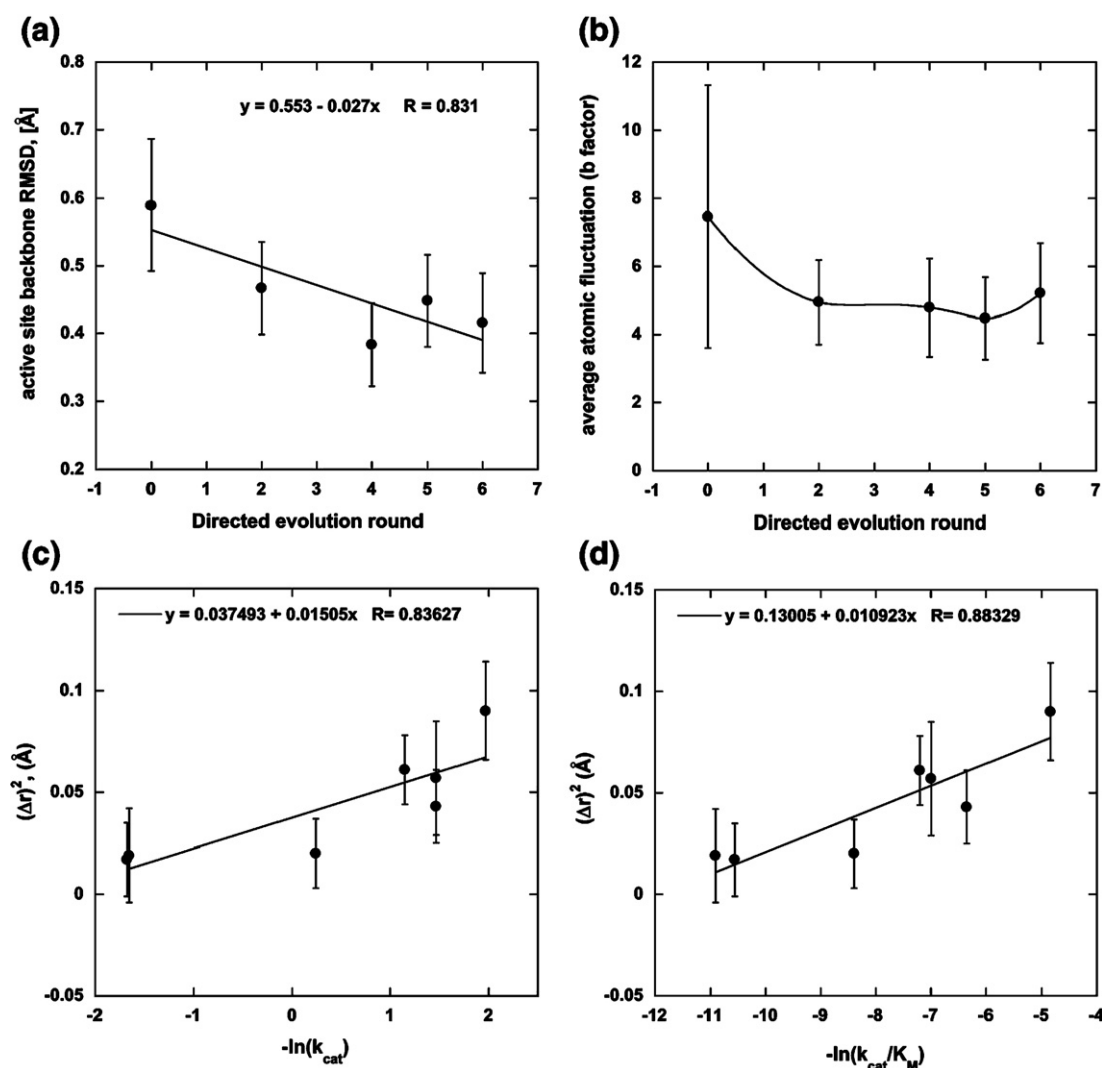


Fig. 7. (a) Active-site backbone RMSD. The full width at half maximum of each RMSD distribution is represented by individual error bars. (b) Active-site backbone B -factor *versus* the round of directed evolution. Error bars depict the spread (standard deviation) in atomic fluctuations across all active-site residues. (c) The square of the deviation $(\Delta r)^2$ from the ideal hydrogen-bond distance (1.8 Å) *versus* $-\ln(k_{cat})$. (d) The square of the deviation $(\Delta r)^2$ from the ideal hydrogen-bond distance (1.8 Å) *versus* $-\ln(k_{cat}/K_M)$.

essentially unchanged and only underwent fine-tuning. Only one of the designed residues was mutated, and even this residue is a second-shell residue that does not directly interact with the substrate. The KE70 design turned out to be both more “evolvable” and redesignable. Several designed residues were spontaneously altered in the first rounds of directed evolution, including the substrate stacking residue (mutation Tyr48Phe). This change suggested that the designed active site could be computationally optimized, as proved to be the case.

Despite the significant changes in the optimized KE70 design, the catalytic dyad His17-Asp45 remained unmodified. Similarly, in the divergent evolution of families and superfamilies of enzymes,

the basic mechanistic features and key active-site residues remain unchanged, while the other parts of the active site are modified to fit specific substrates and reactions.²² These results also highlight the strengths and weaknesses of the current computational enzyme design methodology—the general outlines of the active site can be created in broad strokes, but the fine-tuning of active-site electrostatics is very difficult to achieve only by computation.

Unlike the case of KE07, optimization of KE70 involved extensive computational design optimizations, rationally designed mutations, and directed evolution that combined the designed mutations with the random ones. In the case of KE07, it was shown that directed evolution, which requires no previous knowledge of structure–function, could be

successfully applied to improve computationally designed enzymes and can therefore bridge some of the gaps created by our limited design skills.¹⁶ However, the bottleneck of directed evolution is the very limited sequence space that can be explored by error-prone mutagenesis (single-nucleotide changes, with strong biases against certain mutations) and by medium-throughput screening. The screening throughput in 96-well plates using the colorimetric 5-nitrobenzisoazole substrate is relatively low, as only ~ 1000 variants were screened per round. Structure-based computational design was therefore used not only to produce the starting point enzyme but also to optimize it. Various combinations of mutations identified by the Rosetta algorithm were introduced into the evolved KE70 variants in a combinatorial manner. The incorporation of backbone flexibility yielded improved solutions compared to the original fixed backbone design, but only in a subset of cases—varying the backbone has associated perils (the protein may not fold or the backbone shift may not be brought about by sequence changes). To minimize the incorporation of mutations that are not beneficial or even deleterious, we first experimentally validated the designed mutations. Those proven advantageous or neutral were then incorporated into the evolved KE70 variants in a combinatorial manner.²⁰ This strategy allowed sampling of the various combinations of mutations while maintaining library sizes that are compatible with the screening capacity. Notably, many of the designed mutations involved changes in more than one nucleotide within the same codon (e.g., Ala21Asn/Gln/Arg, Ala238Met, and Ser239His), which are very unlikely to occur upon random mutagenesis. Likewise, insertions are not a likely outcome of error-prone mutagenesis.

While several of the computationally designed mutations have been incorporated into the evolving KE70 variants, it remains unclear which design strategy may yield the best designs. The procedures applied here followed a ‘trail-and-error’ approach, rather than a systematic exploration of various approaches. Nonetheless, the approach of including computationally designed mutations in a ‘spiking’ manner enabled us to explore a relatively large number of possibilities and to ‘hedge the bets’ by including mutations that may actually reduce rather than increase activity.

Changes in the designed residues of KE70, complemented by random mutations introduced by directed evolution, led to >400 -fold improvement of KE70's catalytic efficiency. Substrate binding (up to a 12-fold decrease in K_m), as well as the turnover number (up to a 53-fold increase in k_{cat}), were improved. The most efficient evolved KE70 variants exhibit k_{cat}/K_m values of $\sim 5 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$, which brings them closer to natural enzymes. Nonetheless, in view of the high reactivity of 5-nitrobenzisoazole,

it is clear that the catalytic powers of the evolved KE70s are far below those of natural enzymes. The k_{cat}/k_{uncat} values of the evolved KE70 variants are $\sim 5 \times 10^6$, which compares very favorably with the other enzyme mimics.²⁶ For comparison, the k_{cat}/K_m of the evolved KE7 is $2.6 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$. The k_{cat}/K_m of the catalytic antibody 34E4, which also catalyzed the Kemp elimination of 5-nitrobenzisoazole, is $5.45 \times 10^3 \text{ s}^{-1} \text{ M}^{-1}$, and its k_{cat}/k_{uncat} is 2.1×10^4 .¹⁰

The improvement in the catalytic activity of the KE70 variants has three main origins:

1. Modification of the active-site electrostatics. While the His17-Asp45 dyad remained unchanged, the electrostatics surrounding it were modified; in particular, the unfavorable bonds between Arg69 and the catalytic dyad (especially Asp45), which likely quenches the dyad's basicity, were broken (Fig. 5). Modification of the electrostatic environment was also a key feature in the optimization of the KE07 design, where an unfavorable bond between the catalytic base (Glu101) and Lys222 was broken, thus leading to higher basicity and activity of the catalytic base.¹⁶ While electrostatic effects play a crucial role in the catalysis of Kemp elimination (and perhaps of most other reactions),²⁷ they are not well modeled by the computational design methodology. The different rotamers adopted by Arg69 (Fig. 5c) exemplify the difficulty in modeling the charged residues at protein surface. The different rotamers seem to be very close in energy, as indicated by their coexistence in the two molecules within the same asymmetric unit (Fig. 5c). A similar situation was observed in the analysis of the KE07 design and its evolved variants.¹⁶ Improvements in the computation of electrostatic interactions, which enable prediction of the energetics of hydrogen-bond networks, would clearly contribute to the improvement of computational design methodology.
2. Reshaping the active-site cavity. The active-site cavity of the designed KE70 was completely reshaped, resulting in a deeper pocket with a narrower entrance (Fig. 4). This led to improved substrate binding (indicated by lower K_m values).
3. Mobility of the active-site residues. MD simulations suggest that the optimization of KE70 design led to increased preorganization of the catalytic dyad and restricted mobility of most active-site residues, particularly of residues His17 and Asp45, in a way that could

enhance the catalytic activity. The simulations further suggest that the increased k_{cat} and k_{cat}/K_m values of the evolved variants correlate with the formation of a more favorable geometry in the catalytic His-Asp dyad as a shorter hydrogen bond between these residues is achieved. This is in line with the primal requirement (emphasized early on by Jencks²⁸) that the catalytic residues and the substrate need to be precisely and tightly positioned to optimize the interactions essential for catalysis.²⁹

In summary, computational and evolutionary optimization methods were combined to yield improved Kemp eliminases. Computational algorithms can both generate the initial designs, which can then serve as starting points for directed evolution, and assist in the optimization of these designs by identifying beneficial amino acid substitutions for testing.³⁰ Further improvements in the computational methods may include better modeling of the electrostatic networks between charged residues, including additional parameters in the calculations (e.g., pK_a of the catalytic base), and MD simulations of the top-ranked designs to assess the relevant parameters such as bond lengths and residue fluctuations. Future work will be aimed at generating catalysts designed and optimized *de novo* not only for model reactions with activated substrates such as 5-nitrobenzisoxazole but also for more challenging reactions of the type catalyzed by natural enzymes.

Materials and Methods

Cloning

The synthetic gene encoding the designed KE70 protein was purchased from Codon Devices, Inc., and the gene was cloned into the His-tag expression vector pET29b (Novagen) using NdeI and XhoI restriction sites. In the libraries, the NdeI site was replaced by NcoI, which involved the addition of an alanine residue at the N-terminus, just after the initial methionine.

Backrub protocol

TS placement was carried out as previously described²¹ on the rigid backbone scaffold (PDB accession code 1JCL) using the RosettaMatch algorithm. In addition to the previously described theozymes used, theozymes lacking a hydrogen-bonding group in the phenoxide oxygen (Ser, Tyr, or Lys in previous theozymes) were allowed.

For each of the matches, gradient-based minimization was used to optimize the catalytic geometrical constraints.

Residues surrounding the TS were then redesigned to maximize the stability of the active-site conformation. Five hundred of the lowest-scoring designed variants were then subjected to a backrub simulation, as previously described,¹⁹ to generate five variants for each starting design. In general, the level of observed divergence of the loop regions was small (Supplementary Table 1), but the resulting subtle changes in $C^\alpha-C^\beta$ bond vectors changed the allowed spectrum of designed residues. Each of the backrub variants was then redesigned as before, using both the standard Rosetta van der Waals function and a soft van der Waals function, for a total of 10 redesigns per starting model. All models (flexible and fixed backbone) were subjected to both energetic and geometric filters (cavities, exposed surface area, and number of hydrogen bonds). The mutations of the eight variants that were finally selected for expression are listed in Supplementary Table 2.

Library making and screening

Several mutagenesis methods were used to create genetic diversity in KE70 genes. Error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit; Stratagene) was used for random mutagenesis; at certain rounds (Table 1), the genes were shuffled as previously described.¹⁶ Selected positions were mutated by incorporation of spiking oligonucleotides during the assembly PCR, as described previously.²⁰ After mutagenesis and/or shuffling, the KE70 genes were recloned into the original pET29b plasmid, and the ligated DNA was transformed into *E. coli* DH5 α cells. Typically, 10^4 – 10^5 transformants were obtained, and the plasmids encoding the libraries were extracted. It should be noted that random mutagenesis and gene shuffling are not exclusive (i.e., during random mutagenesis, the genes are shuffled, and the shuffling procedure incorporated a certain level of random mutations created by polymerase errors; ~ 1 mutation per gene, on average). The libraries were screened with 5-nitrobenzisoxazole, as previously described.¹⁶

Enzymatic characterization

Variants subjected to detailed analysis were produced and purified as described previously.¹⁶ The enzymes were stored at 4 °C in storage buffer [25 mM Hepes (pH 7.25) with 100 mM NaCl, 5% glycerol, and 0.02% sodium azide]. For kinetic characterization, the reactions were started by adding 150 μ l of substrate [final concentration: 0.13–1.05 mM or 0.065–0.525 mM in 25 mM Hepes (pH 7.25) with 100 mM NaCl] to 50 μ l of enzyme in storage buffer (various concentrations were used for different variants). 5-Nitrobenzisoxazole was used from a 0.1 M stock in acetonitrile. The cosolvent percentage was 1.5%, and the glycerol percentage was 1.25% in all reaction mixtures. Storage buffer without protein was used for the background reaction. Product formation was monitored spectrophotometrically at 380 nm (Power HT microtiter scanning spectrophotometer) in 200- μ l reaction volumes using 96-well plates. The reported results are the average of at least three independent measurements, and the error ranges between the repeated experiments were <20%.

Table 3. Summary of crystallization, data collection, and refinement of the KE70 variant structures

Variant	KE70 design	KE70 design	KE70 R2 2/7E	KE70 R2 7/12F	KE70 R2 3/5G	KE70 R4 8/5A	KE70 R5 7/4A	KE70 R6 6/10A	KE70 R6 6/10A with 5-nitrobenzotriazole
<i>Crystal parameters</i>									
Crystallization conditions	0.2 M NaF, 0.1 M Bistris, propane (pH 7.5), 20% PEG 3350	0.2 M NaF, 0.1 M Bistris, propane (pH 7.5), 20% PEG 3350	0.2 M NH ₄ Cl, 0.1 M 4-morpholineethanesulfonic acid (pH 6), 20% PEG 6000	0.1 M Na-citrate (pH 5.5), 20% PEG 3000	0.1 M PCB (pH 5), 20% PEG 1500	0.2 M NH ₄ NO ₃ , 20% PEG 3350 (pH 6.3)	0.2 M (NH ₄) ₂ SO ₄ , 0.2 M NaAc (pH 4.6), 30% PEG MME 2000, 3% trimethylamine N-oxide	0.2 M MgCl ₂ , 0.1 M NaAc (pH 5), 20% PEG 6000	0.1 M Na-malonate, imidazole, and boric acid (pH 4); 25% PEG 1500
Crystallization method	Sitting-drop vapor diffusion	Sitting-drop vapor diffusion	Sitting-drop vapor diffusion	Microbatch under oil	Microbatch under oil	Sitting-drop vapor diffusion	Microbatch under oil	Sitting-drop vapor diffusion	Sitting-drop vapor diffusion
Temperature (°C)	19	19	19	19	4	19	4	19	19
Space group	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁ ₂ ₁	<i>P</i> ₂ ₁ ₂ ₁	<i>C</i> ₂	<i>P</i> ₄ ₃ ₂ ₁	<i>C</i> ₂	<i>P</i> ₂ ₁	<i>P</i> ₂ ₁
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	62.22, 53.29, 81.84	62.59, 53.51, 81.07	52.39, 82.87, 127.96	52.13, 83.33, 129.20	99.01, 96.23, 71.27	52.82, 52.82, 153.81	90.52, 64.72, 49.03	49.10, 41.22, 139.44	48.84, 41.33, 138.81
Cell dimensions α , β , γ (°)	90.00, 110.19, 90.00	90.00, 110.34, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 125.88, 90.00	90.00, 90.00, 90.00	90.00, 104.69, 90.00	90.00, 94.96, 90.00	90.00, 95.35, 90.00
Number of copies in the asymmetric unit	2	2	2	2	2	1	1	2	2
PDB accession code	3NPU	3NPV	3NPW	3NPX	3NQ2	3NQ8	3NQV	3NRO	3Q2D
<i>Data collection</i>									
Data collected	In-house source	ESRF beamline ID14-4	In-house source	ESRF beamline ID14-4	In-house source	ESRF beamline ID14-4	In-house source	In-house source	In-house source
Wavelength (Å)	1.54178	0.9765	1.54178	0.975	1.54178	0.939	1.54178	1.54178	1.54178
Resolution range (Å)	50.00–2.25	50.00–1.48	50.00–2.15	50.00–1.80	50.00–2.02	50.00–1.40	50.00–1.70	50–2.20	50.00–2.20
Last resolution shell (Å)	2.33–2.25	1.51–1.48	2.23–2.15	1.86–1.80	2.09–2.02	1.45–1.42	1.73–1.70	2.24–2.20	2.24–2.20
Number of observations	154,480	751,597	220,501	506,485	177,142	881,041	150,631	109,838	493,378
Number of unique reflections	23,599	80,055	31,353	53,243	34,934	43,027	28,862	27,694	27,722
Completeness (%) ^a	97.8 (83.8)	98.9 (98.6)	99.9 (98.9)	99.6 (99.1)	97.9 (95.3)	97.9(100.0)	95.8 (92.4)	95.4 (95.7)	96.4 (79.9)
Redundancy	6.5 (4.9)	5.4 (5.3)	7.0 (5.5)	9.5 (8.4)	5.1 (4.4)	20.5 (23.1)	5.2 (3.5)	4.0 (4.0)	4.6 (3.9)
$\langle I \rangle / \langle \sigma(I) \rangle$ ^a	22.3 (4.9)	24.6 (6.6)	24.2 (2.9)	28.7 (5.8)	19.6 (3.1)	57.28 (12.5)	14.5 (2.6)	15.5 (3.0)	15.1 (2.5)
R_{merge} ^b on I (%) ^a	8.1 (21.7)	5.2 (24.2)	8.1 (39.4)	9.1 (31.0)	7.0 (32.0)	5.8 (24.3)	13.1 (37.0)	9.8 (49.6)	9.3 (34.2)

<i>Refinement and model statistics</i>									
Total number of reflections	22,380	78,893	29,720	50,467	33,171	40,785	27,408	26,268	26,309
Number of reflections in the test set	1292	4150	1574	2703	1757	2155	1451	1416	1403
Water molecules	40	263	83	228	159	166	124	67	35
Mean <i>B</i> value of protein (Å ²)	30.2	17.6	28.9	19.0	19.3	14.6	26.1	31.4	28.6
Mean <i>B</i> value of water (Å ²)	24.2	25.8	25.9	24.6	19.3	22.7	30.7	26.7	20.0
<i>R</i> _{cryst} (%) ^f	18.9	14.3	21.96	19.69	20.54	17.4	20.5	20.7	21.81
<i>R</i> _{free} (%) ^d	24.1	17.01	25.8	22.31	25.6	19.0	23.4	26.8	27.12
RMSD bond length (Å)	0.020	0.020	0.019	0.011	0.010	0.007	0.012	0.026	0.026
RMSD bond angle (°)	1.73	2.02	1.63	1.26	1.35	1.13	1.44	2.08	2.01
<i>Stereochemical parameters</i>									
Ramachandran plot									
Residues in most favored regions (%)	95.2	95.7	95.6	95.4	94.5	95.4	94.0	94.2	94.0
Residues in additionally allowed regions (%)	4.1	3.9	4.0	3.6	5.1	4.2	5.6	5.3	5.5
Residues in generously allowed regions (%)	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0
Residues in disallowed regions (%)	0.7	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

PEG, polyethylene glycol; ESRF, European Synchrotron Radiation Facility.

^a Values in parentheses are for the last resolution shell.

^b $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where *I* denotes the observed intensities.

^c $R_{\text{cryst}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where *F*_o denotes the observed structure factor amplitude, and *F*_c is the structure factor calculated from the model.

^d *R*_{free} is for 5% of randomly chosen reflections excluded from the refinement.

pH-rate profiles

k_{cat} and K_m values were determined with 5-nitrobenzoxazole at pH 4.5–8.0 for the KE70 variants Round 2 7/12F, Round 4 8/5A, and Round 6 6/10A, and at pH 5.0–8.0 for the KE70 design. Initial velocities (v_0) were determined at eight different substrate concentrations (0.13–1.05 mM). The buffers used (at 50 mM) were as follows: citrate (pH 5.0–5.5), 4-morpholineethanesulfonic acid (pH 5.5–6.5), and 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (Bistris) propane (pH 6.5–8.0) at 50 mM. At pH 5.5 and pH 6.5, the activity was measured with both relevant buffers to eliminate the buffer effects. Measurements at pH >8 were precluded due to substrate decomposition and the low activity of the KE70 design. The ionic strength was adjusted to a total of 0.1 M with NaCl. The enzyme stocks were kept in 10 mM Hepes (pH 7.25) containing 100 mM NaCl and 5% glycerol.

Data analysis

Kinetic parameters (k_{cat} , K_m , and k_{cat}/K_m) were obtained by fitting the data to the Michaelis–Menten equation [$v_0 = k_{\text{cat}}[E]_0[S]_0 / ([S]_0 + K_m)$], using the program Kaleidagraph 5.0. When working with low substrate concentrations, we fitted the data to the linear regime of the Michaelis–Menten model [$v_0 = [S]_0[E]_0 k_{\text{cat}}/K_m$], and we deduced k_{cat}/K_m from the slope. All the data presented are the averages of at least three independent experiments with standard deviations. The pH-rate data [k_{cat} and k_{cat}/K_m values for each pH value: $(k_{\text{cat}})^{\text{H}}$ and $(k_{\text{cat}}/K_m)^{\text{H}}$] were fitted using the equations: $(k_{\text{cat}})^{\text{H}} = (k_{\text{cat}})^{\text{max}} \times 10^{-\text{p}K_a} / (10^{-\text{pH}} + 10^{-\text{p}K_a})$ and $(k_{\text{cat}}/K_m)^{\text{H}} = (k_{\text{cat}}/K_m)^{\text{max}} \times 10^{-\text{p}K_a} / (10^{-\text{pH}} + 10^{-\text{p}K_a})$, where $(k_{\text{cat}})^{\text{max}}$ and $(k_{\text{cat}}/K_m)^{\text{max}}$ are the plateau values of k_{cat} and k_{cat}/K_m , and $\text{p}K_a$ is the apparent $\text{p}K_a$ value for the acidic group. The pH-rate data of the KE70 design were fitted to a “bell-shaped” model using the equations $(k_{\text{cat}})^{\text{H}} = (k_{\text{cat}})^{\text{max}} / [(10^{-\text{pH}} / 10^{-\text{p}K_{a1}}) + (10^{-\text{p}K_{a2}} / 10^{-\text{pH}}) + 1]$ and $(k_{\text{cat}}/K_m)^{\text{H}} = (k_{\text{cat}}/K_m)^{\text{max}} / [(10^{-\text{pH}} / 10^{-\text{p}K_{a1}}) + (10^{-\text{p}K_{a2}} / 10^{-\text{pH}}) + 1]$, where $(k_{\text{cat}})^{\text{max}}$ and $(k_{\text{cat}}/K_m)^{\text{max}}$ are the plateau values of k_{cat} and k_{cat}/K_m , and $\text{p}K_{a1}$ and $\text{p}K_{a2}$ are the apparent $\text{p}K_a$ values for the acidic and basic groups, respectively.

MD simulations

The crystal structures of KE70 and its evolved variants R2 2/7E, R2 3/5G, R2 7/12F, R4 8/5A, R5 7/4A, and R6 6/10A were used as starting geometries. Cocrystallized water molecules were removed. The substrate was docked into each active site with AutoDock4.³¹ Substrate parameters were generated within the antechamber module of AMBER 10³² using the general AMBER force field, with partial charges set to fit the electrostatic potential generated at HF/6-31G* by RESP.³³ The charges were calculated according to the Merz–Singh–Kollman scheme^{34,35} using Gaussian 03.³⁶ The electron density of the D-loop could not be resolved for all crystal structures. Here, the geometry of the KE70 wild-type chain B D-loop (Ala18 through Thr25) was used and grafted onto affected structures prior to MD.

The structures were immersed in a truncated octahedral box with a 10-Å buffer of TIP3P water molecules.³⁷ The

systems were neutralized by addition of explicit counterions. All subsequent calculations were performed using the widely tested Stony Brook modification of the AMBER 99 force field.³⁸ A two-stage geometry optimization approach was utilized: initial minimization of the positions of water molecules and ions, followed by an unrestrained minimization of all atoms. The systems were heated gently in six 50-K increments, using 50-ps steps from 0 K to 300 K at constant-volume periodic boundary conditions. Harmonic restraints of 30 kcal mol⁻¹ were applied to the solute, and the Langevin equilibration scheme was used to control and equalize the temperature. The time step was kept at 1 fs during the heating stages, allowing potential inhomogeneities to self-adjust. Each system was then equilibrated for 2 ns with a 2-fs time step at a constant pressure of 1 atm. Water molecules were triangulated with the SHAKE algorithm such that the angle between the hydrogen atoms is kept fixed. A 20-ns production MD simulation was performed for each of the systems (with and without the substrate bound to the active site) using PMEMD.³⁹ Geometries and velocities were saved every 100 steps (0.2 ps), resulting in a total of 100,000 frames from each production run. Long-range electrostatic effects were modeled using the particle mesh Ewald method.⁴⁰ Post-MD data extraction and analysis were performed using the ptraj module of AMBER 10 and the statistical analysis software OriginPro8.⁴¹

Crystallization, data collection, and refinement

For crystallization, the designed KE70 and its evolved variants were purified by a Ni-NTA HiTrap chelating HP column (Amersham), followed by gel filtration (HiLoad 16/60 SuperdexTM; Amersham).

Some of the crystals of KE70-evolved variants and the design were obtained by the microbatch-under-oil method using the Oryx6 robot (Douglas Instruments Ltd., East Garston, Hungerford, Berkshire, UK), and others were obtained by the sitting-drop vapor diffusion method using the Mosquito robot (TTP LabTech, Inc., Cambridge, MA, USA) (Table 3). The protein concentration used for the crystallization of all variants was 20–25 mg ml⁻¹. Diffraction data were integrated, scaled, and reduced using the HKL2000 program package.⁴² The structure of the R2 2/7E KE70 variant was solved by molecular replacement with the program PHASER,⁴³ using the refined structure of the deoxyribose phosphate aldolase of *E. coli* (PDB accession code 1JCL) as model. The structures of the remaining KE70 variants were solved using the R2 2/7E KE70 variant structure as model. All the steps of atomic refinement were carried out with the program CCP4/Refmac5.⁴⁴ The models were built to σ -weighted, $2F_{\text{obs}} - F_{\text{calc}}$, and $F_{\text{obs}} - F_{\text{calc}}$ maps using the program Coot.⁴⁵ The coordinates of all variants have been deposited in the PDB, and their accession codes are listed in Table 3. The KE70 models of the design and variants were evaluated with the program PROCHECK.⁴⁶

Accession numbers

Coordinates and structure factors for the structures of the KE70 design and its variants have been deposited in the PDB with the following accession numbers: 3NPU (KE70

design at 2.25 Å resolution), 3NPV (KE70 design at 1.4 Å resolution), 3NPW (R2 2/7E), 3NPX (R2 7/12F), 3NQ2 (R2 3/5G), 3NQ8 (R4 8/5A), 3NQV (R5 7/4A), 3NR0 (R6 6/10A), and 3Q2D (R6 6/10A with 5-nitrobenzotriazole).

Acknowledgements

We gratefully acknowledge financial support from the EU network BioModularH2 and the Defense Advances Research Projects Agency, a research grant from Meil de Botton Aynsley, and an Adams Fellowship from the Israel Academy of Science (to O.K.). D.S.T. is the incumbent of the Nella and Leo Benozio Professorial Chair.

Supplementary Data

Supplementary data to this article can be found online at [doi:10.1016/j.jmb.2011.01.041](https://doi.org/10.1016/j.jmb.2011.01.041)

References

1. Das, R. & Baker, D. (2008). Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382.
2. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.
3. 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.
4. Mandell, D. J. & Kortemme, T. (2009). Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* **5**, 797–807.
5. Baker, D. (2010). An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819.
6. 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.
7. Casey, M. L., Kemp, D. S., Paul, K. C. & Cox, D. D. (1973). The physical organic chemistry of benzisoxazoles: I. The mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294–2301.
8. Kemp, D. S., Cox, D. D. & Paul, K. G. (1975). The physical organic chemistry of benzisoxazoles: IV. The origins and catalytic nature of the solvent rate acceleration for the decarboxylation of 3-carboxybenzisoxazoles. *J. Am. Chem. Soc.* **97**, 7312–7318.
9. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. (2001). On the magnitude and specificity of medium effects in enzyme-like catalysts for proton transfer. *J. Org. Chem.* **66**, 5866–5874.
10. Thorn, S. N., Daniels, R. G., Auditor, M. T. & Hilvert, D. (1995). Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature*, **373**, 228–230.
11. Debler, E. W., Ito, S., Seebeck, F. P., Heine, A., Hilvert, D. & Wilson, I. A. (2005). Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl Acad. Sci. USA*, **102**, 4984–4989.
12. Hollfelder, F. & Tawfik, D. (1997). Efficient catalysis of proton transfer by synzymes. *J. Am. Chem. Soc.* **119**, 9578–9579.
13. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. (1996). Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature*, **383**, 60–62.
14. Hollfelder, F., Kirby, A. J., Tawfik, D. S., Kikuchi, K. & Hilvert, D. (2000). Characterization of proton-transfer catalysis by serum albumins. *J. Am. Chem. Soc.* **122**, 1022–1029.
15. Hu, Y., Houk, K. N., Kikuchi, K., Hotta, K. & Hilvert, D. (2004). Nonspecific medium effects *versus* specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J. Am. Chem. Soc.* **126**, 8197–8205.
16. 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.
17. Bar-Even, A., Noora, E., Davidia, D., Savirb, Y., Tawfik, D. C. & Milo R. (submitted for publication). The moderately efficient enzyme: evolutionary and physico-chemical trends shaping enzyme parameters. *Biochem.*
18. Davis, I. W., Arendall, W. B., III, Richardson, D. C. & Richardson, J. S. (2006). The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure*, **14**, 265–274.
19. Smith, C. A. & Kortemme, T. (2008). Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* **380**, 742–756.
20. Herman, A. & Tawfik, D. S. (2007). Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226.
21. 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.
22. Glasner, M. E., Gerlt, J. A. & Babbitt, P. C. (2006). Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol.* **10**, 492–497.
23. Murphy, P. M., Bolduc, J. M., Gallaher, J. L., Stoddard, B. L. & Baker, D. (2009). Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl Acad. Sci. USA*, **106**, 9215–9220.
24. Kiss, G., Röthlisberger, D., Baker, D. & Houk, K. N. (2010). Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773.
25. Bartlett, G. J., Porter, C. T., Borkakoti, N. & Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **324**, 105–121.
26. Kirby, A. J. (1996). Enzyme mechanisms, models, and mimics. *Angew. Chem. Int. Ed. Engl.* **35**, 707–724.
27. Roca, M., Liu, H., Messer, B. & Warshel, A. (2007). On the relationship between thermal stability and

- catalytic power of enzymes. *Biochemistry*, **46**, 15076–15088.
28. Jencks, W. P. (1975). Binding energy, specificity, and enzymic catalysis: the Circe effect. *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 219–410.
 29. Warshel, A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **273**, 27035–27038.
 30. Mena, M. A., Treynor, T. P., Mayo, S. L. & Daugherty, P. S. (2006). Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat. Biotechnol.* **24**, 1569–1571.
 31. Huey, R., Morris, G. M., Olson, A. J. & Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145–1152.
 32. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E. *et al.* (2008). AMBER10 University of California, San Francisco, CA.
 33. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280.
 34. Besler, B. H., Merz, K. M. & Kollman, P. A. (1990). Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439.
 35. Singh, U. C. & Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145.
 36. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R. *et al.* (2004). Gaussian 03, Revision C.02 Gaussian, Inc., Wallingford, CT.
 37. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926.
 38. Wang, J. M., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **21**, 1049–1074.
 39. Duke, R. E. & Pedersen, L. (2003). *PMEMD*. University of North Carolina, Chapel Hill, NC.
 40. Darden, T. A., York, D. & Pedersen, L. (1993). Particle mesh Ewald an $N \log(n)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
 41. Origin (OriginLab, Northampton, MA).
 42. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
 43. Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallogr. Sect. D*, **60**, 432–438.
 44. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
 45. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
 46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.