ARTICLES

Kemp elimination catalysts by computational enzyme design

Daniela Röthlisberger[1]*, Olga Khersonsky[4]*, Andrew M. Wollacott[1]*, Lin Jiang[1,2], Jason DeChancie[6], Jamie Betker[3], Jasmine L. Gallaher[3], Eric A. Althoff[1], Alexandre Zanghellini[1,2], Orly Dym[5], Shira Albeck[5], Kendall N. Houk[6], Dan S. Tawfik[4] & David Baker[1,2,3]

The design of new enzymes for reactions not catalysed by naturally occurring biocatalysts is a challenge for protein engineering and is a critical test of our understanding of enzyme catalysis. Here we describe the computational design of eight enzymes that use two different catalytic motifs to catalyse the Kemp elimination—a model reaction for proton transfer from carbon—with measured rate enhancements of up to 10^5 and multiple turnovers. Mutational analysis confirms that catalysis depends on the computationally designed active sites, and a high-resolution crystal structure suggests that the designs have close to atomic accuracy. Application of *in vitro* evolution to enhance the computational designs produced a >200-fold increase in k_{cat}/K_m (k_{cat}/K_m of 2,600 $M^{-1}s^{-1}$ and k_{cat}/k_{uncat} of >10^6). These results demonstrate the power of combining computational protein design with directed evolution for creating new enzymes, and we anticipate the creation of a wide range of useful new catalysts in the future.

Naturally occurring enzymes are extraordinarily efficient catalysts[1]. They bind their substrates in a well-defined active site with precisely aligned catalytic residues to form highly active and selective catalysts for a wide range of chemical reactions under mild conditions. Nevertheless, many important synthetic reactions lack a naturally occurring enzymatic counterpart. Hence, the design of stable enzymes with new catalytic activities is of great practical interest, with potential applications in biotechnology, biomedicine and industrial processes. Furthermore, the computational design of new enzymes provides a stringent test of our understanding of how naturally occurring enzymes work. In the past several years, there has been exciting progress in designing new biocatalysts[2,3].

Here we describe the use of our recently developed computational enzyme design methodology[4] to create new enzyme catalysts for a reaction for which no naturally occurring enzyme exists: the Kemp elimination[5,6]. The reaction, shown in Fig. 1a, has been extensively studied as an activated model system for understanding the catalysis of proton abstraction from carbon—a process that is normally restricted by high activation-energy barriers[7,8].

Computational design method

The first step in our protocol for designing new enzymes is to choose a catalytic mechanism and then to use quantum mechanical transition state calculations to create an idealized active site with protein functional groups positioned so as to maximize transition state stabilization (Fig. 1b). The key step for the Kemp elimination is deprotonation of a carbon by a general base. We chose two different catalytic bases for this purpose: first, the carboxyl group of an aspartate or glutamate side chain, and, second, the imidazole of a histidine positioned and polarized by the carboxyl group of an aspartate or glutamate (we refer to this combination as a His–Asp dyad). The two choices have complementary strengths and weaknesses. The advantage of the carboxylate is that it is likely to be in the basic

(deprotonated) form, but partial desolvation of the charged group in an apolar environment (to increase its relatively weak basicity) could destabilize the protein and further desolvation by the substrate could oppose binding. Although histidine is a better general base than a carboxylate, it is necessary to regulate both its pK_a and its tautomeric state. Coupling the histidine with a base such as aspartate in a dyad serves to both position the histidine and increase its basicity. If the pK_a of histidine is raised too high, however, it can become doubly protonated, rendering it ineffective as a base.
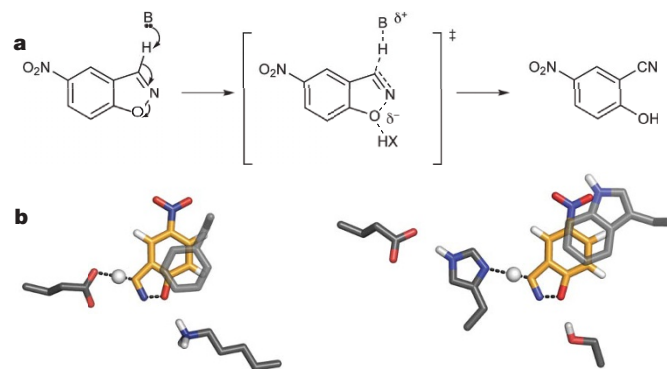
Figure 1 | Reaction scheme and catalytic motifs used in design. a, The Kemp elimination proceeds by means of a single transition state, which can be stabilized by a base deprotonating the carbon and the dispersion of the resulting negative charge; a hydrogen bond donor can also be used to stabilize the partial negative charge on the phenolic oxygen. **b,** Examples of active site motifs highlighting the two choices for the catalytic base (a carboxylate (left) or a His–Asp dyad (right)) used for deprotonation, and a π-stacking aromatic residue for transition state stabilization. For each catalytic base, all combinations of hydrogen bond donor groups (Lys, Arg, Ser, Tyr, His, water or none) and π-stacking interactions (Phe, Tyr, Trp) were input as active site motifs into RosettaMatch.

[1]Department of Biochemistry, [2]Biomolecular Structure and Design, and [3]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [4]Department of Biological Chemistry, and [5]Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel. [6]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.
*These authors contributed equally to this work.

For both the carboxylate- and histidine-based mechanisms, we included additional functional groups in the idealized active sites to further facilitate catalysis using both quantum mechanical and classical methods[9]. A hydrogen bond donor was used to stabilize the developing negative charge on the phenolic oxygen in the otherwise hydrophobic active site. Catalytic motifs lacking the H-bond donor were also tested, because the developing negative charge is relatively small in the transition state and can be easily solvated by water[9,10]. For each choice of catalytic site composition, density functional theory quantum-mechanical methods[11–13] were used to optimize the placement and orientations of the catalytic groups around the transition state for maximal stabilization (see Methods). Finally, because stabilization of the transition state by charge delocalization is a key factor in catalysis of the Kemp elimination[5–7,10,14], we chose to stack aromatic amino acid side chains on the planar transition state (Fig. 1b) using idealized π-stacking geometries[15].

We next used the RosettaMatch hashing algorithm[4] to search for constellations of protein backbone positions capable of supporting these idealized active sites in a large set of stable protein scaffolds with ligand-binding pockets and high-resolution crystal structures. As described in the Methods, the His–Asp dyad required generalizing RosettaMatch to handle side chains, such as the Asp, for which the range of allowed positions are referenced to another catalytic side chain rather than to the transition state; this was accomplished by identifying, for each His rotamer in a scaffold, the set of Asp rotamers that can provide the supporting hydrogen bond. The scaffold set spans a broad range of protein folds, including TIM barrels, β-propellers, jelly rolls, Rossman folds and lipocalins, amongst others (Supplementary Table 3). In a typical search, more than 100,000 possible realizations of the input idealized active site were found in the scaffold set. For each of these 'matches', gradient-based minimization[16] was used to optimize the rigid body orientation of the transition state and the torsional degrees of freedom of the catalytic side chains to best satisfy all catalytic geometrical constraints. Subsequently, residues surrounding the transition state were redesigned both to maximize the stability of the active site conformation and the affinity to the transition state and to maintain protein stability using the Rosetta design methodology for proteins[17] and small molecules[18]. Designs were screened for compatibility with substrate and product and were ranked on the basis of the catalytic geometry and the computed transition-state-binding energy.

A steady enrichment of the fraction of designs in the TIM barrel scaffold was observed throughout the enzyme design process. TIM barrel scaffolds represent 25% of the proteins in the input scaffold set, 43% of the initial matches, and 71% of the low-energy designs. Inspection of the designs suggests that the binding pockets in TIM barrel scaffolds were favoured because of the large number of take-off positions (all positions around the barrel pointing towards the cavity) for both the catalytic residues and the additional transition-state-binding residues optimized in the design process; the former favoured TIM barrel matches, and the latter favoured low-energy designs in TIM scaffolds. The TIM barrel is the most widespread and catalytically diverse fold in naturally occurring enzymes; our *in silico* design process seems to be drawn towards the same structural features as naturally occurring enzyme evolution.

Experimental characterization

Following the active site design, a total of 59 designs in 17 different scaffolds were selected for experimental characterization. Out of the 59 designs, 39 use an Asp or Glu as the generalized base and 20 use a His–Asp or His–Glu dyad. Eight of the designs showed measurable activity in Kemp elimination assays in an initial activity screen (Table 1; see Supplementary Table 4 for sequence information and Methods for experimental details). For each of these eight designs, mutation of the catalytic base (to Ala or Gln/Asn) markedly decreased the activity or abolished catalysis completely, suggesting that the observed activity results from the designed active site (Table 1; for some examples, see Fig. 2a). The designs have k_{cat}/K_m values in the range of 6 to 160 $M^{-1} s^{-1}$ (Table 1 and Fig. 2b); it was not possible to obtain saturation kinetics in all cases (for example, see KE10 (open squares) and KE61 (open triangles) in Fig. 2b) owing to low substrate solubility. Both catalytic motifs were used in active designs; of the two most active catalysts, which show a rate acceleration of roughly 10^5 and a k_{cat}/K_m of about 100, one uses the Glu as the base and the other uses the His–Asp dyad. All designs exhibited multiple turnovers (≥ 7)—a prerequisite for efficient catalysis.

Models for these two most active designs are shown in Fig. 3. In the KE59 design (Fig. 3a), which is in a TIM barrel scaffold, Glu 231 is the catalytic base and Trp 110 facilitates charge delocalization by π-stacking to the transition state. Additionally, Leu 108, Ile 133, Ile 178, Val 159 and Ala 210 create a tightly packed hydrophobic pocket that envelops the non-polar substrate. The polar residues Ser 180 and Ser 211 provide hydrogen-bonding interactions with the nitro group of the transition state. Mutation of the catalytic base Glu 231 to Gln abolished catalytic activity (Table 1 and Fig. 2a, open triangles). Attempts to add a hydrogen bond donor to stabilize the negative charge developing at the phenolic oxygen through a Gly 131

Table 1 | Kinetic parameters of designed enzymes

Design	PDB code of template	Base	Hydrogen-bond donor	π-stack	k_{cat} (s^{-1}; mean ± s.d.)	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat}/k_{uncat}*
KE07	1thf	E101	K222	W50	0.018 ± 0.001	1.4 ± 0.1	12.2 ± 0.1	1.6 x 10^4
KE07	1thf	E101A	K222	W50	0.0009 ± 0.0004	0.29 ± 0.14	3.5 ± 1.0	7.8 x 10^2
KE07	1thf	E101	K222A	W50	0.030 ± 0.004	1.3 ± 0.2	22.7 ± 2.7	2.6 x 10^4
KE10	1a53	E178	None	W210	NA	NA	51.6 ± 4.0	NA
KE10	1a53	E178Q	None	W210	NA	NA	0.17 ± 0.17	NA
KE15	1thf	D48	None	Y126	0.022 ± 0.003	0.63 ± 0.09	35.1 ± 4.8	1.9 x 10^4
KE15	1thf	D48A	None	Y126	NA	NA	0.05 ± 0.07	NA
KE16	1thf	D48	K201	Y126	0.006 ± 0.001	4.2 ± 0.1	13.4 ± 1.4	5.2 x 10^3
KE16	1thf	D48A	K201	Y126	ND	ND	ND	ND
KE59	1a53	E231	None	W110	0.29 ± 0.11	1.8 ± 0.6	163 ± 21	2.5 x 10^5
KE59	1a53	E231Q	None	W110	NA	NA	0.003 ± 0.002	NA
KE59	1a53	E231	G131S	W110	NA	NA	17.9 ± 0.7	NA
KE61	1h61	E100	None	W184	NA	NA	7.9 ± 0.6	NA
KE61	1h61	E100A	None	W184	NA	NA	0.7 ± 0.1	NA
KE70	1jcl	H16–D44	S137	Y47	0.16 ± 0.05	2.1 ± 0.8	78.3 ± 13.7	1.4 x 10^5
KE70	1jcl	H16A–D44	S137	Y47	NA	NA	0.02 ± 0.03	NA
KE70	1jcl	H16–D44N	S137	Y47	0.06 ± 0.03	2.3 ± 1.5	29.8 ± 4.6	5.2 x 10^4
KE71	1a53	H89–D85	Y210	W184	NA	NA	5.9 ± 0.2	NA
KE71	1a53	H89A–D85	Y210	W184	NA	NA	0.4 ± 0.2	NA
KE71	1a53	H89–D85N	Y210	W184	NA	NA	1.0 ± 0.2	NA

* k_{uncat} (1.16 x 10^{-6} s^{-1}) was determined in HEPES buffer at pH 7.25, and extrapolated to zero buffer concentration. The means and standard deviations of the kinetic parameters were calculated from at least three independent measurements. NA, not applicable; ND, not determined.
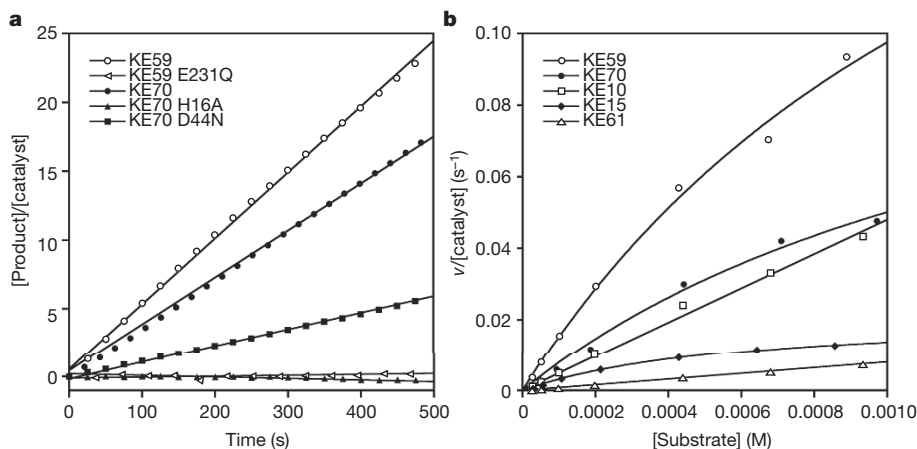
Figure 2 | Kinetic characterization of designed catalysts. a, Catalytic activity was measured by monitoring the product formation over time for KE59 (open circles) and KE70 (filled circles) at 400 μM substrate concentration. The y axis is the product concentration divided by the catalyst concentration that corresponds to the number of substrate turnovers. Deleting the catalytic base in both designs largely eliminates catalytic activity (open and filled triangles). Mutating Asp 44 of the catalytic dyad of KE70 to Asn (filled squares) causes a 2.5-fold reduction in activity. **b,** Michaelis–Menten plots for a representative selection of designed catalysts. The reaction velocity v divided by catalyst concentration is plotted on the y axis and the substrate concentration on the x axis. Some designs (for example, KE10 (open squares) and KE61 (open triangles)) show no saturation up to the maximal substrate solubility.

to Ser mutation caused a ninefold reduction in k_{cat}/K_m (Table 1), perhaps owing to unfavourable electrostatic interactions between the oxygen atoms on the serine and substrate; this large effect suggests that the transition-state-binding site is quite well defined. The aromatic-rich pocket and carboxylate base are reminiscent of the active site of the Kemp catalytic antibody 34E4 (ref. 10).

The KE70 design (Fig. 3b) uses the His–Asp dyad mechanism. Asp 44 positions and polarizes His 16 to optimally deprotonate the substrate. Tyr 47 π-stacks above the transition state, and together with Ile 201, Ile 139, Val 167, Ala 18, Ala 102 and Trp 71 creates a tight hydrophobic pocket around the transition state. The active site is again in a TIM barrel scaffold with the His–Asp dyad near the bottom of the site. Mutation of the catalytic base His 16 to Ala abolished catalytic activity (Table 1 and Fig. 2a, filled triangles), whereas mutating Asp 44 of the catalytic dyad to Asn produced an approximately 2.5-fold reduction (Table 1 and Fig. 2a, filled squares). In another design using a His–Asp dyad as general base (KE71), the analogous

Asp-to-Asn mutation reduced activity sixfold (Table 1) whereas the His-to-Ala mutation abolished catalysis (Table 1).

High-resolution structural information on designed proteins is essential to validate the accuracy of the design methodology. We were able to grow crystals and obtain a high-resolution structure of one of the early Glu-based designs, KE07 (see Supplementary Information for details). As shown in Fig. 4, the crystal structure and design model are virtually superimposable, with an active site (6.0 Å around the transition state) root mean squared deviation (r.m.s.d.) of 0.95 Å mostly reflecting modest side-chain rearrangements. The similarity between the design model and the crystal structure suggests that the active sites in our new enzymes resemble those in the corresponding design models. The subtle deviations in the backbone indicate loop regions in which explicitly modelling backbone flexibility may yield improved designs.

The crystal structure also revealed that Lys 222 makes a salt bridge to the catalytic Glu 101 in the absence of substrate, whereas in the
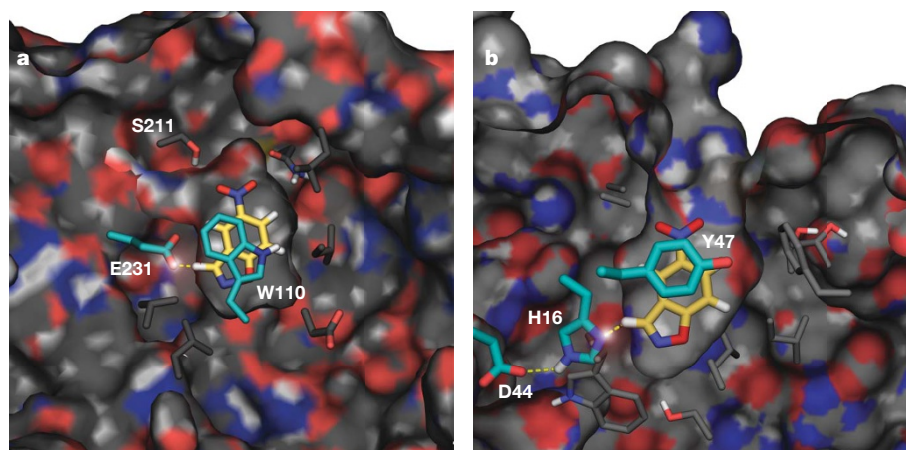
Figure 3 | Computational design models of the two most active catalysts. a, KE59 uses indole-3-glycerolphosphate synthase from *Sulfolobus solfataricus* as a scaffold. The transition state model is almost completely buried, with loops covering the active site. The mostly hydrophobic residues in the active site pocket pack the transition state model tightly, providing high shape complementarity (shape complementarity = 0.84; ref. 29). The polar residue Ser 211 interacts with the nitro group of the transition state to promote binding. The key catalytic residues (Glu 231 and Trp 110) are depicted in cyan. **b,** The deoxyribose-phosphate aldolase from *E. coli* is the scaffold for KE70. The shorter loops leave the active-site pocket freely accessible for the substrate. The transition state is surrounded by hydrophobic residues that provide high shape complementarity (shape complementarity = 0.77; ref. 29). His 16 and Asp 44 (in cyan) constitute the catalytic dyad whereas Tyr 47 (in cyan) provides π-stacking interactions.

designed model the ammonium of the lysine stabilizes the developing phenoxide in the transition state. Forming the productive transition state complex thus requires breaking of the salt bridge, and therefore elimination of the salt bridge in the unbound state would be expected to improve catalysis. We tested this prediction by substituting the lysine with an alanine, and this resulted in a 2.5-fold increase in k_{cat}/K_m (Table 1).

Directed evolution

In vitro evolution has been shown to markedly improve the stability, expression and activity of enzymes, and is currently the most widely used and successful approach for refining biocatalysts[19]. However, *in vitro* enzyme evolution generally requires a starting point with at least a low level of the desired activity, which is then optimized by repeated rounds of mutation and selection (for a notable exception, see ref. 20). We reasoned that *in vitro* evolution would be an excellent complement to our computational design efforts. The design calculations ensure that key catalytic functional groups are correctly positioned around the transition state, and, as demonstrated above, can generate active catalysts without requiring any starting activity. Thus, computational design can potentially provide excellent starting points for *in vitro* evolution. In contrast, the design process does not explicitly model configurational entropy changes, longer range second-shell interactions, and dynamics effects that can be important for efficient turnover; these shortcomings can potentially be remedied by directed evolution. Directed evolution can be valuable both in improving the designed catalysts and in stimulating improvements in the computational design methodology by shedding light on what is missing from the designs.

To investigate the extent to which *in vitro* evolution methods can improve computationally designed enzymes, we initiated evolution experiments on KE07—the early design for which the crystal structure was determined. Seven rounds of random mutagenesis and
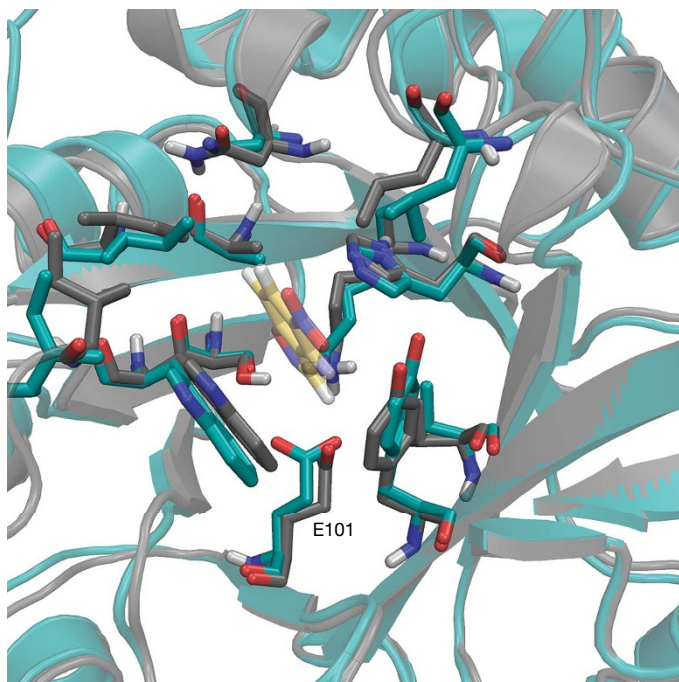
Figure 4 | Comparison of the designed model of KE07 and the crystal structure. The crystal structure (cyan) was solved in the unbound state and shows only modest rearrangement of active site side chains compared to the designed structure (grey) modelled in the presence of the transition state (yellow, transparent). (Backbone r.m.s.d. for the active site is 0.32 Å versus 0.95 Å for the active site including the side chains.) The observed electron density around relevant amino acids in the active site is shown in Supplementary Fig. 6. KE07 contains 13 mutations compared to the starting template scaffold (PDB code 1thf).

shuffling (also including synthetic oligonucleotides that expanded the diversity at selected residues), followed by screens in microtitre plates, yielded variants that had 4–8 mutations relative to KE07 and an improvement of >200-fold in k_{cat}/K_m (Table 2). Notably, the key aspects of the computational design, including the identities of the catalytic side chains, were not altered by the evolutionary process (indeed, mutating the catalytic base Glu 101 abolished the catalytic activity of both the designed template and its evolved variants; Table 2). Instead, the mutations were often seen in residues adjacent to designed positions (for example, Val 12, Ile 102, Gly 202), and thus provide subtle fine-tuning of the designed enzyme. Some mutations, such as Gly202Arg, are likely to increase the flexibility of regions neighbouring the active site. The hydrophobic residues Ile 7 and Ile 199 at the bottom of the active site were frequently mutated to polar or charged residues (the most common mutation being Ile7Asp), which may hold Lys 222 in position to stabilize the developing negative charge in the transition state while preventing interaction of Lys 222 with Glu 101. Consistent with this idea, the pK_a of the catalytic Glu 101 shifts from <4.5 to 5.9 in the evolved variant with the Ile7Asp mutation (for details, see Supplementary Information). Although the Lys222Ala mutation increases the activity of the original KE07, it significantly decreases the activity of the evolved variants, perhaps owing to the uncompensated additional negative charge.

Table 2 | Kinetic parameters of KE07 variants

Variant	Mutations	k_{cat} (s^{-1})	K_m, (mM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat}/k_{uncat}*
KE07 WT	-	0.018 ± 0.001	1.4 ± 0.1	12.2 ± 0.1	1.55×10^4
R2 11/10D†	K19E Q123R K146T G202R N224D	0.021 ± 0.001	0.31 ± 0.02	66 ± 2	8.38×10^4
R3 I3/10A	I7Q F86L K146T G202R N224D F229S	0.206 ± 0.003	0.48 ± 0.03	425 ± 16	1.78×10^5
R3 I3/10A E101A				≤ 3.9	
R4 1E/11H	I7D K146E G202R N224D	0.699 ± 0.001	2.40 ± 0.07	291 ± 9	6.02×10^5
R4 1E/11H E101A				≤ 2.4	
R5 10/3B	I7D V12M G202R N224D	0.49 ± 0.01	0.59 ± 0.03	836 ± 18	4.22×10^5
R6 3/7F	I7D K19E K146T G202R N224D	0.60 ± 0.07	0.69 ± 0.09	872 ± 25	5.17×10^5
R7 2/5B	I7D F77I G202R N224D	1.20 ± 0.08	0.86 ± 0.08	$1,388 \pm 44$	1.03×10^6
R7 10/11G	I7D V12M F77I I102F K146T G202R N224D F229S	1.37 ± 0.14	0.54 ± 0.12	$2,590 \pm 302$	1.18×10^6

* k_{uncat} (1.16×10^{-6} s^{-1}) was determined in HEPES buffer at pH 7.25, and extrapolated to zero buffer concentration. The means and standard deviations of the kinetic parameters were calculated from at least three independent measurements.
† The nomenclature of the evolved variants states the directed evolution round (R2, R3, and so on), the microtitre plate number and the location in the microtitre plate.

Conclusions

The marked increase in catalytic activity and in turnover (>1,000 catalytic cycles were observed for the evolved variants), achieved through screening a relatively small number of variants (800–1,600 clones per round) by molecular evolution standards bodes well for future combinations of computational design and molecular evolution. In particular, the *in vitro* evolution of the most active of the computational designs, for example, KE59 or KE70, has the potential to yield highly active catalysts for the Kemp elimination reaction. We anticipate the successful use of the combination of computational design and molecular evolution that we have described here for a wide range of important reactions in the years to come.

The challenge of generating new biocatalysts has led to several successful experimental strategies[20–22]. In particular, the Kemp elimination comprises a well-defined model for catalysis of proton transfer from carbon—a highly demanding reaction and a rate-determining step in numerous enzymes. It has therefore been the subject of several attempts to generate enzyme-mimics and models (such as catalytic antibodies[23], promiscuous protein catalysts[24] and enzyme-like polymers[14]). The catalytic parameters of the new enzymes described here are comparable to the most active catalysts of the Kemp elimination of 5-nitro-benzisoxazole described thus far, and provide further insights into the makings of an enzyme. Comparison with the catalytic antibodies[23] highlights the major shortcoming of many of the designs noted above—that is, their relatively weak binding of the substrate. Although the computational design methodology has the advantage of being able to explicitly place key catalytic residues, this may come at a cost of overall substrate and transition-state binding affinity. Consistently achieving high affinity to the transition state and high turnover numbers is a challenge that we are currently approaching by introducing scaffold backbone flexibility into the design process. This should enable us to create higher affinity binding sites formed by more precisely positioned constellations of binding and catalytic residues.

The computational methodology described here can be readily generalized to design catalysts for more complex multistep reactions[25]. The combination of computational enzyme design to create the overall active site framework for catalysing a synthetic chemical reaction with molecular evolution to fine-tune and incorporate subtleties not yet modelled in the design methodology is a powerful route to create new enzyme catalysts for the very wide range of chemical reactions for which naturally occurring enzymes do not exist. Equally importantly, computational design provides a critical testing ground for evaluating and refining our understanding of how enzymes work.

METHODS SUMMARY

Computational design. Transition state geometries were computed at the B3LYP/6-31G(d) level for idealized active sites containing either a carboxylate or an imidazole-carboxylate dyad as the general base. Aromatic side chains were placed above and below the transition state using idealized π-stacking geometries[15]. A six-dimensional hashing procedure[4] was applied to find transition state placements in a large set of protein scaffolds (Supplementary Table 3) that were consistent with the catalytic geometry. Residues surrounding the catalytic side chains and transition state were repacked and redesigned[17,18] to optimize steric, coulombic and hydrogen-bonding interactions with the transition state and associated catalytic residues.

Experimental characterization. The proteins were expressed in *Escherichia coli* BL21(DE3) using pET29b (Novagen) and purified over a Ni-NTA column (Qiagen). The proteins (1 μM to 10 μM) were assayed in 25 mM HEPES (pH 7.25) and 100 mM NaCl at 250 μM substrate concentration for the initial screening, and substrate dilutions from 1 mM to 11 μM were used for kinetic characterization. Kinetic parameters were determined in at least three independent measurements. Fitted K_m values above 1 mM (and their corresponding k_{cat} values) are necessarily approximate. Site-directed mutagenesis of catalytic residues and independent protein purifications by different protocols/laboratories were carried out to exclude possible contaminating enzymes (Supplementary Information).

***In vitro* evolution.** Gene libraries of KE07 were created by random mutagenesis using error-prone PCR with 'wobble' base analogues dPTP and 8-oxo-dGTP[26] using the Genemorph PCR mutagenesis kit (Stratagene), and by DNA shuffling of the most active variants[27]. In certain rounds, shuffling included the spiking of synthetic oligonucleotides that expanded the diversity at selected residues[28]. In each round, the cleared lysates of 800–1,600 individual colonies were assayed for hydrolysis of 5-nitrobenzisoxazole (0.125 mM) by following product formation at 380 nm. The most active clones were sub-cloned and sequenced, and the encoding plasmids were used as templates for subsequent rounds of mutagenesis and screening.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
2. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
3. Kaplan, J. & DeGrado, W. F. *De novo* design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **101**, 11566–11570 (2004).
4. Zanghellini, A. *et al.* New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).
5. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. The physical organic chemistry of benzisoxazoles I. The mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294–2301 (1973).
6. Kemp, D. S. & Casey, M. L. Physical organic chemistry of benzisoxazoles II. Linearity of the brønsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **95**, 6670–6680 (1973).
7. Hu, Y., Houk, K. N., Kikuchi, K., Hotta, K. & Hilvert, D. 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 (2004).
8. Hollfelder, F., Kirby, A. J., Tawfik, D. S., Kikuchi, K. & Hilvert, D. Characterization of proton-transfer catalysis by serum albumins. *J. Am. Chem. Soc.* **122**, 1022–1029 (2000).
9. Na, J., Houk, K. N. & Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **118**, 6462–6471 (1996).
10. Debler, E. W. *et al.* Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl Acad. Sci. USA* **102**, 4984–4989 (2005).
11. Lee, C., Yang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785–789 (1988).
12. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
13. Frisch, M. J. *et al. Gaussian 03, revision C. 02* (Gaussian, Inc., Wallingford, Connecticut, 2004).
14. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Efficient catalysis of proton transfer by synzymes. *J. Am. Chem. Soc.* **119**, 9578–9579 (1997).
15. Misura, K. M., Morozov, A. V. & Baker, D. Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. *J. Mol. Biol.* **342**, 651–664 (2004).
16. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C++* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 2002).
17. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
18. Meiler, J. & Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **65**, 538–548 (2006).
19. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**, 378–384 (2005).
20. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**, 828–831 (2007).
21. Cesaro-Tadic, S. *et al.* Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library. *Nature Biotechnol.* **21**, 679–685 (2003).
22. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl Acad. Sci. USA* **102**, 6855–6860 (2005).
23. Thorn, S. N., Daniels, R. G., Auditor, M. T. & Hilvert, D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **373**, 228–230 (1995).
24. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* **383**, 60–62 (1996).
25. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
26. Vartanian, J. P., Henry, M. & Wain-Hobson, S. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.* **24**, 2627–2631 (1996).
27. Abecassis, V., Pompon, D. & Truan, G. High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional

analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28**, E88 (2000).

28. Herman, A. & Tawfik, D. S. Incorporating synthetic oligonucleotides via gene reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226 (2007).

29. The. CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **50**, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions D.R. performed computational design using carboxylate and the His–Asp motif, and purified and experimentally characterized designed catalysts; O.K. synthesized the substrate, performed *in vitro* evolution and experimentally characterized evolved variants; A.M.W. performed computational design on the His–Asp motif; L.J. performed initial computational design on the carboxylate motif; J.D. and K.N.H. computed idealized active sites using quantum mechanics; J.B. and J.L.G. expressed and purified designed catalysts; E.A.A. helped with enzyme design set-up; A.Z. wrote RosettaMatch and helped with computational set-up; O.D. and S.A. crystallized KE07; and D.R., A.M.W., D.B., K.N.H., O.K. and D.S.T. designed the experiment and wrote the manuscript.

Author Information The crystal structure of KE07 has been deposited in the RCSB Protein Data Bank (www.rcsb.org) under the accession number 2rkx. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.B. (dabaker@u.washington.edu) or D.S.T. (dan.tawfik@weizmann.ac.il).

METHODS

Quantum mechanical transition state calculation. Quantum mechanical calculations using density functional theory with the B3LYP functional and the 6-31G(d) basis set[11,12] were used to locate transition structures (confirmed by vibrational frequency analysis) for the acetate- and imidazole/acetate-catalysed reactions in the gas phase. Lysine, serine, threonine and tyrosine functional groups were included in the calculations as hydrogen bond donors to stabilize the developing negative charge on the phenolic oxygen of the transition state. All calculations were carried using Gaussian03 (ref. 13).

Aromatic side chains (Phe, Tyr and Trp) were also modelled to stabilize charge delocalization of the transition state and to provide favourable π-stacking interactions. These side chains were placed using idealized π-stacking geometries[15] in a parallel configuration (4 Å separation) with the aromatic centre offset from the transition state rings by 1 Å. The aromatic groups were placed above either the five- or the six-membered ring and were allowed on both the top and the bottom faces of the transition state. Full rotation about the normal to the aromatic plane was permitted, allowing for variable Cβ–Cγ bond vector placement. The optimal catalytic geometry and the associated constraints for both reaction mechanisms are shown in Supplementary Fig. 1.

Scaffold selection. A large set of protein scaffolds were chosen as candidates for transition state placement. The selection criteria for these scaffolds were as follows: that a high-resolution crystal structure is available; that expression in *E. coli* is possible; that they are stable proteins; that they contain a preexisting pocket; and that they span a variety of protein folds. The protein scaffolds used in this study are listed in Supplementary Table 3.

For each scaffold, a three-dimensional grid representing the pre-existing pocket was mapped out using an in-house pymol plugin (Supplementary Fig. 2). This was used to reduce the extremely large search space for transition state placement (see below). The positions of potential catalytic residues near the active site were then compiled for each scaffold. In addition, a three-dimensional grid representing the protein backbone was created for each scaffold to allow for a fast clash check.

Transition state placement. To find active site placements in the input scaffolds, it is necessary to consider many alternative geometries for each catalytic motif. As described below, by varying the precise orientations of the catalytic side chains relative to the transition state, we generated very large ensembles of active site geometries. For each of these active site geometric variants, RosettaMatch[4] was used to position simultaneously transition state and catalytic residues into the set of pre-selected protein scaffolds so as to satisfy all catalytic constraints without steric overlap (only scaffold backbone atoms were modelled for clashes). Supplementary Figs 3–5 show the geometric descriptors used for catalytic side chain–transition state placement and the corresponding number of alternative conformations to be sampled. The His-based mechanism is shown as an example. The Glu/Asp-based mechanism was diversified similarly.

The geometric parameters for the catalytic base–transition state interaction were sampled much more finely because the relative geometry of the general base was considered to be more important than π-stacking or negative charge stabilization. Using the geometric parameters specified in Supplementary Fig. 3, there were 77,472,288 histidine–transition state conformations per position, 52,488 serine–transition state conformations per position, and 27,216 π-stacking–transition state conformations per position.

For a typical matching run, such as the TIM barrel protein scaffold, histidines were sampled at 41 positions around the barrel, and serine and π-stacking residues were placed at 119 residues to allow for catalytic side chains at second-shell residues. For this example, there are more than 1.5×10^{21} possible combinations for creating the catalytic motif, which would be computationally intractable to enumerate. By using the linear-scaling RosettaMatch algorithm, this number was reduced to a much more manageable number (8.7×10^{7}). The use of three-dimensional grids allows for rapid pruning of this large number of transition state conformations, as described above.

For the catalytic mechanism using histidine as the base, we prefiltered each scaffold to identify pairs of positions at which histidine and aspartate/glutamate rotamers can be placed to achieve the dyad geometry. Rotamer pairs with a van der Waals repulsive energy less than $2.0\,kcal\,mol^{-1}$ and hydrogen bonding energy less than $-0.5\,kcal\,mol^{-1}$ were stored in an 'interaction graph'. Matching was carried out using histidine as the catalytic residue, iterating only over histidine rotamers in the interaction graph of His–Asp and His–Glu pairs. For a given match, each Asp or Glu rotamer in the interaction graph that interact with the matched His rotamer was grafted onto the match, and the result screened to remove clashes between the transition state and the backing-up residue. Using the interaction graph decreases the number of potential histidine rotamers that must be modelled in the active site, and thus allows for even finer sampling of ligand rotamer sets. In the TIM barrel example described above, the

number of histidine rotamers sampled at the 41 residue positions was decreased from 3,321 (81×41) to 253 by precalculating and filtering only the subset of histidine rotamers that can form hydrogen bonds to Asp/Glu. This reduces the number of histidine–transition state conformations from 7.7×10^{7} to 5.9×10^{6}.

Geometric filters were applied to remove matches unlikely to produce good designs. Matches for which transition state poses clashed with more than four modelled Cβ atoms were removed as they would require too many Gly mutations to be introduced to accommodate the bound pose, potentially destabilizing the folded state. Matches with an insufficient number of neighbouring residues around the transition state would be expected to lead to underpacking during the design stage and were also removed.

Protein design. Residues surrounding the transition state and catalytic residues were selected for redesign, and the Rosetta protein design methodology[17,18,30] was used to create a pocket with high affinity for the transition state. Residue selection was carried out using a shell-based method. Residues with Cβ atoms within 8 Å were redesigned, those within 10 Å for which the Cα–Cβ bond vector pointed towards the transition state were redesigned, and all other residues within 12 Å were repacked. A rigid-body minimization of the transition state as well as side-chain relaxation of the protein was performed for each designed model.

Design filtering. A geometric filter was applied to choose models for which catalytic geometry was consistent with the specified constraints (tables in Supplementary Figs 3–5). The van der Waals interaction energy for the transition state and catalytic residues was a useful filter for choosing designs that were roughly well packed; designs with a transition state–protein van der Waals energy greater than $-5.0\,kcal\,mol^{-1}$ were removed. Filters were used to select for high transition state–protein shape complementarity[29], and to choose models with minimal small cavities surrounding the transition state (W. Sheffler and D.B., submitted). Solvent accessibility measures were used to remove models that completely buried the transition state. For the His–Asp dyad mechanism, an additional filter was added, requiring the His–Asp hydrogen bond remain on repacking of all residues in the presence of the transition state.

Protein expression and purification. Genes encoding the designs in the pET29b expression vector (Novagen) were purchased from Codon Devices, Inc. The catalytic-side-chain knockout mutations to Ala or Asn/Gln were introduced by site-directed mutagenesis as described[31]. After transformation into BL21 Star (Invitrogen), a one litre culture of auto-induction media[32] was inoculated with a single colony and shaken at 37 °C for 8 h. Expression was continued at 18 °C for 24 h. The cells were harvested, resuspended in 25 mM HEPES (pH 7.5) and 100 mM NaCl, and lysed by sonication. The soluble fraction was applied to a Ni-NTA column (Qiagen), washed with 20 mM imidazole, and the protein was eluted with 250 mM imidazole. The proteins were concentrated and the buffer was exchanged to 25 mM HEPES (pH 7.25) and 100 mM NaCl using a 5 ml Hi-Trap desalting column (GE Healthcare). For KE59, an additional size-exclusion chromatography step (Superdex75 10/300 GL from GE Healthcare) was performed. Protein concentrations were determined by measuring the absorbance at 280 nm using the calculated extinction coefficient[33]. To eliminate the possibility of observing the activity from a contaminating natural enzyme, further purification steps were carried out for KE07 and the evolved KE07 variants, for KE59 and for KE70 as described in the Supplementary Information section 10, validating the Kemp elimination activity of the designed and evolved enzymes.

Kinetic measurements. For the initial activity screen, 100 μl of the designed proteins (10 μM final concentration) were mixed with 100 μl of 500 μM substrate (freshly diluted from a 50 mM stock solution in acetonitrile) in 25 mM HEPES (pH 7.25) and 100 mM NaCl in a 96-well plate. For the kinetic characterization, the reactions were started by adding 150 μl of substrate dilutions (1 mM to 11 μM final concentration) in 25 mM HEPES (pH 7.25), 100 mM NaCl and 2% acetonitrile to 50 μl of protein (1 μM to 10 μM final concentration) in 25 mM HEPES (pH 7.25) and 100 mM NaCl (or no protein for the background reaction) in a 96-well plate. Product formation was followed at 380 nm in a SpectraMax M5e (Molecular Devices) plate reader at 27 °C in at least three independent experiments. The initial rates divided by the catalyst concentration were plotted against substrate concentration, and k_{cat} and K_{m} were determined by fitting the data to the Michaelis–Menten equation (equation (1)) using Kaleidagraph.

$$v/[\text{catalyst}] = k_{cat} \cdot [\text{substrate}]/(K_{m} + [\text{substrate}]) \qquad (1)$$

If saturation kinetics were not observed, k_{cat}/K_{m} values were calculated from a linear fit from the data.

Screening procedure. The libraries were screened by growing the cultures of *E. coli* BL21 cells in 96-deep-well plates and checking the activity of the lysates with 5-nitrobenzisoxazole. In brief, *E. coli* BL21 cells transformed with the libraries were grown on luria broth (LB) agar plates (containing 100 μg ml[−1]

kanamycin). Individual colonies were inoculated into 2YT supplemented with 50 μg ml^{-1} kanamycin (300 μl) in 96-deep-well plates, and grown for ~15 hours at 37 °C. Overnight cultures (20 μl) were inoculated into 2YT supplemented with 50 μg ml^{-1} kanamycin (500 μl) in 96-deep-well plates and grown to $A_{600\,nm}$ of ~0.6. Overexpression was induced by adding 1 mM isopropyl-β-D-thiogalactoside (IPTG), and the cultures were grown for another 5 h, centrifuged, and the pellet frozen overnight at −20 °C. The cells were lysed with lysis buffer, 250 μl well^{-1} (50 mM HEPES (pH 7.25), 0.2% Triton, 0.1 mg ml^{-1} lysozyme), and the lysates were cleared by centrifugation and assayed for hydrolysis of 5-nitrobenzisoxazole (0.125 mM) by following the release of the phenol product at 380 nm (Power HT microtitre scanning spectrophotometer). Overnight cultures of the most active clones were plated on LB agar plates containing 100 μg ml^{-1} kanamycin. To ensure monoclonality, and to verify the activity of the selected variants, the hydrolysis rates were re-assayed after growing two sub-clones from each original colony in the same conditions. Plasmids were extracted and used for sequencing and as templates for subsequent mutagenesis and screening rounds. Variants subjected to detailed analysis were re-transformed into *E. coli* BL21 cells, and the protein overexpressed and purified as described above.

Round 1. First-generation libraries were constructed from the designed KE07 gene by an error-prone PCR method using the 'wobble' base analogues dPTP and 8-oxo-dGTP[26]. The rate of mutations was 5 ± 3 per gene, and mutations were mainly of the transition type. The first round of KE07 evolution yielded active variants with lysate activity up to fivefold higher than of that of the starting point KE07.

Round 2. The 23 most active variants isolated in the first round of screening were subjected to DNA shuffling in the presence of the designed template (20%)[27] to yield second-generation libraries. The most active variants of round 2 had lysate activities up to 15-fold higher than that of the KE07. Analysis by SDS–PAGE demonstrated that the improvements in the activity were partially caused by enhancing the expression of KE07. Four active variants from round 2 were purified, and their kinetic parameters determined (Supplementary Table 1). Several dominant mutations in round 2 clones were identified; these can be divided into three groups: 1) Lys19Glu/Thr or Lys146Glu/Thr—mutations on the surface of the protein that seem to increase the expression levels of KE07. 2) Gly202Arg or Asn224Asp—mutations at the active site, probably interacting with the substrate-binding residues. Two other mutations in the helix 223–233 (Val226Ala and Phe229Ser), which are adjacent to Asp 224, were obtained. 3) Ile7Thr or Ile199Thr—residues located at the bottom of the active site, but not in direct contact with the substrate.

Round 3. The third-generation libraries were created by shuffling the four active variants of round 2 while randomizing various positions by incorporating spiking oligonucleotides during assembly of DNA fragments[28]:

Library 1: positions Ile 7 and Ile 199 were randomized (to Ile, Thr, Val, Ala, Phe, Ser, Glu, Asp, Gln, His), with the aim of finding the optimal combination of these residues at the bottom of the active site.

Library 2: positions Tyr 128 and His 201 were randomized (His 201 to Cys, Ser, Tyr, Ser, Thr, Asn; Tyr 128 to Leu, Pro, Ile, Thr, Val, Ala, Phe, Ser) to probe other residues at these designed positions that are responsible for benzisoxazole ring stacking.

Library 3: one or two amino acids were inserted between residues 224–225 and 225–226 to probe the variations of the helix 223–233, which seemed to be a target of many round 2 mutations.

Library 1 yielded clones with lysate activity up to 70-fold higher than that of KE07. Libraries 2 and 3 did not yield any improved variants, thus demonstrating that the designed stacking residues His 201 and Tyr 128 are at their optimal

configurations, and that the length of the helix 223–233 does not need to be further optimized.

Round 4. At round 4, randomization of Ile 199 was continued because it was not changed in most of the clones of round 3. Positions Ile 173 and Leu 176 were randomized as well (to Ala, Asp, Glu, Val, Leu, Ile, Thr, Asn, Lys, Pro, His and Gln) because these residues interact with Gly 202, which in most of the improved variants was mutated into arginine.

Round 4 yielded active variants with crude lysate activities up to 200-fold higher than that of KE07. The most active variants of rounds 3 and 4 were purified, and their catalytic parameters determined (Supplementary Table 1).

Sequencing of round 3 and 4 variants confirmed the importance of the mutations found in round 2. Lys19Glu/Thr and Lys146Glu/Thr mutations increased the expression levels, and Gly202Arg and Asn224Asp optimized the top part of the active site. Randomization of positions Ile 7 and Ile 199 at the bottom of the active site demonstrated that, in the optimal combination, Ile 7 is changed to a more polar residue and Ile 199 remains intact. In several improved variants, the residues Ile 173 and Leu 176 were mutated as well, but their effect is relatively minor.

Because the mutation Asn224Asp was found in all the improved variants of rounds 3 and 4 (with the exception of R4 2F/2G), we wanted to ensure that this mutation did not alter the initial design, by acting, for example, as a general base, thus replacing the designed base Glu 101. Thus, we created Glu101Ala mutants of the variants R3 I3/10A, R4 1E/11H and R4 2F/2G, and of the KE07. Mutagenesis of Glu 101 caused a significant decrease in the activity of all the variants (up to 1%). These results demonstrated that the initial design, in which Glu 101 acts as a general base, was maintained (Supplementary Table 2).

Round 5. The active variants from round 4 were subjected to random mutagenesis by error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit, Stratagene[34]) to yield the fifth-generation libraries, which contained 1 ± 1 mutations per gene and a large portion of shuffled genes. Mild lysate activity improvements (up to 1.5-fold) were observed, and the 12 most active variants from round 5 were subjected to another round of mutagenesis at a higher mutational load.

Round 6. At round 6, the 12 most active variants from round 5 were subjected to random mutagenesis by error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit, Stratagene[34]) to yield the sixth-generation libraries, which contained 3 ± 1 mutations per gene and a large portion of shuffled genes. Lysate activity improvements of up to 1.5-fold were observed.

Round 7. Seventh-generation libraries were created by shuffling the 20 active variants of round 6, and lysate activity improvements of up to threefold were observed.

The xyz coordinates of the design KE07, KE59 and KE70 are available in the Supplementary Information.

30. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332,** 449–460 (2003).

31. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154,** 367–382 (1987).

32. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).

33. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423 (1995).

34. Barlow, M. & Hall, B. G. Predicting evolutionary potential: *in vitro* evolution accurately reproduces natural evolution of the tem β-lactamase. *Genetics* **160,** 823–832 (2002).