

Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59

Olga Khersonsky^{a,1}, Gert Kiss^{b,2}, Daniela Röthlisberger^{c,3}, Orly Dym^d, Shira Albeck^d, Kendall N. Houk^b, David Baker^{c,e,f}, and Dan S. Tawfik^{a,4}

^aDepartment of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel; ^bDepartment of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; ^cDepartment of Biochemistry, University of Washington, Seattle, WA 98195; ^dIsrael Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel; ^eBiomolecular Structure and Design, University of Washington, Seattle, WA 98195; and ^fHoward Hughes Medical Institute, University of Washington, Seattle, WA 98195

Edited by William F. DeGrado, UCSF School of Pharmacy, San Francisco, CA, and approved May 4, 2012 (received for review December 26, 2011)

Computational design is a test of our understanding of enzyme catalysis and a means of engineering novel, tailor-made enzymes. While the *de novo* computational design of catalytically efficient enzymes remains a challenge, designed enzymes may comprise unique starting points for further optimization by directed evolution. Directed evolution of two computationally designed Kemp eliminases, KE07 and KE70, led to low to moderately efficient enzymes (k_{cat}/K_m values of $\leq 5 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$). Here we describe the optimization of a third design, KE59. Although KE59 was the most catalytically efficient Kemp eliminase from this design series (by k_{cat}/K_m and by catalyzing the elimination of nonactivated benzisoxazoles), its impaired stability prevented its evolutionary optimization. To boost KE59's evolvability, stabilizing consensus mutations were included in the libraries throughout the directed evolution process. The libraries were also screened with less activated substrates. Sixteen rounds of mutation and selection led to >2,000-fold increase in catalytic efficiency, mainly via higher k_{cat} values. The best KE59 variants exhibited k_{cat}/K_m values up to $0.6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, and $k_{\text{cat}}/k_{\text{uncat}}$ values of $\leq 10^7$ almost regardless of substrate reactivity. Biochemical, structural, and molecular dynamics (MD) simulation studies provided insights regarding the optimization of KE59. Overall, the directed evolution of three different designed Kemp eliminases, KE07, KE70, and KE59, demonstrates that computational designs are highly evolvable and can be optimized to high catalytic efficiencies.

computational protein design | enzyme mimic

The endeavor of making enzyme-like catalysts spans several decades (1) with the Kemp elimination being a thoroughly explored model (2–7). In this activated model system, base-catalyzed proton elimination from carbon is concerted with the cleavage of nitrogen-oxygen bond, thus leading to the cyanophenol product (Fig. 1A). Activation of a carboxylate base catalyst by desolvation is efficiently mimicked by aprotic dipolar solvents such as acetonitrile and by various enzyme mimics. Effective alignment of the substrate and charge-dispersing interactions that stabilize the negatively charged transition state (TS) have also been achieved by various enzyme mimics that catalyze the Kemp elimination (8, 9). However, generation of a (wo)manmade active site that exhibits all these features and performs as well as natural enzymes, remains a challenge.

Computational methods for predicting structure from sequence at atomic accuracy provide a new approach to enzyme engineering (10–12). The design involves two steps: (i) designing an active-site configuration that may confer efficient catalysis, (ii) computing a sequence that confers the desired configuration. Both steps are currently far from optimal, as the catalytic efficiency of designed enzymes falls far behind that of natural enzymes. Further, as with other enzyme mimics, computational designs tackle activated model systems and fail to catalyze challenging reactions (13). In prin-

ciple, directed evolution—protein optimization by cycles of random mutagenesis and selection—could be applied to optimize computationally designed enzymes. However, can directed evolution bridge the orders-of-magnitude gap between computational design and enzyme-like catalytic efficiencies?

We have previously generated a series of computationally designed Kemp eliminases (12). These were designed to have apolar active sites and a base (carboxylate or histidine) aligned against the C-H bond. The relative locations of the catalytic groups around the reaction's TS were optimized by quantum mechanical calculations. Apart from the catalytic base, additional residues were included: aromatic residue(s) to promote substrate binding and delocalize the TS's negative charge, and hydrogen donors for the stabilization of the phenolate's negative charge. The designed Kemp eliminases were obtained by identifying protein scaffolds in which the designed active site configurations could be realized using the RosettaMatch algorithm (14) and then optimizing TS binding with RosettaDesign (12). The active designs exhibited k_{cat}/K_m values in the range of $\leq 10 \text{ M}^{-1}\text{s}^{-1}$ (KE07) up to $160 \text{ M}^{-1}\text{s}^{-1}$ (KE59)—orders of magnitude lower than the catalytic efficiency exhibited by natural enzymes [k_{cat}/K_m values of 10^4 – $10^8 \text{ M}^{-1}\text{s}^{-1}$ (15)].

The KE07 design was optimized by directed evolution (16), and KE70 by a combination of directed evolution and computational design (17). The k_{cat}/K_m values of optimized KE07 and KE70 variants (2.4×10^3 – $5 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$) compare well with other enzyme mimics but are still inferior to natural enzymes. Further, these Kemp eliminases catalyze only the elimination of the activated 5-nitro-benzisoxazole substrate (5-nitro BI). Here we describe the optimization of KE59. This design was based on the α/β barrel scaffold of indole-3-glycerolphosphate synthase [IGPS, Protein Data Bank (PDB) entry 1A53]. The designed active site comprised a tightly packed hydrophobic pocket that envelops the nonpolar substrate (Fig. 1B). The designed E230 acts as the catalytic base, and W109 was introduced to facilitate

Author contributions: O.K., K.N.H., D.B., and D.S.T. designed research; O.K., G.K., and D.R. performed research; O.K., G.K., O.D., S.A., D.B., and D.S.T. analyzed data; and O.K., G.K., O.D., D.B., and D.S.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3UXA, 3UXD, 3UY8, 3UYC, 3UZ5, and 3UZJ).

¹Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195.

²Present address: Department of Chemistry, Stanford University, Stanford, CA 94305.

³Present address: Arzeda Corp., Seattle, WA 98102.

⁴To whom correspondence should be addressed. E-mail: tawfik@weizmann.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121063109/-DCSupplemental.

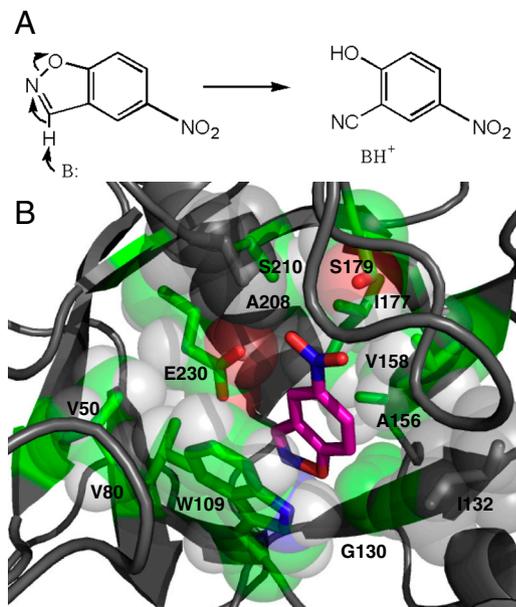


Fig. 1. The active site of the KE59 design. (A) Kemp elimination of 5-nitrobenzisoxazole. (B) The residues that were designed to create the Kemp eliminase active site (green), and the modeled 5-nitro BI (magenta). [Numbering of residues in this paper is $n - 1$ relative to the numbering in the previous publication, due to omission of N-terminal M (12)].

charge delocalization by π -stacking to the transition state. S179 and S210 were included to provide hydrogen-bonding interactions with the substrate's nitro group. KE59's long loops resulted in a deep pocket, which may have made KE59 the most active design within this series ($k_{\text{cat}}/K_m \geq 160 \text{ M}^{-1}\text{s}^{-1}$).

As other KE designs, KE59 was designed to catalyze the elimination of 5-nitro BI, an activated substrate with a pK_a^{LG} of 4.1. However, unlike the other designed Kemp eliminases (KE07 and KE70), KE59 also catalyzes the elimination of substrates with pK_a^{LG} values of up to 6.9, whose spontaneous rates of elimination are ≤ 27 -fold slower than 5-nitro BI. In this paper, we describe the improvement by directed evolution of both the stability of KE59 and its catalytic activity with nonactivated substrates.

Results

Boosting Evolvability by Spiking Consensus Mutations. The KE59 design was unstable, its soluble expression in *Escherichia coli* was extremely poor ($< 2 \text{ mg/L}$ culture), and it could not be purified to homogeneity. This was unexpected, given that KE59 design was based on a scaffold of hyperthermophilic enzyme, and the design process involved protein free energy minimization. Because mutations are destabilizing on average and function-modifying mutations tend to be more destabilizing than neutral ones (18, 19), proteins with marginal stability show low tolerance to mutations and hence limited evolvability (20, 21). Indeed, our attempts of directed evolution of unstable KE designs such as KE59 were unsuccessful. We therefore sought to increase the functional expression and evolvability of KE59 design by incorporation of consensus mutations.

Substitution of an amino acid residue for the one predominant in the protein family often leads to stabilization (22–24). Such consensus mutations were previously shown to promote protein evolvability (25). The sequence of the template protein of KE59 was aligned to sequences of related IGPSs (42–57% identity to KE59 design; see *SI Appendix, Fig. S1 and Table S1*), and several amino acids that deviate from the consensus were identified. At positions with no clear consensus, amino acids present in a significant number of family members were probed (see *SI Appendix, Table S2*). Although technically all these residues could

be mutated simultaneously, this could lead to protein destabilization, because many consensus mutations are deleterious (24). Therefore, the consensus mutations were spiked into the KE59 libraries, to allow various combinations of stabilizing mutations (26). The importance of combinatorial incorporation is highlighted by the fact that consensus mutations at only 10 out of 23 positions were incorporated into the evolved KE59 variants.

The mutations were incorporated in the first two rounds of directed evolution of KE59, and the variants were screened for increased activity. Incorporation of 6–10 consensus mutations significantly improved the soluble expression of KE59 (from $< 2 \text{ mg protein/L}$ culture to $> 30 \text{ mg protein/L}$ culture; see *SI Appendix, Fig. S2*), and enabled the accumulation of other mutations that led to higher catalytic activity. The stabilizing mutations were located at the protein surface, either in the loops or helices, and they had little effect on the kinetic parameters of KE59 (Table 1). Following another two rounds of directed evolution, there was a decrease in the expression of the evolved variants due to the destabilizing effect of function-modifying mutations (19). Therefore, at Round 5, consensus mutations were included again and combinations that were not observed in the first stabilization attempt were incorporated. The incorporated mutations included mutations of solvent-exposed aromatic residues to smaller hydrophobic residues (W7A, F21L, Y75G, Y151L, F245L), mutations of solvent-exposed leucines to charged/polar residues (L14R/K and L247Q), and mutations resulting in opposite charges (K9E, E142K) (see *SI Appendix, Fig. S3A*).

Due to the irreversibility of thermal denaturation of KE59 variants, the effect of the consensus mutations on KE59 stability could be examined only by measuring the melting temperatures (T_M) of KE59 variants by residual activity (see *SI Appendix, Fig. S4*). These apparent T_M values did not correlate with soluble expression levels (see *SI Appendix, Fig. S2*), possibly due to the fact that many consensus mutations affect kinetic rather than thermodynamic stability (25, 27). The primary outcome of the incorporated consensus mutations was therefore an increased tolerance to mutations, which enabled the directed evolution of KE59 towards higher catalytic activity.

Optimization of Catalytic Activity. Sixteen rounds of directed evolution were carried out by screening the libraries for Kemp elimination activity in crude lysates (see *SI Appendix, Table S4*). To maintain the ability of KE59 to catalyze elimination of nonactivated substrates (Table 1), at each evolution round, variants exhibiting improved rates with 5-nitro BI were also screened with less activated benzisoxazoles. Variants with similar, or enhanced, activity for less activated substrates, relative to the rate with 5-nitrobenzisoxazole, were taken for the next evolution round. After Round 7, the libraries showed no further improvement with 5-nitro BI, and we therefore focused on less activated substrates. The libraries were split to 5-NO₂ library, screened with 5-nitro BI ($pK_a^{LG} = 4.1$), and 6-Cl library, screened with 6-chloro BI ($pK_a^{LG} = 6.1$, approximately 10-fold slower spontaneous elimination rate). After two rounds (8 and 9), there was no improvement in the 5-NO₂ library but the activity of variants in the 6-Cl library improved significantly. Following rounds 10 and 11, at which there was no improvement with 6-chloro BI, the libraries were screened with 5-fluoro BI ($pK_a^{LG} = 6.8$, 5-F library). Rounds 12 and 13 yielded modest improvement with 5-fluoro BI, accompanied by a significant decrease in expression. We therefore switched at Round 14 to an *E. coli* strain overexpressing the GroEL/GroES chaperone that was shown to facilitate the accumulation of destabilizing yet function-enhancing mutations (28). However, Rounds 14–16 did not yield any significant improvement in catalytic activity, and it therefore seemed that the catalytic potential of KE59 had been exhausted.

Table 1. Mutations and kinetic parameters of representative KE59 variants

variant	Catalytic parameters: k_{cat} , s^{-1}				Consensus mutations	Other mutations:
	5-nitro BI	5,7-dichloro BI	6-chloro BI	6-fluoro BI		
KE59 design	$k_{cat}/K_M \sim 160^*$	not measured	below detection limit	below detection limit		
R1-7/10H	ND [†] ND [†] 328 ± 1	not measured	ND [†] ND [†] 16.4 ± 0.4	below detection limit	W7A, F21L, N33R, S69A, T94A, N163E, F175I, F245L	
R2-4/3D	0.528 ± 0.002 0.29 ± 0.01 $1,833 \pm 75$	ND [†] ND [†] $5,820 \pm 141$	ND [†] ND [†] 5.7 ± 0.7	ND [†] ND [†] 0.149 ± 0.002	K9E, L14R, F21V, N33K, S69A, T94D, E142K, N160H	V80A
R4-5/11B	4.5 ± 0.3 0.48 ± 0.03 $9,524 \pm 335$	ND [†] ND [†] $10,465 \pm 129$	0.041 ± 0.002 1.14 ± 0.14 36 ± 3	ND [†] ND [†] 3.91 ± 0.02	K9E, N33K, S69A, T94D	V80A, R181H, A208V, R222Y, L247Q
R5-11/5F	2.43 ± 0.04 0.36 ± 0.02 $6,706 \pm 365$	not measured	0.0256 ± 0.0003 0.80 ± 0.06 32 ± 2	not measured	K9E, F21L, N33K, S69A, T94D	I44N, V80A, R181H, A208V, R222Y, L247Q
R8-2/7A	5.4 ± 0.8 0.44 ± 0.04 $12,350 \pm 774$	not measured	0.0185 ± 0.0001 0.435 ± 0.005 42.3 ± 0.3	0.0060 ± 0.0002 1.03 ± 0.01 5.84 ± 0.01	K9E, F21L, N33R, D60N, S69A, Y75G, T94D, E142K, L247G	R22H, I44N, A76V, V80A, L107M, F111I, R181H, I200V, N203D, A208V
R9-1/4A	1.99 ± 0.08 0.32 ± 0.03 $6,147 \pm 399$	ND [†] ND [†] $28,650 \pm 270$	0.109 ± 0.005 0.60 ± 0.07 182 ± 13	ND [†] ND [†] 25.0 ± 0.7	K9E, L14R, F21V, N33K, S69A, Y75G, T94D, Y151L, N160H	L16Q, I48M, A76V, V80A, F111I, R181H, A208V, R222Y, L247Q
R13-3/11H	9.53 ± 0.62 0.16 ± 0.02 $60,430 \pm 2,000$	21.2 ± 0.4 0.037 ± 0.002 $573,090 \pm 19,160$	2.11 ± 0.31 0.98 ± 0.16 $2,150 \pm 39$	0.31 ± 0.02 0.98 ± 0.16 315 ± 32	K9E, L14R, F21V, N33K, S69A, Y75G, T94D, Y151L, N160H	L16Q, I48M, A76V, V80A, I104V, F111I, S179T, R181H, K190N, A208V, R222Y, S233T, L247Q

*The KE59 design was unstable, and the measured catalytic parameters can be an underestimate due to a fraction of soluble yet misfolded enzyme.

[†]ND, not determined. Due to limited substrate solubility, the data were fitted to the linear phase of the Michaelis-Menten model [$v_0 = [E]_0 k_{cat} [S]_{0t} / K_M$], and k_{cat}/K_M was deduced from the slope.

Kinetic Parameters and Mutations in the Evolved KE59 Variants. After 13 rounds of directed evolution, the catalytic efficiency of the initial design improved ≤ 380 -fold with 5-nitro BI. The rate improvement with the less activated 5-fluoro BI was even larger, $\geq 2,000$ -fold. For 5-nitro BI, both increased turnover numbers (k_{cat}) and decreased K_m values were observed, while for less activated substrates, the increase was primarily in k_{cat} (Table 1). The most active evolved variants acquired up to 23 mutations, of which nine came from the spiked consensus mutations, and the rest from random mutagenesis. The improvement was gradual and smooth thus making it hard to ascribe specific effects to individual mutations. Apart from the stabilizing consensus mutations, the mutations observed in the most active KE59 variants fall into the following categories (Table 1; see *SI Appendix, Fig. S3*):

1. Mutations at the protein surface (see *SI Appendix, Fig. S3B*) that are likely to improve the foldability and stability of KE59.
2. Mutations in the residues of the β strands that underline the designed active site (see *SI Appendix, Fig. S3C*), including mutations of designed residues (V80A and S179T) and mutations spatially adjacent to designed residues (I48M, L107M, and A208V).
3. Mutations in the loops that comprise the active site's top (see *SI Appendix, Fig. S3D*). These residues are in proximity to the substrate and could affect its positioning (e.g., F111I, R181H, and S233T).

Kinetic Analyses. The most distinct feature of the designed Kemp eliminases, and of other enzyme-mimics that catalyze this reaction, is an activated base catalyst (8, 29). To probe the reactivity of the catalytic E230, pH-rate profiles were obtained for R2-4/3D variant, representing the KE59 design, and the evolved variants from R4 to R13. The pH-rate profiles for both k_{cat} and k_{cat}/K_M exhibit an acidic shoulder that is likely to represent the titration of

E230 (pK_a values 6.1–6.7, Table 2; see *SI Appendix, Fig. S7*). In the case of KE07, directed evolution led to a significant increase in the basicity (≥ 1.7 pK_a units) and reactivity of its catalytic Glu (16). For KE59, the increase in E230's basicity was smaller [around 0.6 units for $pK_a(k_{cat})$]. However, E230's pK_a for the original design (approximately 6.0) is comparable to the best evolved KE07 variants, thus accounting for the design being a priori more reactive ($k_{cat}/K_m \geq 160$ $M^{-1}s^{-1}$ versus approximately 10 for KE07).

To probe the ability of KE59 to act on nonactivated substrates, the rates of four representative KE59 variants were measured for a series of benzisoxazoles with decreasing reactivity (pK_a^{LG} 4.3–6.9; see *SI Appendix, Fig. S8* and Table S5). The Brønsted plots demonstrate that substrate structure seems to be a determining factor for both k_{cat} and K_m . Certain substituents enable a more effective positioning than others. For example, all variants exhibited the highest activity with 5,7-dichloro BI, and not with 5-nitro BI. Conversely, the k_{cat} values for 5-nitro-6-chloro BI and unsubstituted benzisoxazole are lower than expected for their pK_a^{LG} (see *SI Appendix, Fig. S8A*). The 6-nitro substituent in 6-nitro BI seems to lead to misalignment with the base, whereas the unsubstituted benzisoxazole might be too loosely bound. The misalignment and/or loose alignment seem to be manifested in low k_{cat} values and high K_m values, and also with high K_i values with the corresponding benzotriazole inhibitors (see *SI Appendix,*

Table 2. pK_a and β_{LG} values of representative KE59 variants

KE variant	$pK_a(k_{cat})/pK_a(k_{cat}/K_M)$	$\beta_{LG}(k_{cat})/\beta_{LG}(k_{cat}/K_M)$
R2-4/3D	6.1/6.3	−1.06/ −1.76
R4-5/11B	5.5/6.7	−0.95/ −1.35
R8-2/7A	6.2/6.5	
R9-1/4A	6.5/6.5	−0.56/ −1.23
R13-3/11H	6.7/6.5	−0.68/ −1.26

Tables S5 and S6). Substrate alignment problems result in not only suboptimal positioning relative to E230 but also in more access for water molecules that disrupt the aprotic environment of the active site and reduce E230's basicity.

The irregularities for various subsituents put aside, the negative Brønsted slopes indicate a change in the active-site environment of the evolved KE59 variants. The β_{LG} (k_{cat}) values of R2 and R4 variants are similar to the β_{LG} previously obtained for acetate in acetonitrile (-0.95 , Table 2; see *SI Appendix*, Fig. S8A) (8). This implies that in the active sites of the early KE59 variants, the dispersion of the TS's negative charge is as ineffective as in acetonitrile. The β_{LG} (k_{cat}) values for R9 and R13 variants (Table 2) are closer to the β_{LG} for acetate in water (-0.67). The Brønsted plots for k_{cat}/K_m (See *SI Appendix*, Fig. S8B) exhibit more negative slopes than the k_{cat} plots because of the general tendency of K_m to increase with pK_a^{LG} , but indicate a similar trend. The decrease in β_{LG} values correlates with the evolution of the R9 to R13 variants towards the elimination of less activated substrates, and suggests that their active sites are better adjusted for the stabilization of the developing negative charge. Although the KE59 starting point lacked a specifically designed proton donor (12), a suitably placed water molecule could fulfill this role (30). Indeed, in some crystal structures of KE59 variants, water molecules solvating the benzotriazole ligand at relevant positions can be observed (see *SI Appendix*, Fig. S9).

Structural Analysis. Structures of several KE59 variants were solved at 1.45–2.50 Å resolution (see *SI Appendix*, Table S7). Structural properties of the unstable KE59 design were examined using variant R1-7/10H that contains only stabilizing mutations and has similar kinetic parameters (Table 1). We obtained the apo structure of R1-7/10H and of its complex with 5,7-dichlorobenzotriazole, the apo structures of R5-11/5F and R8-2/7A variants, and the structures of R13-3/11H with benzotriazole and 5,7-dichlorobenzotriazole (see *SI Appendix*, Fig. S10).

The backbones of the KE59 computational model and of the actual structures overlap quite well (See *SI Appendix*, Fig. S11 and Table S8). The main deviations occurred in the long loops (residues 51–65, and 179–190). The active-site entrance of the most advanced R13-3/11H variant is wider than in the design-like variant R1-7/10H, primarily due to R181H mutation and the movement of loop 51–65 (see *SI Appendix*, Fig. S12A and B). No major changes in the configuration of the bottom part of the active site were observed despite several mutations in this region (see *SI Appendix*, Fig. S12C and D).

At the level of side-chain rotamers, however, the active sites of the evolved KE59 variants differ from the designed model. In the apo structures of KE59 variants, the stacking W109 adopts rotamers different from the designed, and these are incompatible with substrate binding (Fig. 2A). Similar “flipping” of the stacking Trp was also observed in the apo structures of catalytic antibody 34E4 (31) and of KE design HG-1 (32). In the structures of KE59 variants with benzotriazole ligands, the position of W109 is more similar to the designed rotamer (Fig. 2B). It is therefore likely that in the presence of substrate, W109 adopts a rotamer similar to the modeled one. The catalytic base E230 adopts similar rotamers in the designed model and in the crystal structures, with some carboxylate rotation (Fig. 2A and B). However, MD simulations show that E230 has a strong preference for alternative conformations that are less catalytically competent. The glutamate side chain can extend towards the bulk solvent with the carboxylate forming hydrogen bonds with the backbone-NH of S210 and with solvent molecules (Fig. 2C). In the case of R13-3/11H, the hydrogen bond between E230 and W109 stabilizes the designed configuration and draws the carboxylate oxygens away from S210 (Fig. 2B). This configuration coincides with higher catalytic efficiency of the R13 variant (see *SI Appendix*, Fig. S13)

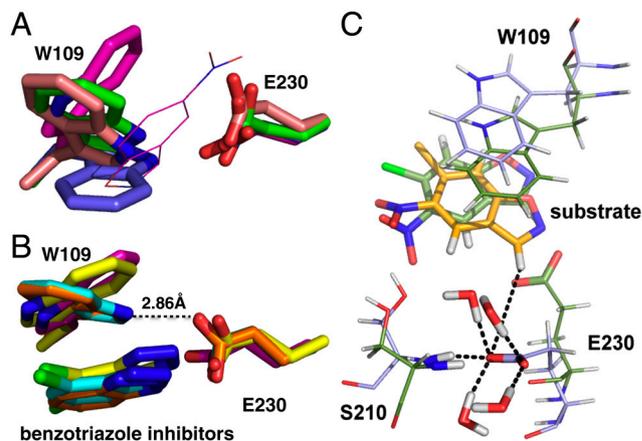


Fig. 2. W109 and E230 rotamers in the KE59 structures. (A) W109 and E230 rotamers in the KE59 model (magenta) and in the apo structures of R1-7/10H (pink), R5-11/5F (green), and R8-2/7A (violet). The 5-nitro BI (magenta) is overlaid from the KE59 model. (B) W109 and E230 rotamers in the KE59 model (magenta) and in the structures of R1-7/10H with 5,7-dichlorobenzotriazole (yellow), R13-3/11H with benzotriazole (orange) and with 5,7-dichlorobenzotriazole (cyan). (C) The predominant conformation of E230 observed in MD simulations (blue) vs. the conformation observed in the crystal structures (green), shown for R13-3/11H variant.

despite the possible unfavorable effect in lowering the basicity of E230 (see *SI Appendix*, Fig. S14).

The benzotriazole ligands in the complex structures are stacked well against W109. However, their orientation is flipped relative to the designed binding mode with the triazole nitrogens facing the top of the active site (Fig. 3). This alternative binding mode does not seem to be the outcome of directed evolution, because the benzotriazole ligands in R1 and R13 structures bind in the same manner. Both binding modes could be productive, as in both cases E230 carboxylate faces the heterocycle's hydrogen (Fig. 3). MD simulations show that both binding modes are equally likely (see *SI Appendix*, Fig. S15), and it is therefore difficult to estimate which of them is more populated. In addition, the mode of binding of the benzotriazole inhibitors may not reflect the catalytic mode. Their binding affinities are much lower than expected from rate accelerations (see *SI Appendix*, Table S6). Unlike the actual TS, benzotriazoles possess a complete positive charge ($pK_a = 8.37$) distributed all around the heterocycle ring. The benzotriazoles' binding mode allows this charge to be more optimally solvated by facing bulk solvent and E230 and may not necessarily mimic the substrate mode of binding.

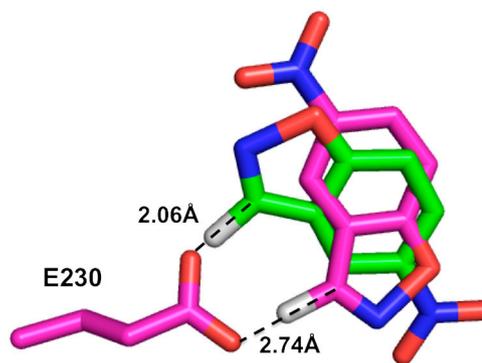


Fig. 3. Comparison of alternative substrate binding modes. KE59 model, showing the E230 and the two binding modes of 5-nitro BI: the designed (magenta) and the alternative (green), based on the benzotriazole orientation in the KE59 structures.

Desolvation and Base Positioning Effects. The increase in the catalytic power of E230 can arise from two sources: activating medium effects, and/or improved positioning of the catalytic base. Both factors play a role in catalysis of the Kemp elimination, but their relative contribution is not easily quantifiable by kinetic measurements (7–9). Thus, MD simulations were used to monitor the solvent accessibility of E230 and assess the impact of base desolvation (medium effect) (33). The simulations suggest that as the KE59 variants become more active, fewer bulk solvent water molecules are observed in direct contact with the catalytic carboxylate in the substrate complex. In addition, substituents at the 5 and 7 positions were found to better shield E230 from bulk solvent than substituents at positions 4 and 6, accounting for the higher k_{cat} values of the former (see *SI Appendix, Fig. S16*).

Although the evolutionary process seems to have disfavored noncatalytic configurations of E230, no correlation is observed between k_{cat} and the relative positioning of E230 versus the substrate (see *SI Appendix, Fig. S17*). Improved positioning can certainly result in higher rates, as it was observed for the catalytic His-Asp dyad in the case of KE70 (17). However, in the case of KE59, no polarizing contact is needed to activate the E230 carboxylate, because its catalytic strength stems from embedment in a hydrophobic pocket with no contacts that could favor a particular rotamer. Further, unlike in the case of KE70, where directed evolution gave rise to an increasingly rigid and precisely positioned active site (17), the KE59 series shows no such trend. The atomic fluctuation profiles remain similar throughout and the active site residues even gain mobility (see *SI Appendix, Fig. S18*). Because in the course of directed evolution the KE59 variants were screened with benzisoxazoles with substituents of various sizes and positions, precise and tight substrate positioning may not be expected.

Discussion

Directed evolution improved the catalytic efficiency of KE59 with 5-nitro BI, and with less activated substrates, by $\leq 2,000$ -fold (Table 1). The catalytic efficiency with the best substrate, 5,7-dichloro BI, reached $0.57 \times 10^6 \text{ s}^{-1} \text{ M}^{-1}$, and the rate acceleration ($k_{\text{cat}}/k_{\text{uncat}}$) for all substrates became $\sim 10^7$ (see *SI Appendix, Table S11*). While the evolved catalytic efficiency compares well with the average k_{cat}/K_m value for natural enzymes (15), the activity with less activated substrates (34) still falls behind.

The Evolvability of Designed Proteins. Optimization of three designed Kemp eliminases, KE07, KE59, and KE70, based on different templates and different active-site configurations, supports the notion that designed proteins are evolvable. The catalytic efficiencies of these designed-evolved eliminases compare very favorably to other enzyme-models, including catalytic antibodies with Kemp eliminase activity (2, 4, 6, 7) (see *SI Appendix, Table S11*). The clear advantage of the designed Kemp eliminases is their evolvability, which allowed increasing the activity by three orders of magnitude. Computationally designed starting points might therefore be as valuable as promiscuous activities in natural enzymes, especially for obtaining catalysts for unnatural substrates and reactions. Enzyme evolvability is largely driven by active site “floppiness”—alternative binding modes of substrates and alternative active site conformations (35). Such floppiness is observed in the case KE59, both in the structures and MD simulations. Thus, although floppy designs exhibit lower activity, they might also be evolvable for the very same reason: for example, the alternative W109 rotamer and the resulting interaction with E230 may have led to a new and more effective active site configuration of KE59.

We also noted a correlation between the initial activity of the KE design and the final activity of its most active evolved variant: the designs KE07, KE70, and KE59 exhibited k_{cat}/K_m values of 12, 126 and $\sim 160 \text{ M}^{-1} \text{ s}^{-1}$, respectively, and the k_{cat}/K_m values of

their best evolved variants reached 2.6×10^3 , 5.5×10^4 , and $5.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$. This trend implies that relatively modest improvements in the design methodologies may give rise to efficient catalysts upon directed evolution.

Stability Promotes Evolvability. The marginal stability of proteins restricts the accumulation of mutations, and especially of mutations that modulate protein function (19, 20). KE07 and KE70 designs possessed an excess of stability that enabled mutations to accumulate (16). However, as is the case with many natural proteins, the stability of KE59 was limited, and so was its evolvability—no improved variants were observed in libraries of random mutations. Consensus mutations comprise the most easily identifiable potential of stabilizing mutations that can boost protein evolvability (25). As less than half of the predicted consensus mutations actually have a stabilizing effect (24, 36), and different mutations become compensatory at different evolutionary stages (as seen in R1-R2 versus R7 and later), spiking these mutations into libraries is a more effective strategy than the use of a pre-stabilized starting point (20).

Design and Mechanistic Insights. A key contribution to efficient catalysis of the Kemp elimination is an activated, high pK_a base, which can be achieved by placing a carboxylate in an aprotic, dipolar environment (8, 29). Indeed, KE59 was the most active amongst eight parallel designs, and >10 -fold more efficient than KE07 in which the pK_a of the catalytic base (E101, $pK_a \leq 4.5$) was similar to that of Glu's γ -carboxylate in solution (16). Directed evolution of KE07 led to detachment of a lysine that hydrogen-bonded E101, and increased the pK_a of E101 to approximately 6. An activating environment was a priori achieved in the KE59 design (E230 $pK_a \leq 6$), and further refined in its evolved variants ($pK_a \leq 6.7$). Better shielding of E230 from solvent molecules was a major identifiable contribution to the increase in catalytic activity.

Efficient desolvation also underlines the best substrates. The less reactive substrates have either small or no substituents and show not only low k_{cat} values but also high K_m and K_i values with the corresponding benzotriazoles (see *SI Appendix, Tables S5 and S6*). Thus, in addition to their low intrinsic reactivity, these substrates exhibit loose positioning and ineffective desolvation of the catalytic base. This, and the possibility of alternative catalytic modes, suggests that the KE59 design is floppy. The case of KE59, and of other designs where the actual mode of binding fundamentally differs from the designed one (37), demonstrate weaknesses of the current design methodologies in designing an active-site with perfect complementarity to the reaction's TS. Surprisingly, catalysis and binding are executed by the designed residues, but not in the designed mode. Such floppiness may be tolerable, or even advantageous in cases of activated reactions such as the Kemp elimination, where precise positioning is not a key factor (4), but not for catalysis of demanding reactions. If better catalysts are to be designed, the design must yield active sites with tighter, unambiguous binding modes. Part of the problem lies in the choice of substrates: small, nonpolar substrates present a challenge, both for designed and natural enzymes. Designs based on large substrates and on substrates with readily recognizable groups such as phosphate (15) are therefore likely to yield more efficient catalysts. More effective designs should also consider the computation of substrate floppiness, of pK_a values of key residues, and of water molecules. Water should be largely but not totally excluded from the enzyme-TS complex. Appropriately positioned waters can exert efficient catalysis (38), such as protonation of the phenoxide leaving group in Kemp elimination (30); see *SI Appendix, Fig. S8*.

Overall, our current knowledge seems to be sufficient not only to explain how natural enzymes work but also to make new enzymes. However, even for a simple and well-defined reaction such

as Kemp elimination, designing highly efficient active-site configurations, and predicting sequences that deliver these configurations at sub-Ångström accuracy, remain a challenge.

Materials and Methods

Substrates/Inhibitors. Preparation of benzisoxazoles and benzotriazoles, and their properties are summarized in the *SI Appendix*.

Cloning and Library Making. The genetic diversity in KE59 genes was generated by: error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit, Stratagene) (39), gene shuffling (40), and mutations incorporation by spiking oligonucleotides during the assembly of DNA fragments (26). After mutagenesis and/or shuffling, the KE59 genes were recloned into the original pET29b plasmid (16).

Screening Procedure. The libraries were transformed into *E. coli* BL21 (DE3) cells ($>10^5$ transformants per library) and screened as previously described (16), with the following variations: the cultures were grown at 30 °C before the induction with IPTG and for 24–28 h at 20–25 °C after the induction. At Rounds 14–16, the libraries were transformed into *E. coli* BL21 (DE3) cells containing the pGro7 (GroEL/GroES), pBAD33-GroEL (GroEL) plasmid (28). The lysates were assayed with 5-nitro BI (0.125 mM), 6-chloro BI (0.25 mM), or 5-fluoro BI (0.25 mM) by following the release of the phenol product (Power HT microtiter scanning spectrophotometer). Only 600–2,000 clones were screened at each directed evolution round and the load of random mutations was kept accordingly low.

- Kirby AJ (1996) Enzyme mechanisms, models, and mimics. *Angew Chem Int Ed Engl* 35:707–724.
- Debler EW, et al. (2005) Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc Natl Acad Sci USA* 102:4984–4989.
- Hollfelder F, Kirby AJ, Tawfik DS, Kikuchi K, Hilvert D (2000) Characterization of proton-transfer catalysis by serum albumins. *J Am Chem Soc* 122:1022–1029.
- Hollfelder F, Kirby AJ, Tawfik DS (1996) Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* 383:60–62.
- Hollfelder F, Tawfik D (1997) Efficient catalysis of proton transfer by synzymes. *J Am Chem Soc* 119:9578–9579.
- Korenovych IV, et al. (2011) Design of a switchable eliminase. *Proc Natl Acad Sci USA* 108:6823–6827.
- Thorn SN, Daniels RG, Auditor MT, Hilvert D (1995) Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* 373:228–230.
- Hu Y, Houk KN, 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.
- Hollfelder F, Kirby AJ, Tawfik DS (2001) On the magnitude and specificity of medium effects in enzyme-like catalysts for proton transfer. *J Org Chem* 66:5866–5874.
- Das R, Baker D (2008) Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:363–382.
- Jiang L, et al. (2008) De novo computational design of retro-aldol enzymes. *Science* 319:1387–1391.
- Rothlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195.
- Baker D (2010) An exciting but challenging road ahead for computational enzyme design. *Protein Sci* 19:1817–1819.
- Zanghellini A, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15:2785–2794.
- Bar-Even A, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–4410.
- Khersonsky O, et al. (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 396:1025–1042.
- Khersonsky O, et al. (2011) Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 407:391–412.
- Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320:85–95.
- Tokuriki N, Stricher F, Serrano L, Tawfik DS (2008) How protein stability and new functions trade off. *PLoS Comput Biol* 4:e1000002.
- Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103:5869–5874.
- Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–932.
- Steipe B, Schiller B, Pluckthun A, Steinbacher S (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol* 240:188–192.
- Lehmann M, Pasamontes L, Lassen SF, Wyss M (2000) The consensus concept for thermostability engineering of proteins. *Biochim Biophys Acta* 1543:408–415.
- Bommarius AS, Broering JM, Chaparro-Riggers JF, Polizzi KM (2006) High-throughput screening for enhanced protein stability. *Curr Opin Biotechnol* 17:606–610.
- Bershtein S, Goldin K, Tawfik DS (2008) Intense neutral drifts yield robust and evolvable consensus proteins. *J Mol Biol* 379:1029–1044.
- Herman A, Tawfik DS (2007) Incorporating synthetic oligonucleotides via gene reassembly (ISOR): A versatile tool for generating targeted libraries. *Protein Eng Des Sel* 20:219–226.
- Godoy-Ruiz R, et al. (2006) Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *J Mol Biol* 362:966–978.
- Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459:668–673.
- Kemp DS, Cox DD, Paul KG (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.
- Debler EW, Muller R, Hilvert D, Wilson IA (2009) An aspartate and a water molecule mediate efficient acid-base catalysis in a tailored antibody pocket. *Proc Natl Acad Sci USA* 106:18539–18544.
- Debler EW, Muller R, Hilvert D, Wilson IA (2008) Conformational isomerism can limit antibody catalysis. *J Biol Chem* 283:16554–16560.
- Privett HK, et al. (2011) An iterative, rational approach to computational enzyme design. *Proc Natl Acad Sci USA* 109:3790–3795.
- Kiss G, Rothlisberger D, Baker D, Houk KN (2010) Evaluation and ranking of enzyme designs. *Protein Sci* 19:1960–1973.
- Wolfenden R, Snider MJ (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res* 34:938–945.
- Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. *Science* 324:203–207.
- Polizzi KM, Chaparro-Riggers JF, Vazquez-Figueroa E, Bommarius AS (2006) Structure-guided consensus approach to create a more thermostable penicillin G acylase. *Bio-technol J* 1:531–536.
- Karanicolas J, et al. (2011) A de novo protein binding pair by computational design and directed evolution. *Mol Cell* 42:250–260.
- Fersht A (1999) *Structure and Mechanism in Protein Science* (WH. Freeman and Company, New York).
- Barlow M, Hall BG (2002) Predicting evolutionary potential: In vitro evolution accurately reproduces natural evolution of the tem beta-lactamase. *Genetics* 160:823–832.
- Abecassis V, Pompon D, Truan G (2000) 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.