Computational Optimization of Electric Fields for Improving Catalysis of a Designed Kemp Eliminase

Valerie Vaissier[1,2,5,†], Sudhir C. Sharma[2†], Karl Schaettle[1,3], Taoran Zhang[3],

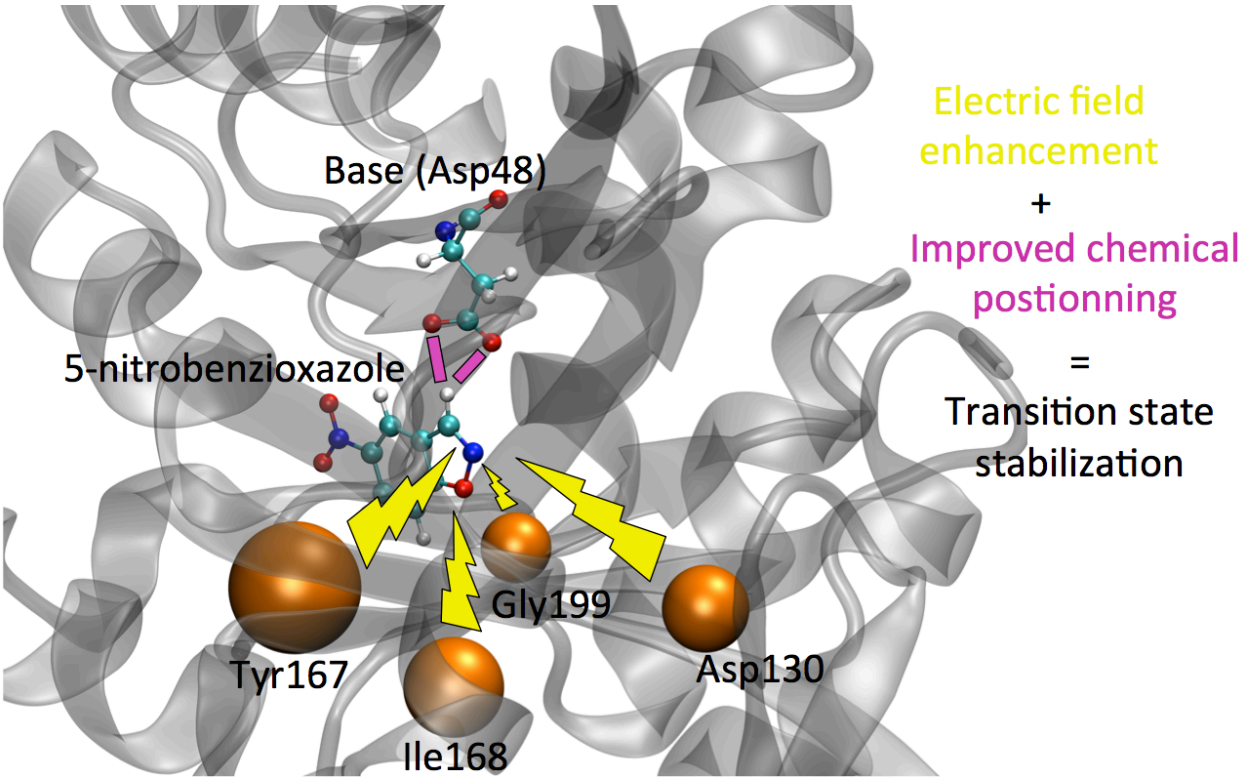
Teresa Head-Gordon[1,2,3,4,5,*]

[1]Kenneth S. Pitzer Center for Theoretical Chemistry, [2]Department of Chemistry, University of California, Berkeley
[3]Department of Chemical and Biomolecular Engineering and [4]Department of Bioengineering, University of California Berkeley
[5]Chemical Sciences Division, Lawrence Berkeley National Labs
Berkeley, California 94720, USA

Here we report a computational method to improve efficiency of a *de novo* designed Kemp Eliminase enzyme KE15, by identifying mutations that enhance electric fields and chemical positioning of the substrate that contribute to free energy stabilization of the transition state. Starting from the design that has a k_{cat}/K_M of 27 $M^{-1}s^{-1}$, the most improved variant introduced 4 computationally targeted mutations to yield a k_{cat}/K_M of 403 $M^{-1}s^{-1}$, with almost all of the enzyme improvement realized through a 43-fold improvement in k_{cat}, indicative of a direct impact on the chemical step. This work raises the prospect of computationally designing enzymes that achieve better efficiency with more minimal experimental intervention using electric field optimization as guidance.

[†]authors contributed equally
[*]Corresponding author: thg@berkeley.edu

Keywords: enzyme, catalysis, laboratory directed evolution, Kemp Eliminase, electric fields, chemical positioning


Base (Asp48)
5-nitrobenzioxazole
Gly199
Tyr167
Asp130
Ile168
Electric field enhancement
+
Improved chemical postionning
=
Transition state stabilization

TOC graphic

INTRODUCTION

De novo designed enzymes via computer modeling of active sites accommodated within a variety of protein scaffolds[1-2], have been reported for reactions such Kemp elimination[3-6], retro-aldol condensation[7] and Diels-Alders chemistry[8]. However, most computational designs to date have exhibited very little catalytic competence compared to native enzymes or enzymes subsequently improved through laboratory directed evolution (LDE). A case in point is the Kemp elimination reaction, a one-step proton transfer reaction from the 5-nitrobenzisoxazole substrate by a catalytic base, leading to breaking of the 5-membered ring and forming the final product, alpha-cyanophenol (Figure 1a). Various Kemp eliminases that were computationally designed catalyze this reaction with efficiencies as measured by k_{cat}/K_M of 12, 126, 160, and 425 $M^{-1}s^{-1}$, for KE07[5], KE70[6], KE59[4], and HG3[9], respectively. However substantial gain in biocatalytic activity is achieved when these minimal designs are subjected to laboratory directed evolution LDE; after undergoing LDE, the Kemp eliminases KE07[5], KE70[6], KE59[4], and HG3[10] yielded k_{cat}/K_M values of $2.0x10^3$, $5.7x10^4$, $6.0x10^5$, and $2.3x10^6$ $M^{-1}s^{-1}$, respectively, which has led to much optimism for achieving efficiencies that rival natural enzymes.

However, there is still great appeal for replacing the labor-intensive and opaque LDE optimization process[11-12] with a rational approach using computation to create more competent enzymes, or at least by starting the LDE process with a much more improved enzyme by computational design. There are a number of possible concepts that can be used to rationally explain the LDE process and/or which can be used for further improving designed enzymes.[13] For example, Frushicheva et. al[14-16] and Labas et. al[17] used simulations based on the empirical valence bond method to show that LDE improved KE07 via ground state destabilization and reorganization energy optimization. DA_20_10, an enzyme that catalyzes the Diers-Alders reaction, was further improved by crowdsourcing it to FoldIt players that led to a remodeling of the backbone and an 18-fold improvement in activity[18], almost all of which was due to improvements in K_M. Others have used computational approaches to augment the construction of LDE libraries, using backrub motion[19], loop redesign[20], and consensus mutations[21-22]. Mayo and co-workers have developed computational techniques that uses a genetic algorithm framework to propose mutations that help avoid unstable folds by conserving the number of contacts that are broken/made before and after doing the proposed shuffling LDE experiments[23]. Other attempts at predicting side directed mutations have tried to minimize absolute entropy of the mutation site[24]

to improve stability. Privett and co-workers utilized an iterative approach of creating a new active site deeper in the scaffold of a designed Kemp Eliminase that is better shielded from solvent, thereby increasing the hydrophobicity of the binding pocket to enhance the pK_a of the catalytic base, and use of molecular dynamics[25] to diagnose problems with side chain orientations that were poorly preorganized[9]. Finally, Warshel and coworkers were the first to identify the preorganization of the electrostatic free energy as a key component of enzyme catalysis[26-28]; in fact many groups have reported evidence for how electric fields contribute to the catalytic power of various natural enzymes.[29-33]

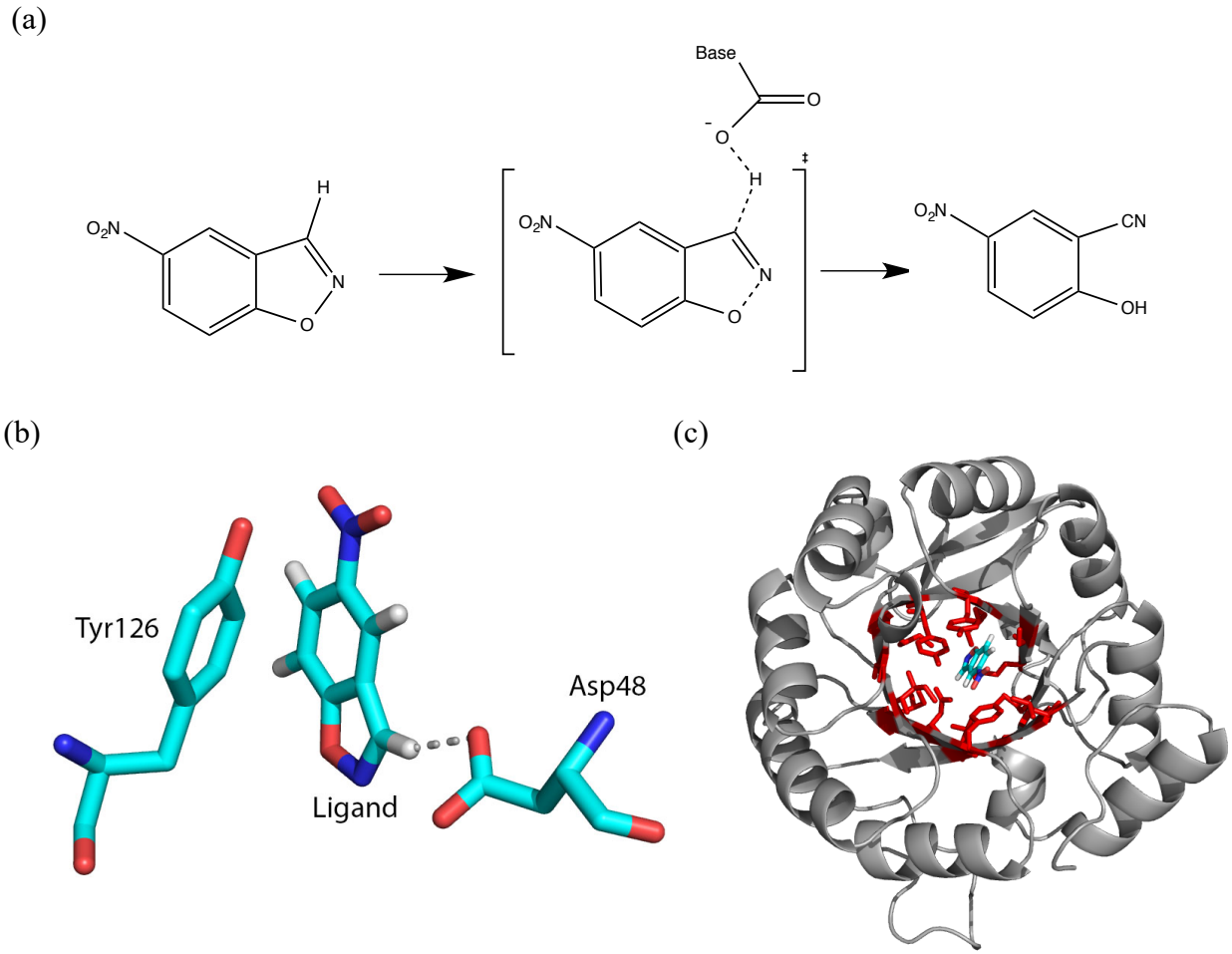


Figure 1: *The Kemp Elimination reaction and designed KE15 enzyme*. (a) The one-step reaction scheme involving the abstraction of hydrogen from the carbon of 5-nitrobenzisoxazole by a catalytic base. Shown is the transition state that has a partial negative charge on the substrate oxygen with cleavage of the O-N bond and nascent formation of a C≡N triple bond (Reprinted with permission from [34] Copyright 2017 American Chemical Society.) (b) Active site of KE15 design utilizing catalytic base Asp-48 and Tyr-126 as π-stacking residue for the residue engineered into the TIM barrel scaffold (PDB ID: 1THF). (c) View of the overall KE15 enzyme. The active site and other designed residues are shown in red with substrate in cyan.[35]

We have recently reported in two separate studies the effects of both side chain conformational entropy and mutual information[36], as well as electrostatic stabilization[34], for improving designed KE07 and KE70 enzymes during the LDE process. It was shown in the first study that improved variants used a combination of free energy destabilization of the reactant enzyme-ligand bound state (EL) (in agreement with Frushicheva et. al[14-16] and Labas et. al[17]) and stabilization of the transition state ($EL^{\dagger}$) to enhance activity, with significant contributions from side chain entropy that supported this trend[36]. In the second study, we demonstrated how electric field stabilization of the three bonds of the substrate that are broken and formed to create the product was greatly improved in the evolved KE07 enzyme, while for KE70 the LDE strategy evolved toward mutations that created a more hydrophobic active site instead[34]. In our study on KE07 and KE70, a breakdown of contribution of the electric field from individual residues further revealed that chemical positioning of the catalytic base and active site residues in the immediate vicinity of the substrate were the primary source of electrostatic improvement for both enzymes. In contrast, the solvent and the remaining scaffold provided electrostatic environments that were detrimental to the active site chemistry for KE07 and KE70, unlike native enzymes[34].

In this paper, we apply the understanding gained from these previous studies to improve the *de novo* enzyme KE15 using only computational predictions. Since no LDE was performed on KE15, it serves as a good test system for improving catalytic efficiency in a rational and systematic way for a *de novo* enzyme designed with the standard Rosetta protocol used for the previous KE07, KE70, KE59, and HG designed enzymes. The active site is built around the catalytic base Asp-48 and a Tyr-126 π-stacking residue to orient the ligand (Figure 1b), which is introduced into a TIM barrel scaffold, with 13 additional design mutations made to accommodate the new active site and stabilize the enzyme (Figure 1c). Experimentally, the designed enzyme was found to have a k_{cat}/K_M of 27 $M^{-1}s^{-1}$ where k_{cat}=0.007 s^{-1} and K_M = 270 μM.

Starting with the KE15 design, we have screened mutants *in silico*[35], primarily targeted by improvements in electrostatic stabilization of the transition state and better chemical positioning of the base in the active site. We tested our predictions experimentally, and with 4 mutations found the best variant that yielded an overall 43X increase in k_{cat} =0.31 s^{-1} with some degradation in K_M =767 μM, leading to 15X improvement of k_{cat}/K_M =403 $M^{-1}s^{-1}$. To put this in perspective, the improvement in k_{cat} for KE15 is equivalent to 6-7 rounds of LDE for KE07 (in

which k_{cat} improved 70X) and KE70 (in which k_{cat} improved 35X). We show from our calculations that while some of the improvement came from destabilization of the reactant state, most improvement came from stabilization of the transition state, as intended.

COMPUTATIONAL METHODS

Our computational screening strategy to determine beneficial mutants for designed enzymes makes use of electric field calculations using short MD trajectories with the AMOEBA force field. We only give a brief description here, and more specific details for each of the simulation techniques have been provided in a previous publication[34].

Sampling approaches to creating structural ensembles for KE15. The designed KE15 starting structure with docked ligand was provided by the Houk group (UCLA), and computational mutations were built using Modeler[37]. For each sequence variant, we computed 10,000 trial moves to create each 25 uncorrelated and low energy backbone structures using backrub simulations provided within the Rosetta modeling software package[38-40]. To generate additional structural ensemble diversity, we used our recently developed Monte Carlo method (MC-SCE)[41], for which we have performed extensive validation across ~60 proteins through comparison to high quality X-ray crystallography data and NMR experiments, to characterize the side chain structural ensemble on the 25 backbone structures for KE15. Briefly, the method uses a Rosenbluth side chain growth protocol sampled from an expanded side chain rotamer library[42-43], which are energy weighted according to a physical energy function based on the AMBERff99SB protein force field combined with a GB-HPMF implicit solvent model; backbone variability is introduced through a backrub algorithm[38-40].

We ran additional MD simulations of the enzyme and docked ligand (EL state) using OpenMM[44] and the AMOEBA polarizable force field[45-46] in the NVT ensemble (T=300K, 1 fs timestep). The system was solvated using Gromacs[47-48] with a pre-equilibrated water box (70 Å × 70 Å × 70 Å). After a 40 ps equilibration time, we collected snapshots every 50 fs across a 10ps production run (200 snapshots for each of the 25 structures so 5000 frames per mutant per state). The binding energy, E_{bind}, was calculated as follows:

$$E_{bind} = < E_{enzyme+ligand} - E_{enzyme} - E_{ligand} >_{NVT} \quad (1)$$

where $E_{enzyme+ligand}$, E_{enzyme} and E_{ligand} are the MM energies of the ligand bound enzyme, enzyme and ligand respectively. Each energy term was found by minimizing the energy of the snapshots in the OBC2 implicit solvent model in OpenMM.

Electric field calculations. Using Gromacs[47], the 25 lowest energy structures were explicitly solvated using a pre-equilibrated water box. We then performed 50 ps equilibration and then 50 ps production runs in the NPT ensemble using the Tinker software package[49-51] and the AMOEBA polarizable force field[45, 49] to provide a high quality description of electrostatics in the active site and overall scaffold and solvent. The substrate geometry for the transition state was the same as in the reactant state, but the atomic multipoles of the transition state are different than the reactant state and were taken from AMOEBA polarization calculations reported in [34]; since this study focuses primarily on electrostatic effects this is an acceptable transition state model. In some of our simulations, we restrained the ligand in place using 2 harmonic potentials with 1000 N/m spring constants, between the carbon of the carbolic acid group of the base and the (i) H (2.0-3.0 Å) and (ii) N (4.0-5.0 Å) atoms of the ligand. Electric fields are calculated every picosecond over the 50ps production run at each of the 4 atoms that make up the 3 bonds of the substrate that are most changed on going from the reactant to transition state (Figure 2). The electric field is projected onto each bond and the value reported is given by the mean of the electric field at the two atoms involved[34].

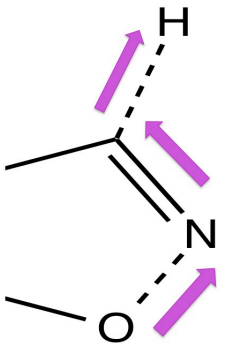


Figure 2: *Electric field projection onto the C-H, C-N, and O-N bond dipoles of 5-nitrobenzisoxazole and sign convention used.* Electric fields are calculated at the C, H, N and O of the ligand, in which the critical chemical step of the reaction is the breaking of the C-H and O-N bonds and the making of the C-N triple bond. Reprinted with permission from [34] Copyright 2017 American Chemical Society.

EXPERIMENTAL METHODS

The 5-nitrobenzisoxazole ligand was synthesized by following an earlier published method[52], and its improved version from the Hilvert laboratory[53]. The KE15 plasmids were kindly provided by the David Baker laboratory at University of Washington, Seattle, WA, and variants studied in this work were generated by site-directed mutagenesis using a Quik Change II site-directed mutagenesis kit (Stratagene; Agilent Technologies, Santa Clara) using appropriate PCR primers. After the mutagenesis PCR reactions, the mutated plasmids were transformed into XL-10 gold cells and the plasmids encoding individual mutations were isolated. The identity of the mutated plasmids were confirmed by sequencing the plasmid from both forward and reverse directions using T7 forward and T7 reverse primers at UC Berkeley Sequencing facility. The individual mutated plasmids were transformed into expression cell line BL21 (DE3) gold.

A single colony from the transformed cells containing individual variant was used to inoculate a starter culture of 20 mL LB medium supplemented with 50 μg/mL kanamycin and the resulting culture incubated with shaking overnight at 37°C. This starter culture was used to inoculate 500 mL LB medium with 50 μg/mL kanamycin and incubated for ~3h at 37°C until OD600 reached ~1.2. The culture was then induced with 1mM IPTG for overproduction and the culture was further grown with shaking at 37°C for 4h. The cells from the liquid culture were harvested and stored at -80°C until used for the isolation. In general, roughly 2 g of the wet cells were routinely obtained from 0.5L culture.

The harvested cells were thawed, re-suspended in 35 mL lysis buffer (25 mM Hepes, pH 7.25 containing 100 mM NaCl, 5% glycerol), lysed by sonication, centrifuged to remove insoluble debris and the soluble fraction loaded into pre-washed Ni-NTA column (5mL resin, His-Pur, Thermo-Fisher). The Ni-NTA resin with the bound proteins were washed first with 10 column volume of lysis buffer followed by 15 column volume of 20 mM NaPi, pH 7.4, 500 mM NaCl, 30 mM Imidazole to remove nonspecific and weakly bound proteins. The bound His-tagged fusion protein was then eluted from the Ni-NTA resin with 20-25 mL of 500mM Imidazole buffer solution (20 mM NaPi pH 8.0, 500 mM NaCl, 500 mM Imidazole). The eluted fusion protein were extensively dialysed in lysis buffer, concentrated through Amicon filters (30,000 MWCO, Millipore), its concentration estimated by measuring the absorbances at 280 nm and stored at -80°C in smaller aliquots. This purification protocol yielded over 90% pure protein

(assessed through the visible bands in SDS-PAGE) and routinely produced 10-20 mg of His-tagged KE15 proteins.

The enzymatic characterization of the KE15 variants was performed similar to previously published work[5] with some modification in the Cary 50 spectrophotometer (Varian) that used a quarz cuvette. In short, the kinetic analysis were performed in 25 mM Hepes, pH 7.25, 100 mM NaCl, 5% glycerol with 5-nitrobenzisoxazole concentration ranging from 5-1500 μM with the co-solvent acetonitrile concentration equalized to 1.5% (v/v) in a micro-cuvette capable of monitoring reaction at 200 μL. A known amount of dry 5-nitroxybenzisoxazole was dissolved in acetonitrile to have 100mM substrate stock. From this stock a series of dilutions of the substrate were made in acetonitrile to achieve the concentration ranges in the kinetic assay. The reaction was initiated by the addition of small amount of the enzyme aliquot (final concentration from 0.5-1.5 μM in the assay) and the product formation was monitored spectrophotometrically at 380 nm ($\Delta\varepsilon$ = 15,800 M^{-1}, cm^{-1}). Steady-state parameters were obtained after fitting the data to the Michelis-Menten equation.

RESULTS

Similar to other computationally designed enzymes such as KE07 and KE70, the efficiency of the KE15 design is quite poor (Table 1). To improve the catalytic activity of the KE15 enzyme, we want to identify sites for possible mutations that would stabilize the transition state through better electric fields at some or all of the bonds under the convention defined in Figure 2.

Table 1*: Kinetics data determined experimentally for the KE15 enzyme design and mutants predicted from electric field and chemical positioning calculations.* The most beneficial mutation from each iteration was added to move to the next round of screening. Only a handful of mutations were tested in the lab, dramatically reducing the labor that usually goes into improving designed enzymes.

Round	Enzyme Mutant	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
1	KE15 Design	0.0072 ± 0.0004	270 ± 39	26.7 ± 5.3
2	Asp130Lys+R1	0.0338 ± 0.002	574 ± 88	58.9 ± 12.5
3	Ile168Met+R2	0.0359 ± 0.00076	227 ± 16	158.1 ± 14.5
4	Gly199Ala+R3	0.1059 ± 0.0038	397 ± 39	266.7 ± 35.7
5	Tyr167Lys+R4	0.30908 ± 0.036	767 ± 188	403.0 ± 145.7

As we have shown previously[34], the electrostatic free energy of stabilization of the transition state, $\Delta G^{\neq}_{elec}$ going from the EL to EL† state can be calculated using Eq. (2)

$$\Delta G^{\neq}_{elec} = -0.048(\bar{\mu}_{EL^{\dagger}} \cdot \bar{E}_{EL^{\dagger}} - \bar{\mu}_{EL} \cdot \bar{E}_{EL}) \quad (2)$$

where $\bar{E}_{EL\dagger}$, $\bar{\mu}_{EL\dagger}$, $\bar{E}_{EL}$, and $\bar{\mu}_{EL}$ is the electric field (in MV/cm) and bond dipole moment (in Debye) in the transition state (EL†) and ligand bound state (EL), respectively, and the factor of 0.048 converts the free energy into units of kcal/mole. The $\Delta G^{\neq}_{elec}$ will primarily direct mutations that will benefit k_{cat} trends as estimated through transition state theory

$$k_{cat} \propto e^{-\beta \Delta G^{\ddagger}_{elec}} = e^{-\beta \Delta G^{\ddagger}_{base}} e^{-\beta \Delta G^{\ddagger}_{active}} e^{-\beta \Delta G^{\ddagger}_{other}} \quad (3)$$

where we have decomposed $\Delta G^{\neq}_{elec}$ into terms that emphasize the free energy stabilization of electric fields from the catalytic base, $\Delta G^{\neq}_{base}$, the active site, $\Delta G^{\neq}_{active}$, and the rest of the protein scaffold as well as solvent, $\Delta G^{\neq}_{other}$. The ligand binding energy terms calculated with Eq. (1) were used to analyze K_M trends.

Figure 3a show the electric field contribution to each bond from the individual residues of the KE15 designed sequence in the EL† state, and Table 2 shows the breakdown of electrostatic free energy changes between the EL and EL† contributions for the catalytic base, active site, remaining protein scaffold, and solvent for the KE15 design; we also include the reference state of the electric field stabilization for each bond of the substrate in aqueous solvent. It is clear that the catalytic base (Asp48) is the only residue with the largest productive electric field contribution to the making and breaking of bonds of the 5-nitrobenzisoxazole substrate, whereas the rest of the scaffold and enzyme solvent plays a small or even counterproductive role for transition state stabilization, consistent with what we found for KE07 and KE70[34].

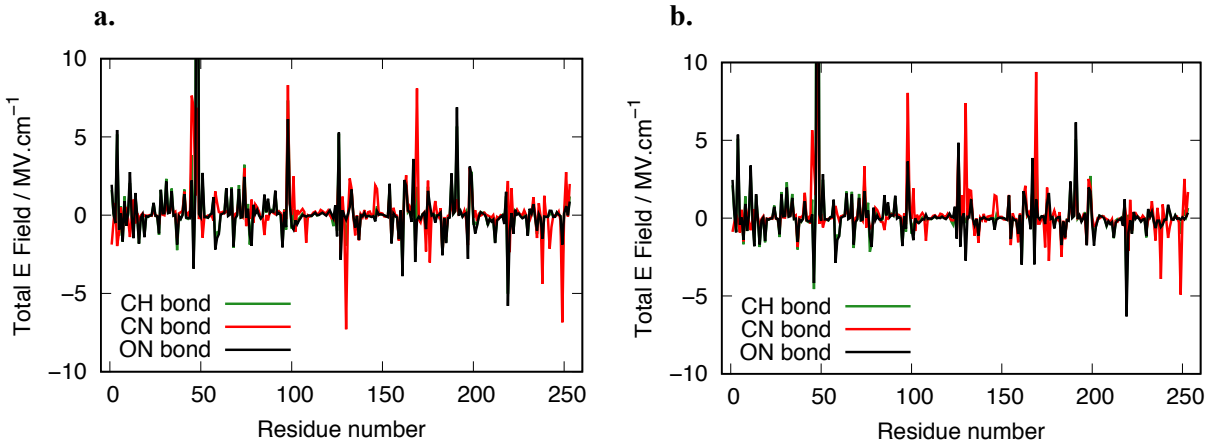


Figure 3. *Individual residue contribution to the electric field, $\bar{E}_{EL\dagger}$ of the (a) KE15 design and (b) KE15 Asp130Lys mutation at the transition state.* Electric projections are shown for the C-H (green), C≡N (red) and O-N (black) bonds of the ligand. The water contributes -3.8, 20.4 and -18.0 MV/cm for the KE15 design CH, CN and ON bonds respectively, while contributing -2.0, 1.0 and -14.4 MV/cm for the KE15 Asp130Lys mutant. KE15 contains a total of 253 residues.

Table 2: *Breakdown of $\Delta G^{\neq}_{elec}$ from the different regions of the protein and solvent for the designed and computationally evolved enzymes and the calculated binding energy, E_{bind}, of the ligand in the enzyme*. The free energy contribution (kcal/mole) from the electric field at each of the substrate bonds (Fig. 2) using Eq. (2), for the designed KE15 and computationally improved variants. The active site is defined by residues within 5 Å from the substrate (see SI for residue numbers), while the protein environment is summed over all residues outside this region. Solvent includes waters in the neck of the TIM barrel as well as the surrounding hydration and bulk water. The binding energy is calculated using Eq. (1). The reference energy is the electric field contribution on each bond from aqueous solution (no enzyme).

Bond	**$\Delta G^{\neq}_{elec}$ (kcal/mole)**				
C-H	**KE15 Design R1**	**Asp130Lys+R1 R2**	**Ile168Met+R2 R3**	**Gly199Ala+R3 R4**	**Tyr167Lys+R4 R5**
Base	-4.6	-4.1	-5.8	-6.7	-8.6
Active	0.1	0.1	-0.1	0.5	-0.6
Protein	-1.0	-0.9	-0.9	-0.5	0.0
Solvent	0.1	-0.3	0.3	0.7	1.4
Total	-5.4	-5.2	-6.5	-6.0	-7.8
Reference	**-3.6**				
C-N	**KE15 Design R1**	**Asp130Lys+R1 R2**	**Ile168Met+R2 R3**	**Gly199Ala+R3 R4**	**Tyr167Lys+R4 R5**
Base	2.0	0.4	0.9	0.9	2.1
Active	0.0	0.3	0.3	0.3	-0.2
Protein	0.8	0.8	0.6	1.1	1.5
Solvent	0.2	0.5	0.6	0.0	-0.5
Total	3.0	2.1	2.4	2.3	2.9
Reference	**2.3**				
O-N	**KE15 Design R1**	**Asp130Lys+R1 R2**	**Ile168Met+R2 R3**	**Gly199Ala+R3 R4**	**Tyr167Lys+R4 R5**
Base	-6.0	-5.2	-6.3	-7.5	-7.9
Active	1.1	0.7	0.8	1.9	0.6
Protein	-1.6	-1.1	-1.3	-0.2	0.7
Solvent	3.2	2.5	3.3	3.2	3.1
Total	-3.3	-3.1	-3.5	-2.6	-3.5
Reference	**-0.9**				
Net Base	-8.6	-8.9	-11.2	-13.3	-14.4
Net Active	1.2	1.1	1.0	2.7	-0.2
Net Protein	-1.8	-1.2	-1.6	0.4	2.2
Net Solvent	3.5	2.7	4.2	3.9	4.0
Net $\Delta G^{\neq}_{elec}$	**-5.7**	**-6.3**	**-7.6**	**-6.3**	**-8.4**
Net reference	**-2.2**				
E_{bind} (kcal/mole)					
unconstrained	-26.9 ± 2.2	-26.1 ± 2.0	-27.8 ± 1.9	-23.0 ± 1.9	-17.2 ± 1.7
constrained		-30.5 ± 1.4	-34.7 ± 1.6		

However, the magnitude of ~35-50 MV/cm, depending on chemical bond of the substrate as seen in Table 3, from Asp48 is under half the electric field value than we found for the corresponding catalytic base Glu101 for the KE07 enzyme, and is still small relative to the values contributed by the His-Asp dyad for KE70, making $\Delta G^{\neq}_{base}$ for KE15 much smaller than the KE07 and KE70 designs[34]. Furthermore, all other KE15 residues and enzyme solvent show only minimal electric field contributions to the 3 substrate bonds, with only a few residues exhibiting contributions of around ~ ±7 MV/cm at most. This is unlike what we previously observed for the KE07 and KE70 designs where many residues and solvent electric field contributions were much larger (anywhere between ±10 MV/cm and ±50 MV/cm), and thereby contributed much larger $\Delta G^{\neq}_{other}$ contributions to stabilization or destabilization.[34]

Table 3: *Electric field contributions by region for KE15 designed and computationally evolved enzymes*. The magnitude of the electric field at the C-H, C-N, and O-N bond in either the EL and EL† states for the designed KE15 and computationally improved variants. The active site is defined by residues within 5 Å from the substrate (see main text for residue numbers), while the protein environment is summed over all residues outside this region. Solvent includes waters in the neck of the TIM barrel as well as the surrounding hydration and bulk water. Positive sign indicates field supporting the reaction (CH and ON bond breaking as well as CN triple bond formation)."Fields are reported in units of MV/cm.

C-H bond	**KE15 Design R1**		**Asp130Lys+R1 R2**		**Ile168Met+R2 R3**		**Gly199Ala+R3 R4**		**Tyr167Lys+R4 R5**	
	EL	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**
Base	46.2	49.5	24.8	60.2	36.0	84.0	35.7	104.1	68.4	111.5
Active	-5.9	4.4	-1.7	3.4	-0.0	2.5	-5.0	-5.0	4.6	7.2
Protein	6.2	15.4	5.2	13.2	6.0	12.6	2.0	9.2	-0.5	1.1
Solvent	-2.4	-3.8	7.5	-2.0	3.5	-8.9	3.3	-18.5	-14.6	-15.3
C-N bond	**KE15 Design R1**		**Asp130Lys+R1 R2**		**Ile168Met+R2 R3**		**Gly199Ala+R3 R4**		**Tyr167Lys+R4 R5**	
	EL	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**
Base	22.1	6.9	7.8	19.2	14.0	24.4	17.2	38.1	31.4	48.1
Active	4.2	22.9	7.4	20.5	8.0	22.3	6.9	16.8	1.0	12.8
Protein	10.1	9.9	13.7	23.9	10.7	22.9	16.4	25.9	21.0	26.3
Solvent	6.3	20.4	5.7	1.0	6.6	3.4	-1.8	-7.7	-6.7	-8.5
O-N bond	**KE15 Design R1**		**Asp130Lys+R1 R2**		**Ile168Met+R2 R3**		**Gly199Ala+R3 R4**		**Tyr167Lys+R4 R5**	
	EL	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**	**EL**	**EL†**
Base	26.0	35.3	15.2	36.1	14.6	46.3	21.7	52.2	31.6	48.4
Active	-12.2	-1.0	-5.7	-1.7	-10.5	0.5	-14.1	-6.6	-4.1	-2.8
Protein	4.4	11.0	4.4	6.5	7.1	6.3	1.2	0.7	-3.8	-3.4
Solvent	-15.3	-18.0	-11.1	-14.4	-12.2	-20.8	-9.9	-21.9	-18.2	-14.9

Although k_{cat} only differs by a factor of 3 for the KE07 and KE15 designs, their active sites are in completely different parts of the TIM barrel (Figure 4a), which likely explains the differences in the electric field magnitudes. Unlike the creation of HG1[9], which sought to move the active site entirely to elsewhere in the protein, we will attempt to maintain the KE15 active site at its designed position and instead seek mutations that further improve electrostatic fields. In addition, we consider the relative importance of the electric fields emanating from the base and active site residues, and compare their free energy contribution to that from the protein scaffold and surrounding solvent as embodied in $\Delta G^{\neq}_{other}$.

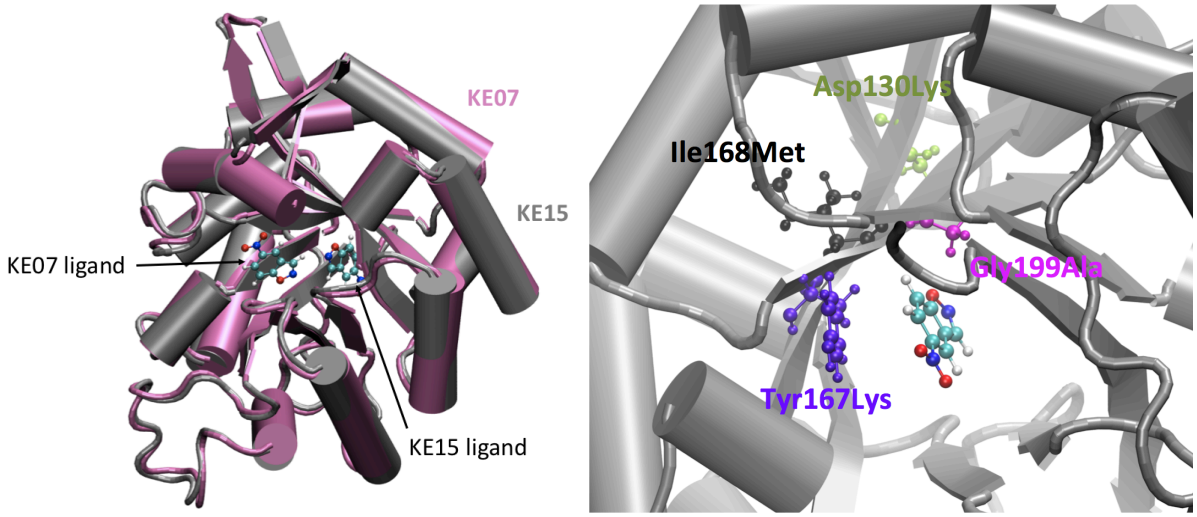


Figure 4. (a) *Substrate and ligand position in the TIM barrel for KE07 (pink) and KE15 (grey) designs. (b) Location of the 4 mutations of KE15 best variant.* Asp130Lys lies in the lower barrel of the scaffold; Ile168Met is closer to the substrate but faces outward and away from the ligand, unlike Tyr167Lys that faces towards it. Gly199Ala is in more direct contact with the bonds that break and form during the reaction and the mutation to a bulkier residue brings the ligand closer to the base."

The largest negative electric field contribution of any residue in the KE15 design is at position Asp130 located at the bottom of the TIM barrel relatively far from the substrate, with E= -7.29 MV/cm when projected onto the C≡N bond. Thus, the most unfavorable $\Delta G^{\neq}_{elec}$ contribution arises from $\Delta G^{\neq}_{other}$. Therefore Asp-130 is expected to have a long-range effect on the chemical step (Figure 4b) and hence a change of charge rather than hydrophobicity and/or size, seemed more likely to increase the electrostatic stabilization of the transition state. Since aspartic acid is negatively charged, we tried a neutral (asparagine) and positively charged (lysine) residue in its

place. Table 2 reports that the mutation Asp130Lys was predicted to significantly increase the electric field stabilization of the transition state but with some small degradation in K_M as estimated from E_{bind}. This qualitative trend was in fact verified experimentally to yield our first best variant with an experimentally measured k_{cat}/K_M = 59 $M^{-1}s^{-1}$, with the bulk of the improvement stemming from a factor of 4-5 in k_{cat} as shown in Table 1. Our calculations further indicate that Asp130Lys stabilized the transition state via improvement in $\Delta G^{\neq}_{elec}$ at the C≡N bond, as intended, but mostly through better positioning of the base as indicated by $\Delta G^{\neq}_{base}$ (Tables 2 and 3).

The electric field projections by individual residues onto the substrate bonds of the transition state for the Asp130Lys mutant enzyme (Figure 3b) does so with even smaller magnitudes than the KE15 design, so that it is not possible to rationally select residues for further mutation based on electric field contributions alone. Instead we restrict ourselves mostly to the residues located near the active site where electrostatic effects will be largest and the search space most restricted. This is also a sensible strategy since we have shown in previous work that the electric fields emanating from the remainder of the TIM barrel scaffold and surrounding solvent are *incommensurate* with the new engineered active site chemistry, such that the electrostatic fields of the remainder of the enzyme and solvent actually *disfavor* the catalytic reaction.[16, 34] In addition, we concluded that chemical positioning of the ligand relative to the catalytic base is a requirement for optimal electric field alignment, and contributed most to changes in catalytic activity, which is largely dictated locally at the active site.[34] Active site residues, defined as within 5 Å of the substrate includes positions 5, 46, 48, 78, 101, 126, 167, 168, 169, 197, 198, 199, 201, 220. We analyzed these positions for both detrimental (albeit small) electrostatic contributions and/or ligand positioning relative to the catalytic base.

One active site residue, Ile-168, contributes negatively to the electric field at the transition state, albeit with small values of E=-1.67 and -2.97 MV/cm, for the C-H and O-N bonds, respectively. Ile-168 is closer to the substrate than Asp-130 but faces away from the ligand as shown in Figure 4b, indicating an indirect effect on catalysis. Steric hindrance prevented the replacement of isoleucine with bulkier residues than methionine, and the Ile168Met mutation was predicted to further stabilize the transition state by ~1.3 kcal/mol and a corresponding improvement of ~1.7 kcal/mol in the calculated E_{bind} (Table 2). Experimentally, Ile168Met showed improved performance, but mostly through K_M, suggesting a tuning of

substrate binding instead. Although our binding calculations do indeed capture a larger effect for changes in K_M for this mutant, the projected electric fields that are primarily designed to capture changes in k_{cat} had predicted more improvements in catalysis than was actually found from experiment.

To better understand this apparent discrepancy, we analyzed the MD trajectories from which we compute the electric fields to determine the distance between the ligand and the catalytic base. We found that, on average, the mutation Ile168Met resulted in the decrease of the distance between the oxygen of the base and the hydrogen of the ligand by almost 1.0 Å through reduction in size of the active site. However, a shorter base-ligand separation might result in both a more favorable binding as well as enhancement of electric fields. To better separate electric field enhancement vs. better substrate binding, we ran a test simulation in which the ligand was restrained to remain in close proximity to the base (see Methods for details). We observe that when the ligand is restrained, there is no apparent electrostatic stabilization going from round 2 (the Asp130Lys enzyme) to round 3 (which builds on round 2 with the addition of Ile168Met) (Table S2 in Supporting Information). However, the binding affinity does indeed increase (Table 2), reasonably reconciling theory with experiments.

It is interesting to note that the magnitude of the electric fields roughly doubles for all three bonds in the simulations when the ligand is restrained close to the base (Table S2). This motivated the design of a mutation that would actually bring the ligand closer to the base, as in the restrained simulations. Active site residue Gly199 is advantageously positioned for that purpose as illustrated in Figure 4b, and we found that mutation of Gly199Ala showed further reduction in the distance between the ligand and the base (an additional 0.5 Å compared to round 3). This is borne out in the large increase in the electric field coming from the base for the C-H and N-O bonds (Table 3) and a corresponding large increase in stabilization of $\Delta G^{\neq}_{base}$~2.0 kcal/mole as given in Table 2.

However, this increase is accompanied by a decrease in the contribution of the rest of the active site and scaffold and solvent where $\Delta G^{\neq}_{active} + \Delta G^{\neq}_{other}$ is destabilizing by ~3.4 kcal/mole such that the overall electrostatic free energy of stabilization is in fact calculated to be lower than for round 3 (Table 2). Experimentally, round 4 has improved performance with a significant increase in k_{cat} and only a slight increase in K_M (in qualitative but not quantitative agreement with E_{bind}). This means that the predicted enhancement of the contribution of the catalytic base

actually overcomes the other detrimental changes, indicating that $\Delta G^{\neq}_{base}$ is more predictive than is $\Delta G^{\neq}_{elec}$ for improving k_{cat}.

For round 5, we set to correct for the disadvantages of Gly199Ala, i.e. reduction in the electric field contribution of the active site, while keeping its benefits from $\Delta G^{\neq}_{base}$, through optimization of $\Delta G^{\neq}_{active}$. Therefore, we target residues that are contributing negatively to the electric field when the ligand is constrained to be close to the base in round 3 to identify a new active site residue, Tyr-167 that is characterized by a negative contribution to the electric field (E= -3.04 MV/cm) for the N-O bond. In Figure 4b, we see that Tyr-167 is located at the other end of the ligand, near the NO_2 group, which is ideal to minimize the impact of this mutation on the position of the Asp48 base that is closest to the C-H ligand bond. Set to improve electrostatic interactions for the active site, we changed the charge of the residue and found that Tyr167Lys is predicted to stabilize the transition state by ~2 kcal/mol compared to round 4 and around ~1 kcal/mol compared to round 3 (Table 2). Furthermore, we see in Table 2 that both $\Delta G^{\neq}_{base}$ and $\Delta G^{\neq}_{active}$ improve significantly, with the latter being much better optimized and contributing a majority to $\Delta G^{\neq}_{elec}$. Experiments are in line with these predictions, and round 5 shows the best efficiency yet, with a 43-fold improvement in k_{cat} compared to the KE15 design and an overall efficiency of 403 $M^{-1}s^{-1}$.

DISCUSSION

With just 4 mutations we have improved KE15 to an efficiency that is comparable to the performance enhancements that was observed for KE07 and KE70 using LDE, and based on the iterative computational design approach used by Privett and co-workers to create HG3[9]. In the latter case, the largest improvement stemmed from repositioning the active site to a different part of the protein, to isolate the base from solvent in a more hydrophobic pocket to increase its pKa, followed by optimization of conformational dynamics in the active site, resulting in ~20 mutations.[9] Our computational approach is very different in the retention of the designed active site position, but evolving the electrostatic environment to create much larger electric fields that are better aligned to promote bond making and breaking of the substrate using much fewer mutations (Table 3). Although we observe modest reactant state destabilization, most of our improvements in catalytic rates come through transition state stabilization as summarized graphically in Figure 5 and in Table S3.

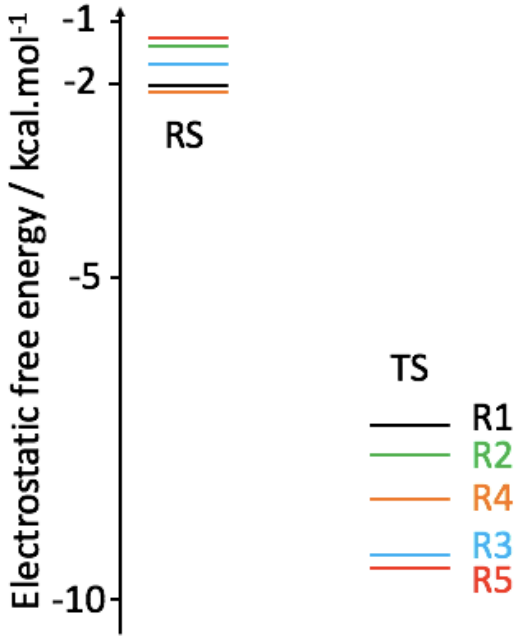


Figure 5: *Electrostatic free energy diagram of KE15 and all improved mutants predicted from electrostatic calculations.* While we observe moderate ground state destabilization going from R1 to R2 and R4 to R5, most of the free energy improvements reported in the main text come from transition state stabilization.

Furthermore, as discussed previously by Fuxreiter and Mones[13], there are other design principles derived from our understanding of the catalytic power of natural enzymes that can be applied to the designed enzyme KE15 and our improved enzymes using optimization of electrostatics and substrate optimization in the active site.. It would be a fruitful line of inquiry in future work to explore how changes in structural reorganization energy[54-55] and reorganization energy based on Marcus theory used by Labas and co-workers[17] could help understand how these quantities vary among our mutants, and how we might consider all design principles simultaneously. It is also interesting that when LDE was applied to HG3, the evolved enzyme was improved by three orders of magnitude.[10] Since our computational improvement in the KE15 enzyme starts from a very different active site and its optimization approach through electrostatics, it would be interesting to explore whether LDE can build upon our computational design of KE15-4 to reach the catalytic efficiency of HG-17. Finally, we note that the concepts and arguments presented in this study to improve KE15 could be used with a different theoretical model. In particular, although basic electrostatic effects are appropriately accounted for with a classical force field approach, a quantum description of at least part of the system would allow a more accurate continuous electric potential to be computed, with appropriate accounting of charge penetration and charge transfer effects not represented in the AMOEBA force field.

CONCLUSIONS

We have used information about electric fields and chemical positioning of the substrate in the active site to create a computationally designed Kemp Eliminase enzyme that has improved k_{cat} from 0.007 s^{-1} to 0.31 s^{-1} using only 4 mutations. We have drawn several conclusions in regards future computational design strategies for *de novo* enzymes in the future. First is that there is a strong interplay between electric field enhancements and chemical positioning, since the latter must be optimized in order for the electric field alignment to be maximized. Second, any active site position can be rescued through electric field enhancements of the key players in the active site, especially from the catalytic base. Finally we have found that computational improvements emanating from the arbitrary scaffold of the designed enzyme continue to be a limitation[36], in contrast to natural enzymes have a scaffold that is pre-organized to be more commensurate with the active site.[26] Our electric field and chemical positioning strategy has proven to be complementary to the iterative approach used by Privett et al.[9] for HG3, which focused on repositioning the active site and increasing its hydrophobicity, and to Labas and co-workers[17] who suggested that LDE could be replaced by computational optimization of reorganization energy instead. There are now many proven computational strategies for optimization of *de novo* enzymes that can be used in future enzyme design work.

SUPPORTING INFORMATION. Bond dipole moments and electric field values for specific regions at the 3 bonds.

ACKNOWLEDGMENTS. This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Hallie Honma for early calculations on KE15.

REFERENCES

1. Tantillo, D. J.; Jiangang, C.; Houk, K. N., *Curr. Opin. Chem. Bio.* **1998,** 2, 743-750.
2. Kiss, G.; Çelebi‐Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K., *Angew. Chem. Int. Ed.* **2013,** 52, 5700-5725.

3. Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. a.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., *Nature* **2008,** 453, 190-5.
4. Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S., *Proc. Natl. Acad. Sci. USA* **2012,** 109, 10358-63.
5. Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S., *J. Mol. Bio.* **2010,** 396, 1025-42.
6. Khersonsky, O.; Röthlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K. N.; Baker, D.; Tawfik, D. S., *J. Mol. Bio.* **2011,** 407, 391-412.
7. Jiang, L.; Althoff, E. a.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., *Science* **2008,** 319, 1387-91.
8. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L., *Science* **2010,** 329, 309-313.
9. Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L., *Proc. Natl. Acad. Sci. USA* **2012,** 109, 3790-3795.
10. Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D., *Nature* **2013,** 503, 418-421.
11. Brustad, E. M.; Arnold, F. H., *Curr. Opin. Chem. Bio.* **2011,** 15, 201-210.
12. Tobin, M. B.; Gustafsson, C.; Huisman, G. W., *Curr. Opin. Struct. Bio.* **2000,** 10, 421-427.
13. Fuxreiter, M.; Mones, L., *Curr. Opin. Chem. Bio.* **2014,** 21, 34-41.
14. Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A., *Proc. Natl. Acad. Sci. USA* **2010,** 107, 16869-74.
15. Frushicheva, M. P.; Cao, J.; Warshel, A., *Biochem.* **2011,** 50 (18), 3849-3858.
16. Jindal, G.; Ramachandran, B.; Bora, R. P.; Warshel, A., *ACS Catalysis* **2017,** 7 (5), 3301-3305.
17. Labas, A.; Szabo, E.; Mones, L.; Fuxreiter, M., *Biochim. Biophys. Acta* **2013,** 1834, 908-17.
18. Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Stoddard, B. L.; Popovic, Z.; Baker, D., *Nature biotechnology* **2012,** 30, 190-192.
19. Tokuriki, N.; Tawfik, D. S., *Curr. Opin. Struct. Bio.* **2009,** 19, 596-604.
20. Bershtein, S.; Tawfik, D. S., *Curr. Opin. Chem. Bio.* **2008,** 12, 151-158.
21. Romero, P. A.; Arnold, F. H., *Nature Revi. Mol. Cell Bio.* **2009,** 10 , 866-876.
22. Goldsmith, M.; Tawfik, D. S., *Curr. Opin. Struct. Bio.* **2012,** 22, 406-412.
23. Voigt, C. A.; Martinez, C.; Wang, Z.-G.; Mayo, S. L.; Arnold, F. H., *Nature Struct. & Mol. Bio.* **2002,** 9, 553-558.
24. Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z. G., *Proc. Natl. Acad. Sci. USA* **2001,** 98, 3778-3783.
25. Ruscio, J. Z.; Kohn, J. E.; Ball, K. A.; Head-Gordon, T., *J. Amer. Chem. Soc.* **2009,** 131, 14111-5.
26. Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M., *Chem. Rev.* **2006,** 106, 3210-3235.
27. Warshel, A., *Proc. Natl. Acad. Sci. USA* **1978,** 75, 5250-5254.
28. Warshel, A.; Weiss, R. M., *J. Amer. Chem. Soc.* **1980,** 102, 6218-6226.

29. Roca, M.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J., *J. Amer. Chem. Soc.* **2005,** 127, 10648-10655.
30. Swiderek, K.; Marti, S.; Tunon, I.; Moliner, V.; Bertran, J., *J. Amer. Chem. Soc.* **2015,** 137, 12024-34.
31. Krzeminska, A.; Moliner, V.; Swiderek, K., *J. Amer. Chem. Soc.* **2016,** 138, 16283-16298.
32. Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J., *ACS Catalysis* **2015,** 5, 2587-2595.
33. Swiderek, K.; Tunon, I.; Moliner, V.; Bertran, J., *Chem. Phys. Chem.* **2017,** 23, 7582-7589.
34. Bhowmick, A.; Sharma, S. C.; Head-Gordon, T., *J. Amer. Chem. Soc.* **2017,** 139, 5793-5800.
35. Bhowmick, A. Understanding and Improving Designed Enzymes by Computer Simulations. Thesis, University of California, Berkeley, 2016.
36. Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T., *Phys. Chem. Chem. Phys.* **2016,** 18, 19386-19396.
37. Šali, A.; Blundell, T., *J. Mol. Bio.* **1993,** 234, 779-815.
38. Friedland, G. D.; Kortemme, T., *Curr. Opin. Struct. Bio.* **2010,** 20, 377-84.
39. Friedland, G. D.; Lakomek, N.-A.; Griesinger, C.; Meiler, J.; Kortemme, T., *PLoS Comp. Bio.* **2009,** 5, e1000393.
40. Friedland, G. D.; Linares, A. J.; Smith, C. a.; Kortemme, T., *J. Mol. Bio.* **2008,** 380, 757-74.
41. Bhowmick, A.; Head-Gordon, T., *Structure* **2015,** 23, 44-55.
42. Shapovalov, M. V.; Dunbrack, R. L., *Structure* **2011,** 19, 844-58.
43. Dunbrack, R. L., *Curr. Opin. Struct. Bio.* **2002,** 12, 431-440.
44. Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S., *J. Chem. Theory Comput.* **2013,** 9, 461-469.
45. Ren, P.; Wu, C.; Ponder, J. W., *J. Chem. Theory Comput.* **2011,** 7, 3143-3161.
46. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., *J. Chem. Theory Comput.* **2013,** 9, 4046-4063.
47. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., *SoftwareX* **2015,** 1–2, 19-25.
48. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., *Bioinform.* **2013,** 29 (7), 845-854.
49. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; Head-Gordon, M.; Clark, G. N.; Johnson, M. E.; Head-Gordon, T., *J. Phys. Chem. B* **2010,** 114 (8), 2549-64.
50. Ponder, J. W. *Tinker--Software Tools for Molecular Design*, 7.0; 2014.
51. Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O. N.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J.; Zeng, Q.; Case, D. A.; Eastman, P.; Wang, L.-P.; Essex, J. W.; Head-Gordon, M.; Pande, V. S.; Ponder, J. W.; Shao, Y.; Skylaris, C.-K.; Todorov, I. T.; Tuckerman, M. E.; Head-Gordon, T., *J. Phys. Chem. B* **2016,** 120, 9811-9832.
52. Lindemann, H.; Thiele, H., *Justus Liebigs Ann. Chem.* **1926,** 449, 63-81.
53. Hollfelder, F.; Kirby, A. J.; Tawfik, D. S.; Kikuchi, K.; Hilvert, D., *J. Amer. Chem. Soc.* **2000,** 122, 1022-1029.

54. Smith, A. J. T.; Müller, R.; Toscano, M. D.; Kast, P.; Hellinga, H. W.; Hilvert, D.; Houk, K. N., *J. Amer. Chem. Soc.* **2008,** 130, 15361-15373.
55. Hammes-Schiffer, S., *Biochem.* **2002,** 41, 13335-13343.