JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

The Importance of the Scaffold for *de Novo* Enzymes: A Case Study with Kemp Eliminase

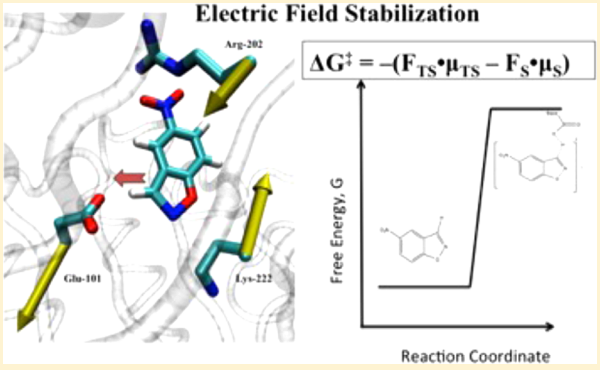
Asmit Bhowmick,[†] Sudhir C. Sharma,[‡] and Teresa Head-Gordon*[,†,‡,§,∥]

[†]Department of Chemical and Biomolecular Engineering, [‡]Department of Chemistry, and [§]Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720, United States

[∥]Chemical Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: We report electric field values relevant to the reactant and transition states of designed Kemp eliminases KE07 and KE70 and their improved variants from laboratory directed evolution (LDE), using atomistic simulations with the AMOEBA polarizable force field. We find that the catalytic base residue contributes the most to the electric field stabilization of the transition state of the LDE variants of the KE07 and KE70 enzymes, whereas the electric fields of the remainder of the enzyme and solvent *disfavor* the catalytic reaction in both cases. By contrast, we show that the electrostatic environment plays a large and stabilizing role for the naturally occurring enzyme ketosteroid isomerase (KSI). These results suggest that LDE is ultimately a limited strategy for improving *de novo* enzymes since it is largely restricted to optimization of chemical positioning in the active site, thus yielding a ∼3 order magnitude improvement over the uncatalyzed reaction, which we suggest may be an absolute upper bound estimate based on LDE applied to comparable *de novo* Kemp eliminases and other enzymes like KSI. Instead *de novo* enzymatic reactions could more productively benefit from optimization of the electrostatics of the protein scaffold in early stages of the computational design, utilizing electric field optimization as guidance.



INTRODUCTION

Although the design of new biocatalysts has not yet reached the level of proficiency of naturally occurring enzymes, there is optimism that further progress toward that goal is realistic and within reach as our understanding deepens on why current efforts have fallen short[1] and what makes natural enzymes so exceptional.[2,3] In this work we consider *de novo* enzyme design whereby a small catalytic "theozyme" is placed into an accommodating native protein scaffold, i.e., one that remains stable.[4,5] While minimal activity is observed for these *de novo* designed enzymes, it is still orders of magnitude below the activity typically seen in natural enzymes. While computation has provided insight[6−9] and useful improvements,[10−12] the majority of the improvement comes from laboratory directed evolution (LDE),[13] which alters the protein sequence through multiple rounds of mutagenesis and selection to isolate the few new sequences that exhibit enhanced catalytic performance.[14−18]

This process is well-illustrated by the *de novo* design of the Kemp elimination (KE) reaction,[5] involving the deprotonation of the ligand substrate 5-nitrobenzisoxazole by a catalytic base (Figure 1),[5] with corresponding electronic rearrangements that break the C−H and N−O bonds while forming a C≡N triple bond. Most designs have been engineered into related TIM barrel scaffolds, although sequence identity among them is low,

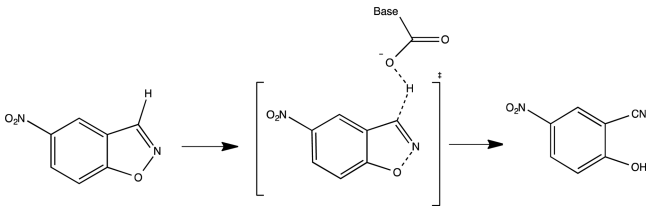


Figure 1. The KE reaction. The one-step reaction scheme involving the abstraction of hydrogen from the carbon of 5-nitrobenzisoxazole by a catalytic base. Shown is the transition state that has a partial negative charge on the substrate oxygen with cleavage of the O−N bond and nascent formation of a C≡N triple bond.

and then are optimized with LDE to create different catalytic motifs such as KE07,[14] KE70,[15] KE59,[18] and HG3.17.[17] For KE07 and KE70, the focus of our study here, the majority of catalytic performance was obtained after 6−7 rounds of LDE, which improved the k_{cat}/K_M by a factor of ∼200 (KE07.R7) and ∼400 (KE70.R6), respectively, in the best evolved enzymes (see Table S1 and Figure S1).[14,15] It is noteworthy that while most of the catalytic improvement for KE07 resulted from increases in k_{cat}, the improvements in KE70 were derived

Received: December 6, 2016
Published: April 6, 2017

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

equally from k_{cat} and K_M and would suggest that LDE took very different strategies in the optimization of the two enzymes.[14,15]

Almost all design protocols for Kemp eliminases[19−22] have taken a minimalist strategy of placing a base in a hydrophobic pocket, thus increasing the pK_a. For example, catalytic antibody 34E4 can catalyze the KE reaction with efficiencies comparable to the KE07.R7 variant using a simplified active site motif of a functional base surrounded by hydrophobic residues.[23] Similar rudimentary Kemp eliminases have also been designed not only into TIM barrels,[5,10,14,15,17,18] but into scaffolds of calmodulin[24] and T4-lysozyme.[25] Thus, regardless of the fold involved, a basic level of activity can be obtained for this reaction.[1] However, to reach the level of natural enzymes, there needs to be synergism between multiple functional groups[23] that includes not only hydrophobicity but also beneficial electrostatic contributions.[2,3] Furthermore, electrostatic stabilization comes not only from the proximity of a few residues in the active site[26,27] but also the rest of the protein scaffold[28−30] as well as the surrounding solvent.[31,32]

Advances in vibrational Stark effect (VSE) spectroscopy[30] have enabled researchers to probe the electric field in the active site of enzymes in order to quantify their contribution to the observed acceleration of reaction rates over the uncatalyzed reaction in aqueous solvent. An electric field can have a catalytic effect if it adopts a sustained direction that specifically stabilizes the transition state in preference to the reactant state—an effect that in principle is better optimized in the preorganized state of an enzyme relative to bulk aqueous environment.[2,3,29,33] Using VSE for the ketosteroid isomerase (KSI) enzyme and its inhibitor 19-nortestosterone (19NT), which has a C═O group located in the same position as the carbonyl group of it is natural substrate in the active site, Fried and co-workers have shown that the large electric fields exerted on this bond were linearly correlated with the activation free energies of the wild type and mutated variants.[29,34] Although precise chemical positioning of the Asp40 base in the active site for proton abstraction from the substrate is important for KSI,[27] leading to transition-state stabilization that contributes 2−3 orders of magnitude to the observed accelerated rate, the analysis of the VSE data suggests that ∼5 orders of magnitude improvement in k_{cat} comes about due to the reduction in the catalytic barrier that arises from the electrostatic *environment* of the KSI protein.[29,35] Although there is disagreement on the relative orders of magnitude that electrostatics contributes to the chemical base positioning vs the scaffold and solvent contributions,[27,36] there is no question that each are highly important for the catalytic performance of naturally occurring enzymes.

Presumably natural enzymes like KSI have developed highly optimized structural folds, including surfaces that invoke additional favorable orientations of solvent dipoles, which together contribute to a long-ranged and organized electrostatic environment for biocatalysis.[2,3] However, for *de novo* designed enzymes, it is reasonable to assume that they suffer from both nonoptimal chemical positioning as well as a poorly concocted electrostatic environment, since the scaffold merely serves as a "backdrop" for containing the designed active site. In this work, we decipher the contribution of electrostatic preorganization to transition-state stabilization in the designed Kemp enzymes KE07 and KE70 and to demonstrate how the electrostatics are further tuned by LDE to improve the catalytic activity for both.

Using atomistic computer simulations with an advanced polarizable force field, we measure the electric field at the three bonds that are made or broken in the ligand bound enzyme (EL) and transition state (EL†) as shown in Figure 2. While

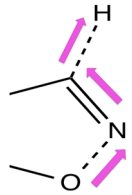


Figure 2. Electric field projection onto the C−H, C−N, and O−N bond dipoles of 5-nitrobenzisoxazole and sign convention used. Electric fields are calculated at the C, H, N, and O of the ligand, in which the critical chemical step of the reaction is the breaking of the C−H and O−N bonds and the making of the C−N triple bond. The positive field direction shown by arrows is chosen to conform to the opposite direction of movement of electrons in the KE reaction, a favorable field direction that supports the transition state.

Labas et al. considered all bonds of the substrate,[37] we focus on only these three bonds since their electronic bonding changes result in bond dipoles that show the greatest difference between the EL and EL† states (Table S2). While the 5-nitro group (Figure 1) might also play a role, under the AMOEBA electrostatic model we have found that the bond dipole difference between the EL and EL† states for the N−O bond of this group are negligible compared to the reactive bonds (Table S2), so that a sizable effect from this group can be ruled out.

Furthermore, transition-state free energy stabilization from bond dipoles and electric fields can be evaluated as

$$\Delta G^{\ddagger}_{elec} = -(\bar{\mu}_{EL^{\dagger}} \cdot \bar{E}_{EL^{\dagger}} - \bar{\mu}_{EL} \cdot \bar{E}_{EL}) \quad (1)$$

where $\bar{\mu}$ is the bond dipole and $\bar{E}$ is the electric field evaluated in the EL and EL† states. Finally, we have decomposed the electric fields into contributions from each residue of the enzyme[33] as well as solvent to better distinguish between "chemical positioning" of the catalytic base in the active site and the contributions that arise from the longer-ranged electrostatic environment from the protein and solvent.

We find that electrostatic fields are far greater in the active site of the enzymes relative to bulk solution when projected onto the relevant bonds, and a significant change in electrostatic preorganization was found when going from the designed enzyme to the most improved variant for KE07 but not KE70. We find that chemical positioning, i.e., the optimization of the active site base that interacts directly with the substrate, contributes the most to the electric field environments for KE07 and KE07.R7, whereas the stabilization effect of the electrostatic field is still present but smaller in the designed KE70 enzyme and does not improve in the LDE variant due to many mutations to hydrophobic amino acids that promote substrate affinity for the active site instead.

But in all cases, whether designed or LDE optimized, the electrostatic fields of the remainder of the enzyme and solvent disfavor the catalytic reaction. By contrast, we show that the electrostatic environment contributes a large stabilizing role for the naturally occurring enzyme KSI bound to the 19NT inhibitor. The underlying premise of the design approach—construction of a new catalytic "theozyme" that is placed into an *arbitrary* protein fold—suggests that one of the primary limitations of the *de novo* strategy is the restriction of the LDE search to optimization of chemical positioning in the active site, with an upper bound of ∼3 orders of magnitude estimated from

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

LDE applied to comparable *de novo* Kemp eliminases[17] and estimates made for natural enzymes like KSI.[26,27,29,34−36] Instead *de novo* enzymatic reactions would most productively benefit from optimization of the protein scaffold[38] in earlier stages of the computational design, utilizing electric field optimization as guidance, to recover the many missing orders magnitude improvements from electric field environments.

METHODS

Generating Backbone and Side Chain Ensembles for EL and EL† States of KE07 and KE70. For both enzymes, the initial design was modeled using the structures reported in refs 14 and 15 with the ligand docked in the appropriate position. Starting structures for improved variants for both cases (R7 for KE07 and R6 for KE70) were generated using Modeler. Using each of these PDB/modeled structures for the backbone in the ligand bound state, we then used the backrub algorithm implemented in Rosetta to run 25 independent simulations, each generating 10,000 trial moves using the C_α atoms as pivot residues, to generate uncorrelated backbone ensembles. From each simulation the lowest-energy structure was saved. Since the backbone scaffolds for KE07 and KE70 are quite rigid, we believe the backbone variations we have generated are adequate.

With these 25 backrub structures, we then used a recently developed Monte Carlo side chain ensemble (MC-SCE) method[39] to create large side chain ensembles for each structure. MC-SCE has been validated across a large number of proteins and protein complexes in which it performed extremely well in predicting observables reported in high-quality X-ray crystallography and NMR J-coupling experiments.[39] We note that for the MC-SCE calculation, the substrate was kept fixed in the docked position in both the EL and EL† state. The substrate geometry for the EL† state was the same as in the EL state with only the charges changed to reflect the transition-state nature. The resulting structural ensembles for KE07 and KE70 represent sampling on the microsecond to millisecond time scale as estimated from repacking of the amino acid side chains on different backbones.

Molecular Dynamics Simulations with AMOEBA. From the MC-SCE simulations on each backbone for the EL and EL† states of KE07 and KE70, we save the lowest-energy structure which is then used as the starting point for molecular dynamics simulations with the AMOEBA polarizable force field.[40−42] The AMOEBA model is described using a permanent multipole expansion up to quadrupoles, and polarization effects are explicitly accounted for by calculating induced dipoles in a self-consistent manner. Due to the sophistication of electrostatics and short-ranged anisotropic interactions, AMOEBA should provide an excellent model for the electric fields in enzymes.

All the molecular dynamics simulations in this study were performed using TINKER software. The tleap module in AMBER was used to solvate the system with a 10 Å spacing between the solute and the nearest box edge. Minimization was then performed using an LBFGS scheme with gradient RMS cutoff of 0.01. After minimization, an NPT simulation was performed with a time step of 1 fs integrated by the Beeman scheme. The temperature was maintained at 298 K with a Nose−Hoover thermostat. The PME real space cutoff and van der Waals cutoff was set to 8 Å. Induced dipoles were iterated until the root-mean-square change was $<10^{-5}$ Debye/atom. Given the ensemble of structures from the molecular dynamics and MC-SCE calculations described above, which provides effective sampling over much longer time scales than an individual and standard tens of nanosecond trajectory, we run 25 independent 100 ps trajectories of which we discard the first 50 ps and then collect statistics for the remaining 50 ps at intervals of 1 ps. Electric field values were calculated at the four atoms in the ligand involved in the breaking and making of chemical bonds in the substrate, namely C, H, N, and O as shown in Figure 2. This was done for EL and EL† states in both designed enzymes and best LDE variants.

Electric Field Calculations. In the AMOEBA framework, the permanent and induced electric fields at atom i due to another atom j can be written as follows:

$$E^{(i,j)}_{\mathrm{perm},\alpha} = -T_\alpha q^{(j)} + T_{\alpha\beta}\mu^{(j,\beta)} - \frac{1}{3}T_{\alpha\beta\gamma}\Theta^{(j,\beta,\gamma)} \tag{2a}$$

$$E^{(i,j)}_{\mathrm{ind},\alpha} = T_{\alpha\beta}\mu^{(j,\beta)}_{\mathrm{ind}} \tag{2b}$$

where $\alpha, \beta, \gamma = x, y, z$ and q, μ, Θ correspond to point charge, point dipole, and point quadrupole permanent electrostatics, μ_{ind} is the polarizable dipole, and the tensor T is expressed in a compact format as

$$T_{\alpha\beta\ldots\nu} = \frac{1}{4\pi\varepsilon_0}\nabla_\alpha\nabla_\beta\ldots\nabla_\nu\left(\frac{1}{R}\right) \tag{2c}$$

Although during the dynamical simulation the long-ranged electrostatics of the many-body polarization are evaluated under Ewald, in order to break down the electric field contributions from specific residues, we do an extra calculation where the induced dipoles are again calculated to convergence but using the real-space interactions only, with no cutoff's, and then eqs 2a−2c are calculated. When we add up all real-space contributions from all residues j to define the total electric field at the i = C, N, O, and H atoms of the substrate,

$$E^{(i)}_\alpha = \sum_{[j]} E^{(i,j)}_\alpha \tag{3}$$

we determine errors of ∼1.0% when we compare to the full Ewald calculation.

Once we know the electric field values at atomic site i due to site j, the electric field values at a bond are then evaluated as the arithmetic mean of the field values at the two atoms forming the bond. For example, along coordinate axis α, the average field at the bond b_{ik} comprised of atoms i and k due to residue j is

$$E^{(b_{ik},j)}_\alpha = (E^{(i,j)}_\alpha + E^{(k,j)}_\alpha)/2 \tag{4}$$

Field values along a bond are then calculated by taking the dot product between the electric field vector at the bond (eq 4) and the unit vector of the bond with positive direction illustrated in Figure 2. These values have been reported in all tables and figures. In all of the three bonds studied here, we chose the positive direction of the field to be opposite to the direction of movement of electron in the bond breaking or bond making process. This is shown in Figure 2 with the arrows illustrating the positive field direction for each bond.

KSI Simulations. In addition to simulations of designed enzymes KE07 and KE70, we also carried out electric field calculations for the natural enzyme KSI in complex with the inhibitor 19NT, the structure of which was recently published (PDB ID: 5KP4). The same protocol as used for KE07/KE70 was followed to calculate the fields. First backrub calculations were done with the inhibitor in the bound state to introduce backbone variability. Next, MC-SCE calculations were done on 25 of these backrub structures to model side chain variability. The parameters for 19NT inhibitor for MC-SCE calculations were generated using Antechamber. The lowest-energy structure from each of the MC-SCE ensembles was used as input for electrostatics calculations with AMOEBA. The same protocol as described above was followed for the field calculation, but now evaluated for the C═O bond of 19NT inhibitor. Parameters for 19NT for AMOEBA simulations were generated using the same protocol described in the SI material for the 5-nitrobenzisoxazole substrate in reactant state.

RESULTS

We first calculate the activation free energy stabilization of the transition state EL† relative to the reactant state EL due to electrostatics, $\Delta G^{\ddagger}_{\mathrm{elec}}$ using eq 1 in order to determine its contribution to the observed rate enhancements, i.e., on k_{cat}:

$$k_{\mathrm{cat}} = \frac{kT}{h}e^{-\beta\Delta G^{\ddagger}_{\mathrm{elec}}}e^{-\beta\Delta G^{\ddagger}_{\mathrm{other}}} \tag{5}$$

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

Table 1. Free Energy Stabilization of the Transition State[a]

enzyme construct and $\Delta G^{\ddagger}$ transition-state stabilization		fields generated for each bond		
		C−H	C≡N	O−N
designed KE07	EL	47.6 (3.9)	43.9 (2.0)	3.7 (2.7)
	EL†	68.7 (7.3)	58.8 (4.1)	22.7 (2.2)
$\Delta G^{\ddagger}$ = −5.3 kcal/mol		−5.6 kcal/mol	3.1 kcal/mol	−2.8 kcal/mol
LDE R7 variant KE07	EL	81.5 (11.0)	49.1 (4.3)	7.2 (3.7)
	EL†	108.2 (12.9)	77.8 (9.6)	30.3 (3.7)
$\Delta G^{\ddagger}$ = −9.9 kcal/mol		−9.1 kcal/mol	3.2 kcal/mol	−4.0 kcal/mol
designed KE70	EL	53.3 (3.6)	48.7 (2.4)	8.8 (1.7)
	EL†	77.6 (3.7)	62.2 (1.8)	28.1 (1.2)
$\Delta G^{\ddagger}$ = −6.6 kcal/mol		−6.3 kcal/mol	3.5 kcal/mol	−3.8 kcal/mol
LDE R6 variant KE70	EL	54.0 (3.8)	29.7 (1.7)	6.9 (1.1)
	EL†	76.7 (2.6)	37.0 (1.6)	16.8 (0.9)
$\Delta G^{\ddagger}$ = −6.6 kcal/mol		−6.3 kcal/mol	2.1 kcal/mol	−2.4 kcal/mol
substrate in water	EL	27.3	36.8	−10.5
	EL†	48.8	66.7	15.8
$\Delta G^{\ddagger}$ = −2.2 kcal/mol		−3.6 kcal/mol	2.3 kcal/mol	−0.9 kcal/mol

[a]Reduction in activated free energies are calculated using $\Delta G^{\ddagger} = -0.048(F_{TS}\cdot\mu_{TS} - F_{S}\cdot\mu_{S})$. Electric field values along the three bonds of the substrate 5-nitrobenzisoxazole in the EL and EL† states of KE07 and KE70 designed enzymes and best LDE variants as well as in aqueous solvent. Positive field indicates favorable contribution, and fields are reported in units of MV/cm. Standard error of the means are in parentheses. Bond dipole moments are estimated from AMOEBA charges and fixed dipoles (see SI material) in the EL and EL† complexes; for C−H μ_{TS} = 1.0D, μ_{S} = −1.0D; for C≡N μ_{TS} = 0.4D, μ_{S} = 2.0D; for O−N μ_{TS} = 2.3D, μ_{S} = −1.7D.

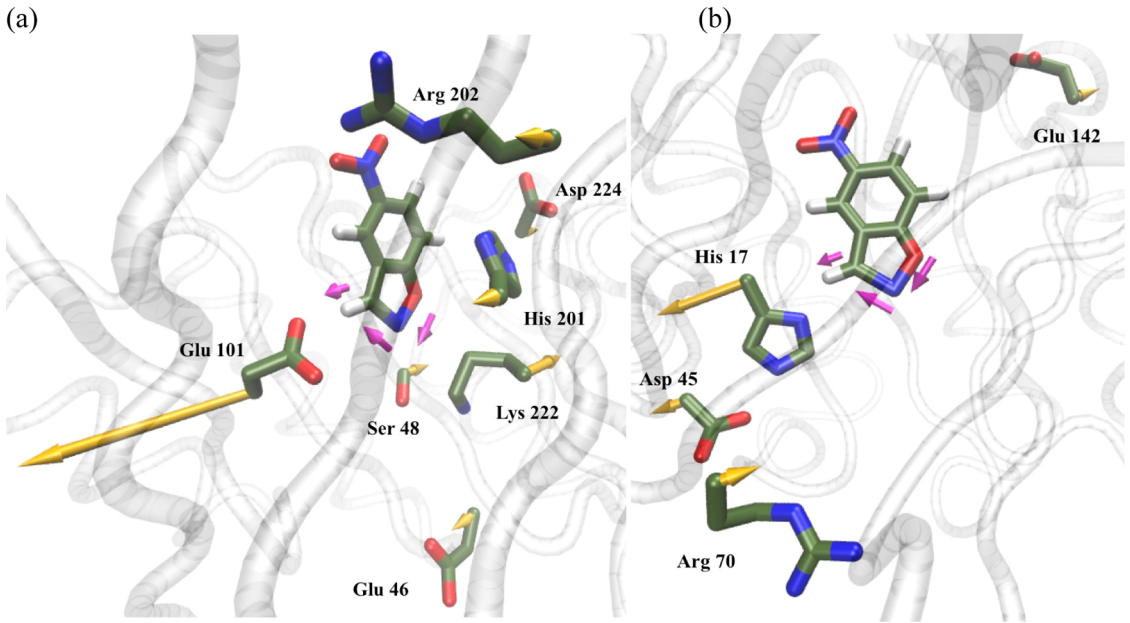


Figure 3. The electric field projection onto the C−H bond dipole of 5-nitrobenzisoxazole from key residues in the active of (a) KE07 and (b) KE70. The yellow arrows indicate the field direction/magnitude, and the ones in magenta indicate dipole directions for each bond studied. All residues shown have a field >10MV/cm ($\sim k_b T$) in the transition state of the best variant.

By convention, field directions that are aligned with the breakage of the C−H and N−O single bonds, and fields aligned in the opposite direction for the formation of a C≡N triple bond, would contribute to free energy stabilization of the transition state through electrostatics (eq 5). Using the transition-state structure reported in ref 43 for an acetate base for the same ligand, and using the AMOEBA electrostatic parameters for charges and fixed dipoles, we can assign a $G^{\ddagger}_{elec}$ contribution to the C−H, N−O, and C≡N bond dipoles in the EL and EL† states (see the SI material and Table S2 for details). The incipient bond in the transition state between the base oxygen/nitrogen and the abstracted hydrogen was not considered for this study.

It is important to note that we are missing other contributions to the total free energy barrier, $\Delta G^{\ddagger}_{other}$, such as the entropic effects arising from desolvation (although the enthalpic interactions are likely accounted for in part by the solvent electrostatic field contributions in $G^{\ddagger}_{elec}$). In addition, we have shown that side chain entropy played a significant role in the observed k_{cat} trends for which the active site of the original KE07 and KE70 enzymes were overdesigned for the binding affinity of the EL state, whereas the LDE optimized enzymes stabilized the EL† complex instead.[9] Therefore, while

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

experimentally the $\Delta\Delta G^{\ddagger}_{total}$= −2.6 kcal/mol accounts for the ∼70× improvement in k_{cat} for the best KE07 variant compared to the design, and the corresponding free energy barrier reduction $\Delta\Delta G^{\ddagger}_{total}$ = −2.1 kcal/mol accounts for the ∼35× improvement in k_{cat} for the best KE70 variant, we are only analyzing electric field contributions $\Delta G^{\ddagger}_{elec}$ and thus do not expect to reproduce these total activation free energy values.

Table 1 reports the total electrostatic field values along the three relevant bonds of the substrate 5-nitrobenzisoxazole in the EL and EL† states of the KE07 and KE70 designed enzymes, and their corresponding best LDE variants, as well as the fields acting on the reactant and transition state in aqueous solvent. If we assume that the electrostatic contribution to $\Delta G^{\neq}_{elec}$ arises from the additive contributions from the three bonds, then we can draw several immediate conclusions. The first is that the designed enzymes and their LDE variants help focus and enhance the electric fields along these bonds relative to the electric fields in bulk solvent, and overall the transition state is stabilized in preference to the reactant state regardless of enzyme variants (and which is true even in bulk solvent). In addition, the electric field stabilization is better for the designed KE70 relative to the designed KE07 enzyme, consistent with the fact that the k_{cat} of the former is an order of magnitude better than the latter. Finally, while for KE07 there is a very clear trend of increasing electric field strength going from the designed enzyme to the best R7 variant in both EL and EL† states for all relevant substrate chemical bonds, the KE70 enzyme exhibits no net activated free energy decrease in going from the designed to the best R6 LDE variant.

While it is apparent that there is transition-state stabilization, we further consider whether the activation free energy is attributable to preorganization, i.e., changes in the bond dipoles[29,30] without any corresponding relaxation of the structural ensemble in response, relative to reorganization that accounts for that relaxation cost. Table S3 shows that the adiabatic contribution, i.e., changes in the transition-state stabilization due to bond dipoles by averaging over the reactant ensemble only, is the dominant contribution to the free energy, with negligible contributions due to reorganization when averaging over the relaxed transition-state ensemble.

In the case of KE07, when we break down the contributions to the total electric field for the C−H, C≡N, and N−O bonds from individual residues (Table S4), we see that the overwhelming contribution comes from the catalytic base Glu-101 (Figure 3a), and the field strength contributed by the Glu-101 in the R7 variant increases by an additional ∼25−40 MV/cm over the designed KE07 enzyme, a huge improvement for transition-state stabilization. There are also significant stabilizing electric field contributions (>k_bT ∼ 10 MV/cm) arising from His201, and in the best LDE R7 variant from the Gly202Arg substitution, which we have shown in previous work interacts directly with the substrate to aid in chemical positioning of the base.[9]

However, in both the KE07 design and R7 variant, the designed residues Lys-222 and Ser-48, originally intended to stabilize the charge of the substrate in the transition state, have electrostatic fields that negatively impact the activation free energy. We and others have shown that Lys222 often forms a hydrogen bond with Ser48, as well as with residues Glu46 and Ile7 or its replacement in LDE R4 with Asp7, that help support the catalytic purpose of KE07 by removing unproductive interference with the base positioning.[7,9] While the Asn224Asp mutation is beneficial for the C≡N bond of the substrate, it is unproductive in regards electric field stabilization[44] for the C−H and N−O bonds, such that the net free energy is found to be detrimental to the catalytic step. Labas et. al also found that the Asn224Asp mutation yielded poor electric fields when they considered the ligand as a whole,[37] however our approach of analyzing the three main bonds of interest separately provides a potential rationale why the mutation might have been introduced in the first place since one important transition-state bond was stabilized. However, a possible alternate purpose for the Asn224Asp mutation is to better complex with water, as seen in the crystal structure of the R7 variant. Even so, these alternate roles for Lys222, Ser48, and Asp224 come with sacrifices to activation free energy stabilization afforded by constructive electric field effects on the substrate.

For KE70, there is virtually no overall optimization of the electrostatic fields going from the designed enzyme to the best R6 variant in both EL and EL† states (Table 1), a result that is largely orthogonal to the LDE optimization path taken for KE07. When we break down the largest contribution to the activation free energy by residue, there is ∼25 MV/cm enhancement from the His-Asp dyad for proton abstraction from carbon in the best R6 enzyme (Figure 3b), although the majority of the net ∼80 MV/cm field strength for the EL† state primarily comes from histidine (Table S5). This is not surprising since the main negative electric field contribution at this bond comes from Arg70 that is known to form unfavorable interactions with Asp45, thus reducing the pK_a of the His-Asp dyad. Otherwise, the electrostatic field due to the His-Asp catalytic base contributes negligibly to the stabilization of the other bonds of the substrate, suggesting that the electric field in KE70 is not as highly optimized as it is in KE07.

What modest gains that are made in electric field stabilization of the EL† state for the primary reactive step of C−H bond breaking in the designed KE70 enzyme are diminished by active site mutations to more hydrophobic groups (Trp72Cys and Ser138Ala) in the best LDE R6 variant. For KE70, it appears that other factors like productive binding of the substrate played a more significant role than electrostatics in the LDE improvement, captured experimentally through an order of magnitude reduction in K_M. In fact the pK_a(k_{cat}) ∼ 6.2 in both the designed KE70 and R6 LDE variant, whereas for KE07, where the majority of the improvement came through electrostatic stabilization, the pK_a(k_{cat}) changed from <4.5 in the design to 5.9 in the best R7 LDE variant.

A lack of stabilization of the oxy-anion is thought to be a bottleneck for the catalytic reaction executed in catalytic antibody 34E4 and appears to be a problem for both Kemp eliminases studied here, as can be seen from the electric fields projected onto the N−O bond dipole (Tables S4 and S5). Although LDE improved the electric fields for the C−H bond considerably in KE07, the improvements in the field for the N−O bond were considerably less, ∼1.2 kcal/mol of additional stabilization for the KE07.R7 variant. For KE70, this bond breaking was destabilized by LDE, quite possibly due to the complete removal of Ser-138 whose primary intent was stabilizing the oxy-anion. As already stated elsewhere,[23] oxy-anion stabilization may be as critical as the chemical positioning involving the proton abstraction step. Taking into consideration the fact that the N−O bond stabilization energy has the largest sensitivity to electric field (due to the large change in dipole moments), future efforts will benefit substantially from designing this feature in earlier stages and is evidently a

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

Table 2. Chemical Positioning vs Electric Field Environment in Designed and Natural Enzymes[a]

	KE07 design		KE07 R7 variant		KE70 design		KE70 R6 variant		KSI natural
region	EL	EL†	EL	EL†	EL	EL†	EL	EL†	EL
base	86.3	103.6	142.2	144.3	46.1	65.1	61.4	80.1	N/A
active	1.0	11.2	2.0	8.3	16.7	23.1	2.3	2.9	41.1
solvent	−15.6	−19.2	−22.6	−20.2	2.9	0.7	1.9	2.8	0.5
protein	−24.1	−26.8	−40.1	−24.1	−12.2	−11.3	−11.6	−9.1	16.0

[a]The magnitude of the electric field at the C−H bond in either the EL and EL† states for the designed KE07 and KE70 enzymes and the best LDE 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. Positive sign indicates field supporting bond breaking. For comparison with a natural enzyme, a similar breakdown for the C═O bond of inhibitor 19NT in KSI has been provided. Fields are reported in units of MV/cm.

primary reason why the HG3.17 enzyme improved under LDE.[17]

However, the important optimization of chemical positioning and active site improvements may also require further electrostatic stabilization of the transition state by the scaffold. For natural enzymes such as KSI, Fried et al. have shown that the major contribution to lowering the activation barrier comes from the electrostatic environment of the enzyme, as opposed to the contributions of residues that interact directly with the substrate or residues that aid in better chemical positioning of the catalytic base. For KSI it was estimated that a $10^{2.5}$-fold improvement in k_{cat} was due to chemical positioning, whereas an additional ∼10^5-fold improvement was attributable to the electrostatic "environment" of the protein scaffold and surrounding solvent. While the relative percent contributions due to chemical positioning vs electrostatic environment may be questioned,[27,36] there is no argument that enzyme folds have optimized an electrostatic environment that aids the catalytic reaction.

This is clearly not the case for the designed and LDE optimized KE07 or KE70 enzymes. Table 2 shows that the electric fields from the protein scaffold and solvent are mostly counterproductively aligned with the C−H bond for KE07 and KE70. To validate this observation, we performed the same calculation and region breakdown for the natural enzyme KSI complexed with the inhibitor 19NT (thereby only relevant for the reactant bound state). Table 2 shows that the electric field environment for the natural enzyme is *qualitatively* different, with the scaffold contributing ∼25% of stabilization of the substrate bound complex, in contrast with the destabilization observed for the reactant state for the designed enzymes. This result generalizes to the other bonds as well, with the exception that the C−N bond receives some support from the scaffold in the evolved KE07 enzyme, whereas the more important N−O bond is negatively impacted by the scaffold (Tables S6 and S7).

These observations on KE07 and KE70 go a long way to explain why *de novo* enzymes are so poor to begin with, and why LDE is such a limited strategy for improving them. By using an "arbitrary" protein scaffold as a container for the active site theozyme, and one that may also orient water solvents in such a way that are optimized for the scaffold and not the reactive chemistry, it should not be surprising the electric field environments are *highly* nonoptimized for stabilizing the reactant or transition state. Hence the only tractable LDE strategy is to optimize the electrostatic fields locally at the active site, as was done for KE07, or utilize other chemical positioning strategies or ways to increase the basicity of the catalytic base through creation of a more nonpolar active site, as found for KE70.

DISCUSSION

At present, computational approaches have yielded *de novo* enzyme designs that are minimally competent, and therefore there is a necessary reliance on LDE to bridge the performance gap to compete at the level of catalytic antibodies, but even then they are nowhere near the catalytic efficiencies of natural enzymes.[1] An important aspect that helps explain the incredible performance of natural enzymes is that they have optimized folded structures that create favorable electric fields from the entire protein and perhaps the surrounding solvent, not just the active site, to stabilize the transition state. In order to understand a natural enzyme's high catalytic proficiency, Warshel has suggested that an enzyme structural fold creates a preorganized electrostatic environment, not found in bulk aqueous solution, that preferentially stabilizes the transition-state charge distribution compared to the substrate reactant.

We conjecture that LDE is ultimately a limited strategy for improving *de novo* enzymes since it would require wholesale reengineering of most of the sequence of the scaffold; if such sequences prove to be unstable for maintaining the fold, it would extend the need to the creation of a new protein fold, that is beyond the capacity of any realistically sized LDE libraries, not to mention human time and patience. Thus, there are more limited options due to the vast reduction in the optimizable sequence space that is now largely restricted to chemical positioning in the active site. While it is true that nothing in principle limits LDE from exploring all mutations, high-throughput assays will still trap active site modifications more often than not, since scaffold optimization would be a small cumulative effect from round to round.

If we were to take KSI as a reference point for the free energy stabilization attributable to local active site contributions,[29] we would expect at most a 3 order of magnitude improvement using LDE. To support that estimate, at present, all known attempts to further optimize the artificial Kemp eliminase biocatalysts using LDE have yielded as little as 1 order of magnitude (the result for KE07 and KE70 after 6−7 LDE rounds beyond which no improvement was realized), to the best result obtained after 17 LDE rounds applied to the *in silico* design Kemp eliminase HG3,[10] yielding a k_{cat} for HG3.17 that is 700 s^{-1} relative to 0.68 s^{-1} for the design.[17] We do not mean to diminish what is clearly a success story, but we hypothesize that it may be unlikely for any designed enzyme to further improve once active site precision using LDE has been optimized, and one must now venture further into the greater protein scaffold to find the next orders of magnitude for improvement.

While *de novo* enzymatic reactions would most productively benefit from optimization of the protein scaffold utilizing

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

electric field optimization as guidance, it should happen in earlier stages of the computational design. For KE07 the cluster of interactions involving Lys222, Ser48, Ile7 (Asp7), and Asp224-water have allowed for better positioning of the Glu101 base to act on the substrate, but with counterproductive electric field effects on the substrate that raises the activation free energy. The primary problem in their removal is that these residues are "baked in" to perform other benefits to support the catalytic purpose of KE07, but the evolved enzyme has to develop even more optimized catalytic base electric fields to compensate. Both KE07 and KE70 may have reached a cul de sac in regards further improvement in the active site after 6−7 rounds of LDE due to such electric field compensations.

CONCLUSIONS

In this study we have used a robust model for electrostatics, the AMOEBA polarizable force field,[40−42] to calculate electric fields for the designed KE07 and KE70 enzymes and for the best variants that were improved under LDE. By calculating the field directions that are productively aligned with the breakage of the C−H and N−O single bonds and the formation of a C≡N triple bond for the small ligand substrate 5-nitrobenzisoxazole, we can assess the electrostatic free energy stabilization of the transition state relative to the reactant state. For KE07, it was found that the enhanced catalytic activity of the best R7 LDE variant stemmed from mutations that improved the electric fields locally in the active site, mostly attributed to the catalytic base, for stabilizing the transition state, while in KE70, the electric field enhancements to the transition state for its best LDE variant were more modest and completely isolated to the catalytic His17-Asp45 dyad. In both enzymes we find that the oxy-anion hole was not optimized by LDE and in fact that deficiency has been a known target for improvement in the design of Kemp eliminases. Finally, regardless of the Kemp eliminase construct (i.e., designed or LDE optimized), we showed that the electrostatic environments of the protein and solvent are counterproductive in their contribution to stabilizing the transition state.

We suggest that LDE is ultimately a limited strategy for improving *de novo* enzymes since we have argued that it may be largely restricted to optimization of chemical positioning in the active site, thus yielding up to a ∼3 order magnitude improvement that we suggest is an upper bound estimate based on the best known *de novo* Kemp eliminase HG3.17[17] as well as based on estimates made on naturally occurring enzymes such as KSI.[29] It might be suggested that the relatively modest LDE throughput of the assays used to optimize these particular Kemp eliminases may not be the best benchmarks for what is actually attainable under LDE. But even in the case of an ultrahigh-throughput assay in which Obexer et al. were able to optimize an aldolase that is >5 orders of magnitude better than the original computational design,[45] the starting activity itself was extremely poor (k_{cat} ∼ 0.00005 s^{-1}) resulting in a final k_{cat} that is ∼10 s^{-1}, similar to that obtained for the designed KE07 enzyme, i.e., the upper bound estimate for LDE improvement for *de novo* enzymes is an absolute one instead of being relative to the starting design. Our upper bound estimate on LDE improvement is drawn from the Kemp eliminases,[14,15,18] an aldolase,[45] and KSI[29] and would require additional tests of other designed enzymes to see if the "speed limit" is operative more generally.

Therefore, *de novo* enzymatic reactions could take a different tack by focusing on optimization of the protein scaffold in early stages of the computational design, utilizing electric field optimization as guidance. One simple optimization strategy would scan a range of known protein scaffolds with the theozyme present and rank them according to their electric field contributions. Widening the repertoire of folds considered for the design of Kemp eliminases beyond the TIM barrel is also likely to be beneficial. For example, in the design process of the Kemp eliminases, TIM barrels show up disproportionately (71% of low-energy structures) compared to their occurrence in natural enzymes (10%). Considering motifs of enzymes like KSI that catalyze a proton transfer reaction involving a labile hydrogen from an aromatic motif with high efficiency might be considered as an alternative scaffold. Even with the current TIM barrels used in the Kemp eliminases, one can imagine a better enzyme scaffold optimization by focusing on polar or charged residue mutations on the protein surface[46] to better preorganize solvent dipoles, whose integrated electric field could be quite large; every ∼30 MV/cm improvement in electric field alignment of the solvent on the active site would result in an order of magnitude of improvement in the catalytic rate.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b12265.

Parameterization, dipole moment and active site residue details. Tables for sequence of KE07/KE70, dipole moments, preorganization/reorganization energy break-up, field values for specific residues, scaffold, and solvent at the three bonds (PDF)

AUTHOR INFORMATION

Corresponding Author

*thg@berkeley.edu

ORCID

Teresa Head-Gordon: 0000-0003-0025-8987

Notes

The authors declare no competing financial interest.

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.

REFERENCES

(1) Korendovych, I. V.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **2014**, *27*, 113−121.

(2) Warshel, A. *J. Biol. Chem.* **1998**, *273* (42), 27035−27038.

(3) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106* (8), 3210−3235.

(4) 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.

(5) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. a.; Zanghellini, A.;

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800

Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190−5.
(6) Ruscio, J. Z.; Kohn, J. E.; Ball, K. A.; Head-Gordon, T. *J. Am. Chem. Soc.* **2009**, *131*, 14111−5.
(7) Alexandrova, A. N.; Röthlisberger, D.; Baker, D.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2008**, *130*, 15907−15.
(8) Boekelheide, N.; Salomón-Ferrer, R.; Miller, T. F., III *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16159.
(9) Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T. *Phys. Chem. Chem. Phys.* **2016**, *18* (28), 19386−19396.
(10) 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. U. S. A.* **2012**, *109* (10), 3790−3795.
(11) Bolon, D. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (25), 14274−9.
(12) Bolon, D. N.; Voigt, C. A.; Mayo, S. L. *Curr. Opin. Chem. Biol.* **2002**, *6* (2), 125−9.
(13) Arnold, F. H. *Acc. Chem. Res.* **1998**, *31*, 125−131.
(14) Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. *J. Mol. Biol.* **2010**, *396*, 1025−42.
(15) 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. Biol.* **2011**, *407*, 391−412.
(16) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. *Nat. Chem. Biol.* **2013**, *9* (8), 494−498.
(17) 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* (7476), 418−421.
(18) Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10358−63.
(19) Malisi, C.; Kohlbacher, O.; Höcker, B. *Proteins: Struct., Funct., Genet.* **2009**, *77* (1), 74−83.
(20) Nosrati, G. R.; Houk, K. *Protein Sci.* **2012**, *21* (5), 697−706.
(21) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15* (12), 2785−2794.
(22) Zhu, X.; Lai, L. *J. Comput. Chem.* **2009**, *30* (2), 256−267.
(23) Seebeck, F. P.; Hilvert, D. *J. Am. Chem. Soc.* **2005**, *127* (4), 1307−1312.
(24) Korendovych, I. V.; Kulp, D. W.; Wu, Y.; Cheng, H.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (17), 6823−6827.
(25) Merski, M.; Shoichet, B. K. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16179−83.
(26) Lamba, V.; Yabukarski, F.; Pinney, M.; Herschlag, D. *J. Am. Chem. Soc.* **2016**, *138* (31), 9902−9909.
(27) Natarajan, A.; Yabukarski, F.; Lamba, V.; Schwans, J. P.; Sunden, F.; Herschlag, D. *Science* **2015**, *349* (6251), 936.
(28) Huang, X.; Xue, J.; Lin, M.; Zhu, Y. *PLoS One* **2016**, *11* (5), e0156559.
(29) Fried, S. D.; Bagchi, S.; Boxer, S. G. *Science* **2014**, *346* (6216), 1510−1514.
(30) Fried, S. D.; Boxer, S. G. *Acc. Chem. Res.* **2015**, *48* (4), 998−1006.
(31) Dielmann-Gessner, J.; Grossman, M.; Nibali, V. C.; Born, B.; Solomonov, I.; Fields, G. B.; Havenith, M.; Sagi, I. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (50), 17857−17862.
(32) Grossman, M.; Born, B.; Heyden, M.; Tworowski, D.; Fields, G. B.; Sagi, I.; Havenith, M. *Nat. Struct. Mol. Biol.* **2011**, *18* (10), 1102−1108.
(33) Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J. *J. Am. Chem. Soc.* **2014**, *136* (29), 10349−10360.
(34) Wu, Y.; Boxer, S. G. *J. Am. Chem. Soc.* **2016**, *138* (36), 11890−11895.
(35) Fried, S. D.; Boxer, S. G. *Science* **2015**, *349* (6251), 936.
(36) Chen, D.; Savidge, T. *Science* **2015**, *349* (6251), 936.
(37) Labas, a.; Szabo, E.; Mones, L.; Fuxreiter, M. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 908−17.
(38) Altamirano, M. M.; Blackburn, J. M.; Aguayo, C.; Fersht, A. R. *Nature* **2000**, *403* (6770), 617−622.
(39) Bhowmick, A.; Head-Gordon, T. *Structure* **2015**, *23* (1), 44−55.
(40) Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J.; Zeng, Q.; Case, D. A.; Eastman, 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.
(41) 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.
(42) Ren, P.; Wu, C.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143−3161.
(43) Hu, Y.; Houk, K. N.; Kikuchi, K.; Hotta, K.; Hilvert, D. *J. Am. Chem. Soc.* **2004**, *126* (26), 8197−8205.
(44) Fuxreiter, M.; Mones, L. *Curr. Opin. Chem. Biol.* **2014**, *21*, 34−41.
(45) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. *Nat. Chem.* **2017**, *9* (1), 50−56.
(46) Russell, A. J.; Fersht, A. R. *Nature* **1987**, *328* (6130), 496−500.

DOI: 10.1021/jacs.6b12265
J. Am. Chem. Soc. 2017, 139, 5793−5800