JCTC
Journal of Chemical Theory and Computation

Article

Bottom-Up Nonempirical Approach To Reducing Search Space in Enzyme Design Guided by Catalytic Fields

Wiktor Beker* and W. Andrzej Sokalski*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 3420−3429

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Currently developed protocols of theozyme design still lead to biocatalysts with much lower catalytic activity than enzymes existing in nature, and, so far, the only avenue of improvement was the in vitro laboratory-directed evolution (LDE) experiments. In this paper, we propose a different strategy based on "reversed" methodology of mutation prediction. Instead of common "top-down" approach, requiring numerous assumptions and vast computational effort, we argue for a "bottom-up" approach that is based on the catalytic fields derived directly from transition state and reactant complex wave functions. This enables direct one-step determination of the general quantitative angular characteristics of optimal catalytic site and simultaneously encompasses both the transition-state stabilization (TSS) and ground-state destabilization (GSD) effects. We further extend the static catalytic field approach by introducing a library of atomic multipoles for amino acid side-chain rotamers, which, together with the catalytic field, allow one to determine the optimal side-chain orientations of charged amino acids constituting the elusive structure of a preorganized catalytic environment. Obtained qualitative agreement with experimental LDE data for Kemp eliminase KE07 mutants validates the proposed procedure, yielding, in addition, a detailed insight into possible dynamic and epistatic effects.
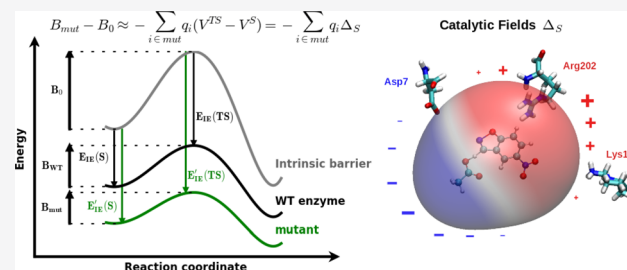
INTRODUCTION

In recent years, intense research has been conducted in pursuit of harnessing the versatile protein capabilites for application in industrial biocatalysis.

The major challenge arises from the fact that living organisms produce enzymes tailored for sustaining their own existence, rather than the needs of human society. Therefore, the key to the future of rational biocatalyst engineering lays in the redesigning of existing enzymes and, ultimately, de novo design with the aid of new computational methods.[1−4]

Major breakthroughs in the theoretical design of enzymes for new, non-natural reactions (so-called "theozymes") have occurred in the past two decades.[5−12] However, after initial success, the progress stalled,[13,14] because theozymes that were designed to maximize transition-state stabilization (TSS) displayed rather low catalytic activity and conventional top-down models requiring numerous assumptions were not able to fully explain the role of additional mutations in the second coordination sphere, introduced by LDE experiments.[13,14] Therefore, such LDE-mutated theozymes represent an excellent proving ground for testing new methodologies and interpreting the deficiencies of older methods. This problem was investigated by several research groups with various methods, including extensive molecular dynamics,[15,16] QM/MM,[17−19] and EVB[20−23] simulations. All previously mentioned methods belonging to the top-down category require consideriation of the entire protein and numerous assumptions regarding protonation states, QM/MM boundaries, always arbitrary selection of force field, etc. Although commonly used methods are able to explain some of the differences between designed and evolved enzymes, they are rather unsuitable for efficient scan over vast combinatorial space. Therefore, a method simple enough for high-throughput prescreening and yet deeply rooted in the physics of the problem without involving any empirical factors is needed to improve the existing protocols.

The catalytic field technique that has been used since we developed it in our laboratory in 1981[24−26] has been recently supplemented by Dittner and Hartke with extensive optimization techniques, allowing one to generate and explore various characteristics of the optimal catalytic environment.[27,28] However, extensive brute force optimization of all variables involved would still make theoretical design of theozyme composed of several hundreds of amino acids extremely costly and impractical. We believe that systematic nonempirical analysis of the physical nature of interactions involved in the catalysis and development of additional techniques specific for enzymes presented in this contribution could lead to rational construction of simpler models useful in theozyme design. Other

rare inverse catalyst design attempts that have been recently summarized in a review[29] are mostly based on artificial intelligence or data mining. Although such methods could be very useful in the automated scanning of vast chemical spaces, it would be surprising to obtain a deeper understanding of physical phenomena responsible for extraordinary catalytic activity of enzymes this way, while many essential structural details are missing in currently available databases.

Recent experimental results that employ Stark effects for a specific band of infrared (IR) spectra for mutated enzymes indicate the electrostatic nature of KSI catalytic activity.[30] This permits one to monitor electric fields exerted by specific mutations on certain reactant bonds in top-down fashion, yielding valuable but still limited information.[16,31] This perspective is reversed here in bottom-up fashion by the use of a catalytic field derived from the reactant and transition state wave functions, yielding general characteristics of charge distribution of the optimal catalytic environment for every point in space outside reagents and enabling one to monitor lowering of activation barrier directly, as a result of amino acid mutations, side-chain rotations, rate-promoting vibrations, or proton transfers in hydrogen bond chains.

Although the catalytic field technique has been derived in our laboratory[24−26] using the theory of intermolecular interactions, it could be also applied to systems involving intramolecular interactions. The ability to predict catalytic effects resulting from hydrogen to fluorine (H → F) substitutions has been illustrated for two simple model organic reactions,[32] which demonstrates qualitative applicability of this technique to model substrate-assisted catalysis.[33]

The present paper constitutes the first application of a catalytic field approach to real case intermolecular catalysis. For this purpose, here, we have chosen the artificial Kemp eliminase KE07 and its mutants.[9] Among other Kemp eliminases based on different scaffolds, this system has the largest number of mutants with gradually increasing activities. Several KE07 mutants introduced by LDE experiments, which apparently were missed in earlier theozyme designs,[13,14] constitute a suitable benchmark to examine new biocatalyst design methodologies.

The mutants obtained from LDE experiments differ mainly by the number of charged amino acids located in the second coordination sphere,[13,14] which may change conformation of its side chain upon docking with the substrate or nearby mutation.[34,35] The importance of considering amino acid side-chain rotamers in biocatalysis has been recently demonstrated for Kemp eliminase[36] and triosephosphate isomerase mutants.[37] A more general catalytic field technique,[25,26] considering simultaneously both TSS[8,9] and ground-state destabilization,[23] combined with the atomic multipole rotamer library scan, allows one to obtain valuable insight into active site preorganization.

This contribution also constitutes the first attempt to test and validate the catalytic field technique and apply it to explore possible mutation nonadditivity, as well as dynamic and epistatic effects in biocatalysis resulting from side-chain rotations.

■ METHODS

Here, we present a bottom-up strategy based on a concept of catalytic fields derived from the DTSS approach.[25,26] Within this ab initio-based framework, activation barrier lowering is partitioned into *intrinsic* (*vacuum* in the original paper[25]) contribution of reactants and the difference of enzyme-transition state(TS) and enzyme−substrate(S) interaction energies E_{DTSS}.

Whenever the DTSS approach is applied to enzyme reactions, the term "substrate" refers to the ground state (alternatively, the reactant complex or Michaelis complex). In principle, the DTSS approach is more general and could be applied to other catalytic systems, such as zeolites.[38]

Technically speaking, the "intrinsic" contribution is a barrier calculated for reaction path coordinates as they are in catalysts, but without inclusion of molecular environment in quantum mechanics (QM) calculation. Despite the virtual character of this reference state and the DTSS energy for a single catalyst, it can be used as the reference in quantitative comparison of a sequence of mutants of the same enzyme. Provided that mutation does not change the reaction mechanism, taking the difference between mutant and wild-type (WT) DTSS energies cancels out the "intrinsic" contribution, leaving a value corresponding to the difference of reaction barriers of mutated and WT enzyme (see Figure 1):
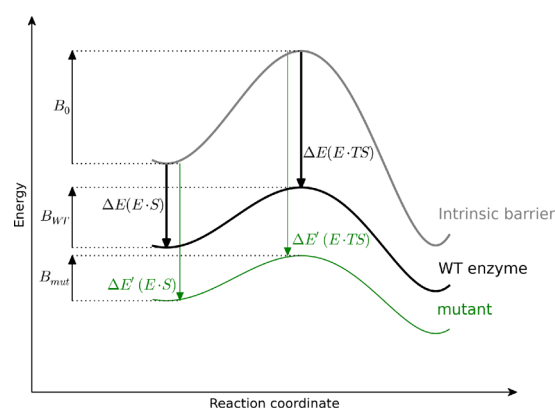

Figure 1. Differential transition state stabilization (DTSS) energy $E_{DTSS} \equiv E_{IE}(TS) - E_{IE}(S)$, as a measure of catalytic efficiency; under the assumption that mutation does not change reaction mechanism, the difference $E_{DTSS}^{mut} - E_{DTSS}^{WT}$ corresponds to the difference of activation barriers $B_{mut} - B_{WT}$.

$$E_{DTSS} \equiv \Delta E(E \cdot TS) - \Delta E(E \cdot S)$$
$$\approx \Delta E_{EL,MTP}(E \cdot TS) - \Delta E_{EL,MTP}(E \cdot S) \quad (1)$$

Furthermore, DTSS could be partitioned into the electrostatic multipole $\Delta E_{EL,MTP}$, electrostatic penetration $\Delta E_{EL,PEN}$ exchange ΔE_{EX}, delocalization (induction) ΔE_{DEL}, and correlation (dispersion) ΔE_{CORR} terms defined within symmetry-adapted perturbation or hybrid variation-perturbation theories of intermolecular interactions.[39,40] This provides not only insight into the physical nature of catalytic interactions induced upon mutation, but it also allows one to derive, in a systematic way, more approximate methods based on leading terms. Recent experimental results[30] and ab initio analysis of several enzyme systems[26,41,42] indicate clearly the dominant contribution of electrostatic term ΔE_{EL} and, in particular, its long-range multipolar electrostatic component $\Delta E_{EL,MTP}$ accurately reflecting, at an atomic level, the molecular charge redistribution during the progress of the reaction.[43] The electrostatic multipolar DTSS term,[26] constituting a more-general approach, represents TSS[44] and, at the same time, ground-state destabilization (GSD)[23] effects, postulated earlier as the

important contributions to enzyme catalysis. Our recent analysis of several mutants of ketosteroid isomerase[45] validated the utility of simple model, based on cumulative atomic multipole expansion (CAMM),[46] in agreement with recent experimental data,[30] confirming the electrostatic nature of catalytic activity.

■ CATALYTIC FIELDS

Further consideration of electrostatic component of DTSS energy leads to the static catalytic field $\Delta_S(\vec{r}_i)$, which is defined[25,26] as the difference between electrostatic potentials $V(\vec{r}_i)$ of TS and substrate S (see eq 1). Its value denotes activation barrier change resulting from the presence of unit point charge in any location of catalytic site outside reactants. Because of the additive nature of electrostatic interactions, catalytic field values obtained in one computational step constitute the solution of inverse catalysis problem in the form of $\Delta_S(\vec{r}_i)$, i.e., complementary charge distribution of optimal molecular environment $q_i(\vec{r}_i)$.[25,26]

$$E_{\text{DTSS}} \simeq \sum_i q_i(\vec{r})(V^{\text{TS}} - V^{\text{S}})(\vec{r}_i) \simeq -\sum_i q_i(\vec{r})\Delta_S(\vec{r}_i)$$

(2)

Therefore, a catalytic field may serve as a general guide for rational selection of mutations or side chain orientations or rate-promoting vibrations by providing clues about optimal static and dynamic charge distribution of catalytic environment reducing the activation barrier. The values of $\Delta_S(\vec{r}_i)$ could be rapidly calculated from CAMM derived directly from transition-state and substrate quantum-chemical wave functions.[46] Therefore, calculations involving catalytic field scale as $O(N_R \cdot N_A \cdot N_{\text{rot}})$ only, where N_R and N_A indicate the number of atoms in reactants and considered mutated amino acid residue, respectively. N_{rot} denotes the average number of rotamers per mutation site (typically a few dozens). In the case when charged amino acid is represented by single unit charge located at the charge gravity center (for example, located near the terminal side chain atom of lysine or arginine), scaling could be further reduced to $O(N_R \cdot N_{\text{rot}})$, which must be calculated only once for the mutation site. The validity of such approximation is illustrated in Figure 2.

This way, the search space in biocatalyst design is considerably reduced within purely nonempirical approach. The use of atomic multipoles is essential to represent the anisotropy of molecular charge distribution properly at the atomic level, in particular for transition states, where bonds are formed or broken.[43] In addition, always arbitrary atomic charge definitions are compensated by the use of higher-order atomic multipoles, considerably reducing the basis set dependency of atomic charges.[46]

Note that Δ_S gradient field defined as dynamic catalytic field Δ_D, indicates direction of the favorable unit probe charge $q_i(\vec{r}_i)$ movement coupled with the reaction coordinate and resulting in additional decrease of the activation barrier.[47] Such coupling may involve proton dislocations in hydrogen-bonded networks. Recent application of the dynamic electrostatic catalytic field (DECF) approach allowed us to determine that each of two steps of reaction proceeding in ketosteroid isomerase (KSI) is catalyzed by concerted proton dislocations proceeding in reverse directions in the network of six hydrogen bonds.[48] Therefore, observed exceptional KSI catalytic activity could be related to ultrafast proton switching.[48]
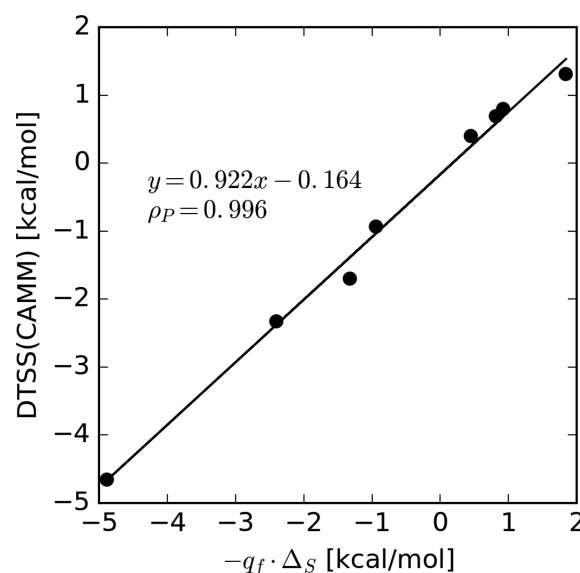

Figure 2. Activation energy changes obtained using cumulative atomic multipole moments DTSS(CAMM) versus products of catalytic fields Δ_S and formal entire amino acid charges $-q_f$ located at its side-chain charge gravity center (Arg- guanidyl carbon, Lys- ^{6}N, Asp- C_γ, Glu- C_δ).

■ EXPLORATION OF ROTAMER SPACE

In this study, we will concentrate on some unexplained mutations introduced by LDE into artificial Kemp eliminase, which are not in direct contact with the substrate.[13,14] According to the aforementioned assumption about second-coordination sphere mutations (that is, their negligible effect on the reaction mechanism), the following calculations base on transition state obtained for KE07 enzyme with Gaussian09[49] ONIOM implementation[50] (details are provided in section S2 of the Supporting Information). The QM region of that simulation, consisting of 5-nitrobenzoixazole (5NBI) and acetate representing part of Glu101 residue, was superimposed on 5NBI-enzyme complexes of the mutants.

The space of possible amino acid conformations was explored in two ways. In the first, conventional molecular dynamics (MD) has been used with solvent effects modeled within the Molecular Mechanics Poisson−Boltzmann Surface Area (MMPBSA) model,[51] which is regarded as a reliable free-energy simulation method to investigate protein−ligand interactions. This provides a baseline representing a commonly used approach to the problem. In the second one, the library of atomic multipoles for amino acid side chain rotamers has been scanned using catalytic fields, providing a proof of concept of a proposed approach (note that, with proper implementation, this involves much less computational time than MD simulation and avoids one having to always use arbitrary force fields).

■ MMPBSA APPROACH

First, 30-ns MD trajectories were prepared for each of eight mutants, followed by MMPBSA calculation with the usage of the gMMPBSA program[51] MD simulations were performed in Gromacs 4.6.[52] For the sake of consistency with ONIOM calculations, Amber94 force field with TIP3 water model was used. Mutant structures were generated from KE07 by a homemade script, according to published data[9] (see Table S1 in the Supporting Information). All structures were submerged in water box at least 1 nm from the protein atoms, with Na and Cl ions for charge neutralization. In all simulations, periodic

boundary conditions and Particle Mesh Ewald long-range electrostatics (1 nm cutoff) were used. For MD simulations, a leapfrog algorithm was used, and the temperature and pressure were set to 300 K and 1 bar, respectively. After initial steepest descent minimization, NVT and NPT equilibration stages were performed, each one 200 ps long. In this phase, a modified Berendsen thermostat and a Parinello-Rahman isotropic barostat were chosen. In subsequent 30 ns production run, distance restraints between Glu101 OE2 and ligand C3 and H3 atoms were used, in order to prevent random dissociation of the complex. Then, trajectories ware used in MMPBSA calculations done with gMMPBSA program.[51]

MD simulations performed for KE07 and its mutants concerned only the structure of the enzyme–substrate complex. It was assumed that the reaction event, when it occurs, is much faster than conformational changes of the surrounding amino acid residues. Therefore, these simulations were supposed to provide an ensemble of protein conformations rather than to simulate the reaction itself. Furthermore, the application of catalytic field Δ_S could be tested in this way. Taking advantage of the minor changes in geometry between reactant and TS (see Figure S3b in the Supporting Information) and keeping the assumption mentioned above, the same trajectory was assumed for TS. Different energies were obtained by substitution of charges assigned to the "QM region" (as defined in ONIOM calculations) in the topology files. Since g_mmpbsa evaluates electrostatic and implicit solvation contributions to binding energy (we included only electrostatic components), DTSS energy with implicit solvation was obtained by putting *differential charges* (the difference between each atomic charge in TS and the reactant), which effectively are an atomic monopole representation of Δ_S. Atomic charges for RS and TS were calculated using the Merz–Kollman scheme[53] in the same basis as that used in ONIOM calculation. The contribution of QM region to solvation energy was subtracted from the result, since this method of calculation provides incorrect value of this term; on the other hand, it may be assumed to be negligible or at least constant among mutants (for a more-detailed derivation, see Section S3 of the Supporting Information).

SCANNING THE CAMM LIBRARY OF ROTAMERS

As the major approach in this study, we used the Dynameonics Rotamer Database (backbone-independent, updated in 2015)[54] to prepare a library of CAMM for each amino acid rotamer.

All CAMMs were calculated with GAMESS (US)[55] with HF method and 6-311G(3d,2p) basis set. CAMMs for the QM layer in both RS and TS states were obtained separately. The CAMM expansion was then used in the calculation of catalytic fields and interaction energies. For each mutation, rotamers with the best and worst TSS/DTSS contribution were determined (hereafter, we will refer to the rotamer with the best DTSS contribution as optimal). Finally, we explored the configuration space resulting from combining rotamers at different sites (mathematically speaking, a Cartesian product of individual site rotamer spaces in a given mutant).

In order to test the importance of solvation effects on electrostatic interaction energies, we examine the version of library incorporating Generalized Born (GB) model, where the solvation energy is modeled according to eq 3. The Born radii were calculated using *bornRadius* software developed by Onufriev's group, utilizing an improved algorithm described in refs 56 and 57. The atomic charges for amino acid residues were taken from Amber94 force field, whereas charges on the QM

region were calculated with the Merz–Kollman method for the HF/6-311G(3d,2p) wavefunction.

$$\Delta G_{solv} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}}\right)\sum_{i,j}\frac{q_i q_j}{f_{ij}}$$

$$f_{ij} = \sqrt{r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)}$$

(3)

ALGORITHMS FOR THE EXPLORATION OF ROTAMER CONFIGURATIONAL SPACE

Since the number of rotamer combinations grows exponentialy with the number of amino acids, it is important to restrict the size search space, especially in the early steps of calculation. In particular, elimination of rotamers that are unlikely to participate in final combination is of great importance. In our experiments, we found that the average number of rotamers per site is in the range of 60–80. Thus, elimination of highly improbable rotamer (e.g., having short contact with the protein backbone) reduces the size of search space by $70(N - 1)$ configurations (with N being the number of included residues). Here, two criteria were applied: (i) geometric and (ii) energetic. The first one basically involves exclusion of rotamers which side chains have close contanct (≤ 1 Å) with protein backbone (neglecting, obviously, contacts corresponding to chemical bonds). The second one is the well-known Goldstein criterion for singles elimination.[58]

$$E_k(r_k^A) - E_k(r_k^B) + \sum_{l=1}^{N}\min_X(E_{kl}(r_k^A, r_l^X) - E_{kl}(r_k^B, r_l^X)) \geq \epsilon$$

(4)

Here, subscripts $k,l \in [1,N]$ denote the position in the amino acid sequence, whereas superscripts "A" and "B" indicate the rotamer at a particular site. However, the value ϵ in the original work was equal to zero; by choosing a positive nonzero value, one can "soften" the criterion and thus obtain a larger final space. Such an approach may be useful when one is interested in a set of low-energy conformers, which may be particularly desirable when considering dynamic effects.

The search algorithm generally can be described by the following steps.

(1) Loading of the amino acid rotamers into specified positions in a protein structure. At this stage, criterion (i) is applied.
(2) Calculation of an interaction energy table.
(3) Elimination of the rotamers according to different criteria (e.g., criterion (ii)).
(4) Systematic search over resulting space.

Throughout the presented work, we refer to the method without additional elimination criteria in point 3 as "multiscan". The second algorithm tested, denoted as Dead End Elimination (DEE), involves criterion (ii) with negligence of the rotamer self-energy ($E_k(r_k^A)$) and ϵ set to either 0.0 or 0.005 hartree. The performance of both approaches was tested on P_6 mutant represented by five charged residues: Asp7, Glu19, Lys146, Arg202, and Asp224. A potential bottleneck may be the calculation of CAMM interaction energies (which can be several times slower than that observed for the point-charge model) or loading CAMM from library (since the rotation of large multipole arrays may be computationally expensive). Therefore, we compared the performance of the RESP model
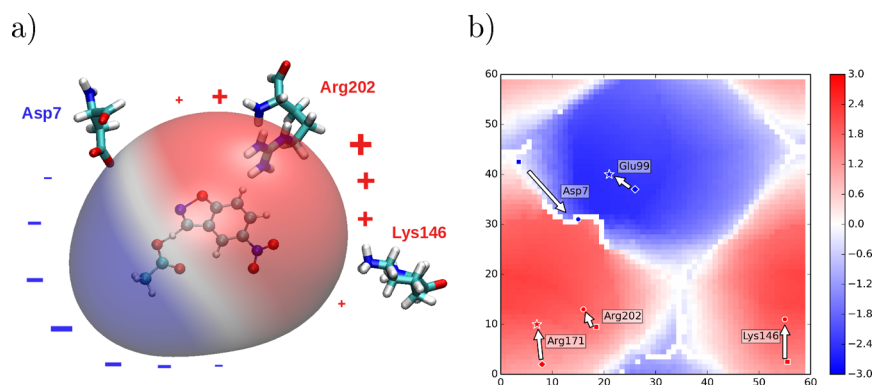
Figure 3. Catalytic field in Kemp eliminase. (a) 3D view of surface colored with the value of Δ_S (red indicates that positively charged catalyst residues will be beneficial, blue indicates that negatively charged residues will be beneficial); (b) Kohonen map[60] of the catalytic field (values given in kcal/(mol e)) (positions of amino acids at a distance of 7 Å above the van der Waals surface are marked with circles (optimal DTSS) and squares (the worst DTSS); residues predicted to provide further improvement are marked with stars (optimum DTSS) and diamonds (the worst DTSS); the arrows indicate the transition from the position with the worst DTSS to that with the best one.

(Table S1 in the Supporting Information) and CAMM (Table S2 in the Supporting Information). We ensured that both approaches lead to the same rotamer configuration.

Two major observations can be drawn from the presented results.

- Elimination of rotamers according to Goldstein criterion reduces the search space by 2−4 orders of magnitude, depending on the assumed threshold. This, in turn, leads to a similar reduction of time spent on scanning the rotamer space.

- Application of CAMM mutipoles leads to an ~2-fold increase of loading time (due to the required transformation of multipole arrays) and an ~6-fold increase in the time spent on the computation of interaction energy table. In fact, this time is comparable to that required to scan the entire (unreduced) rotamer space.

However, for larger sets of amino acid residues, the problems with application of CAMM expansion can be tackled by initial reduction of search space, based on atomic charges, maybe with a nonzero *epsilon*. This should allow one to avoid situations when inclusion of the charge anisotropy leads to different configurations.

Furthermore, there are at least two possibilities, with regard to further improvement of the presented algorithm:

- The elimination of unlikely pairs from the search space (analogous to the Goldstein criterion for singles).

- Further approximation of the energy table. For larger rotamer sets, this could become a serious bottleneck, since its full preparation is required in all presented algorithms. However, for residues distant in space, there should be a little difference in interaction energy between their different rotamers (in other words, the rotation of side chains should not change the interaction energy). In such a case, an entire block of the interaction energy table could be filled at once.

■ FLOWCHART OF COMPUTATIONAL PROCEDURE

(A) Structure preparation

(1) Preparation of CAMM amino acid library (HF/6-31G*); structures of rotamers taken from the Dynameonics database[54]

(2) Calculation of reaction path in KE07 (ONIOM B3LYP/6-311+G**:Amber94)

(3) MD simulations of KE07 with distance restraint put on hydrogen bonds between Asp and 5-nitrobenzoisoxazole (5NBI)

(4) Clusterization of production trajectory with g-clust algorithm (available in Gromacs). The main cluster was then used for analysis of the conformational space

(B) Exploration of rotamer space

(1) Only amino acid residues that varied in directed evolution experiments[9] were included. Rotamers from the CAMM library were loaded into these positions, excluding those having short contacts with either backbone or ligand (short-contact is defined as the shortest interatomic distance below 1.25 Å)

(2) Evaluation of interaction energy table in either the GB model (see below) or with the corresponding point charge model without solvation

(3) Elimination of pairs of rotamers (of different residues) yielding short contacts

(4) Search of the lowest energy configuration with different algorithms
  - brute-force algorithm (direct iteraton over all possible configurations) (referred to as "multiscan")
  - Dead End Elimination, DEE (the elimination of rotamers, according to Goldstein criterion for singles.[58]
  - Mean-Field algorithm.[59]

(5) Evaluation of DTSS(CAMM) or DTSS(GB) energy for the found configurations

■ RESULTS

The Catalytic Field and Optimal Locations of Crucial Amino Acids. The three charged residues—Asp7, Lys146, and Arg202, conserved in directed evolution[9] (see Table S1)—seem to play the most important role. What's even more remarkable, these findings are consistent with the catalytic field corresponding to this reaction. As can be seen in Figure 3, charged amino acids conserved in directed evolution reside exactly where they should be located according to catalytic field Δ_S, and the range of their possible rotamers fits the corresponding "basins" well. The position of Asp7 on the edge between the two areas of Δ_S

provides a straightforward exploration of both its low contribution and range varying from positive to negative (the best and worst DTSS from the rotamer scan are −1.1 kcal/mol and 0.0 kcal/mol, respectively).

Figure 3a presents catalytic field $\Delta_S(\vec{r}_i)$ for the Kemp eliminase, calculated according to eq 2 on the van der Waals surface of reactants, as well as optimal rotamers of crucial charged residues.

Side-Chain Movements and the Ranges of DTSS Contributions. Charged amino acid side chains, especially of lysines and arginines, belong to most flexible upon ligand binding,[34,35] so it is necessary to consider entire range of their dislocations in more detail. Results obtained using library of atomic multipoles for amino acid side chains indicate that activation energy changes may vary between 1 kcal/mol and 4 kcal/mol for charged residues, in contrast to neutral amino acids (0−0.5 kcal/mol). The possible movements of charged amino acid terminal atoms representing extreme values of catalytic activity measured by DTSS are illustrated on a 2D Kohonen map of the Kemp eliminase catalytic field in Figure 3b.

In order to make a 2D representation of catalytic field, we used a self-organizing map (SOM) based on the Kohonen neural network.[60] First, a surface at a distance of 7 Å from the vdW surface of TS was generated with the point density set to 25 points/Å². It was then used to train a neural network consisting of 3600 neurons arranged into a square grid with a toroidal topology. This special topology means that the grid (and thus the resulting map) does not have an edge: the "rightmost" point is topologically adjacent to the "leftmost" point, while the "topmost" point neighbors the "bottom-most" point. Another consequence of this fact is that replicas of such map may be placed next to each other, similar to tiles, showing the topological surroundings of each point. After training, the map was colored according to the value of catalytic field Δ_S, and terminal atoms of amino acids (chosen in the same manner as presented on Figure 2) were projected onto resulting SOM.

These side chain conformation changes could be regarded as the important part of a preorganized catalytic environment. Relatively large activation barrier changes resulting from amino acid side chain movements reinforce the conclusions from previous experimental and computational studies, indicating the importance of the precise positioning of catalytic residues.[61]

Multiscan Results Including Solvent Effects—Comparison with Experiment. Enzyme catalytic activity could be expressed in terms of TS and S interaction energies with the enzyme treated as a rigid body. However, the docking substrate may induce significant changes in side-chain conformation of charged amino acids,[34,35] which may strongly interact between themselves as well as with TS and S. Whereas the interaction energy for typical neutral hydrogen-bonded dimer is 3−5 kcal/mol, it can be 1 or 2 orders of magnitude higher, if at least one of the monomers bear nonzero electric charge (specifically, the absolute value of interaction energy may reach 20 kcal/mol in the case of neutral molecules bonded to a charged one or even 100 kcal/mol when both partners are charged). Therefore, the interactions involving charged amino acids are expected to be dominant and a model relying solely on this contribution seems to be a reasonable first approximation. In the case of Kemp eliminase, six positions varied in LDE experiments involved charged residues: 7, 19, 123, 146, 202, and 224 as shown in Tables 1 and 2. The change of the charged state of these residues located beyond the first coordination sphere and directly involved in reactant binding has the greatest chance to modulate

Table 1. DTSS (CAMM) Values Resulting from Multiscan*

Enzyme	Residue 7		Residue 19		Residue 123		Residue 146		Residue 202		Residue 224		Total
P$_0$	I	0.00	K	-0.85	Q	0.00	K	-1.36	G	0.00	N	0.00	-2.21
P$_1$	I	0.00	E	0.40	R	1.75	T	0.00	R	-4.66	D	2.66	0.15
P$_2$	Q	0.00	K	-0.93	Q	0.00	T	0.00	R	-4.66	D	2.26	-3.32
P$_4$	D	0.02	K	-0.93	Q	0.00	K	-1.36	R	-4.66	D	2.26	-4.66
P$_5$	D	-1.12	E	0.40	Q	0.00	T	0.00	R	-4.66	D	2.66	-2.80
P$_3$	D	-1.12	K	-0.93	Q	0.00	E	1.19	R	-4.66	D	2.26	-3.32
P$_6$	D	0.02	K	-0.93	Q	0.00	K	-1.36	R	-4.66	D	2.26	-4.66
P$_7$	D	-1.12	T	0.00	Q	0.00	K	-2.13	R	-4.66	D	2.66	-5.33

*Values given in units of kcal/mol. For noncharged residues, DTSS has been assumed to be zero. The residue charge code: red, positive; blue, negative; black, neutral.

Table 2. Contributions of Individual Amino Acids to Total Activation Barrier Lowering DTSS (GB) Including Solvent Effects Estimated via the Generalized Born Approach*

Enzyme	Residue 7		Residue 19		Residue 123		Residue 146		Residue 202		Residue 224		Total
P$_0$	I	0.00[1]	K	-0.12	Q	0.00	K	-0.60	G	0.00	N	0.00	-0.72
P$_1$	I	0.00	E	0.04	R	0.40	T	0.00	R	-1.95	D	0.73	-0.78
P$_2$	Q	0.00	K	-0.15	Q	0.00	T	0.00	R	-0.98	D	0.21	-0.92
P$_4$	D	-0.24	K	-0.15	Q	0.00	K	-0.89	R	-0.70	D	0.21	-1.75
P$_5$	D	-0.47	E	0.04	Q	0.00	T	0.00	R	-1.95	D	0.73	-1.65
P$_3$	D	-0.24	K	-0.15	Q	0.00	E	0.33	R	-1.95	D	0.21	-1.79
P$_6$	D	-0.24	K	-0.15	Q	0.00	K	-0.89	R	-0.70	D	0.21	-1.75
P$_7$	D	-0.47	T	0.00	Q	0.00	K	-0.89	R	-1.95	D	0.73	-2.53

*Values given in units of kcal/mol. For noncharged residues, DTSS has been assumed to be zero. The residue charge code: red, positive; blue, negative; black, neutral.

a catalytic field without affecting the reactant binding pose. In addition, they may strongly interact with each other and substrate binding may affect the conformation of their charged side chains. Therefore, multiscan procedure seems to a reasonable way to model these effects, which could not be determined by currently available conventional experimental or theoretical methods.

For all mutants, residues at these positions have been taken into account in a search for the lowest-energy rotamer configuration using a CAMM rotamer library (http://camm.pwr.edu.pl/CAMM).

In order to take into account solvent presence, the GB model has been used with selection of effective Born radii computed with the *bornRadius* program developed by Onufriev and co-workers.[56,62] With this energy function (denoted henceforth as GB), a search over rotamer space was performed in exactly the same way as the CAMM model. Interactions between amino acid residues were included.

Finally, we compare those results to MMPBSA[51] calculations based on 30-ns MD simulations performed for each mutant separately (more details are given in the section devoted to the MMPBSA approach). Experimental activation free energies for Kemp eliminase mutants are compared in Figure 4 with all three aforementioned models. Remarkably, both models involving a scan over rotamer space—that is, DTSS(CAMM) and DTSS-(GB)—outperform the DTSS(MMPBSA) approach, in terms of the number of correctly ordered pairs N_{pred} (See Tables 1 and 2). N_{pred} of a model[63] is a ratio of correctly predicted pairs to all pairs in a test set. A "correctly predicted" or concordant pair is a pair of mutants P,Q for which $E^P_{\text{DTSS}} < E^Q_{\text{DTSS}}$ and $k^P_{\text{cat}} > k^Q_{\text{cat}}$ or $E^P_{\text{DTSS}} > E^Q_{\text{DTSS}}$ and $k^P_{\text{cat}} < k^Q_{\text{cat}}$; otherwise, such a pair is counted as nonconcordant. Given the total number of pairs N_{tot} in a test set consisting of N enzymes (eq 5), the value of N_{pred} (usually
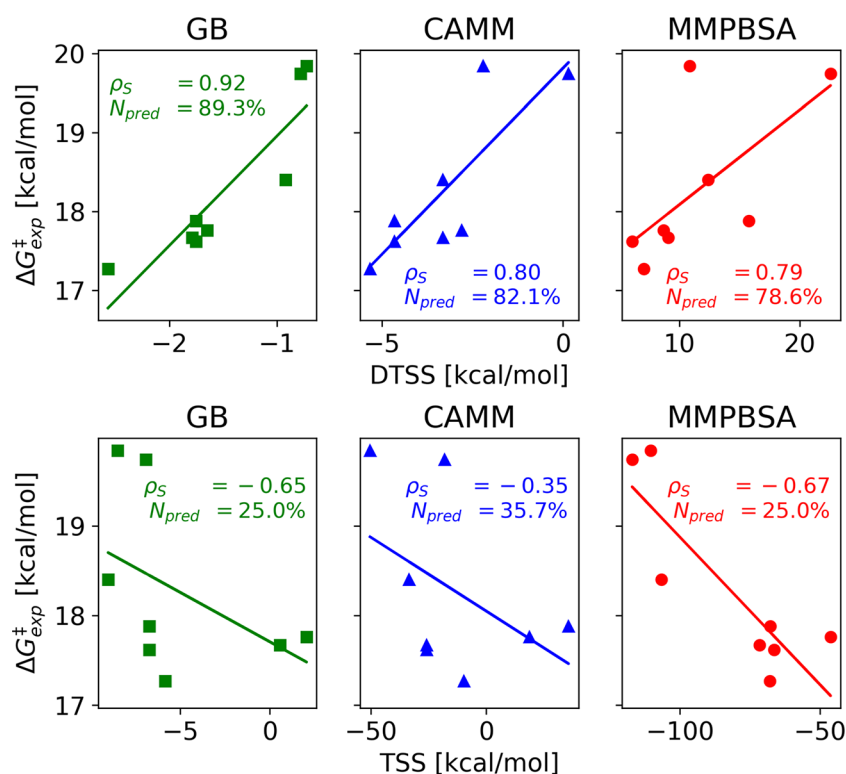
Figure 4. Comparison of theoretical DTSS and TSS results with experimental data for Kemp eliminase mutants obtained from directed evolution experiments.[8] In the CAMM approach, reactant−amino acid side-chain interactions were calculated by employing cumulative atomic multipole moment expansion. In the GB approach, corresponding ESP atomic charges have been used, including solvent effects estimated from the Generalized Born method. MMPBSA denote mean interaction energies using MD trajectories where solvent effects were obrained from the Poisson−Boltzmann model.[51]

expressed as a percentage) is given by eq 6. It is related to Kendall's tau τ_K by the relation described in eq 7).

$$N_{\text{tot}} = \frac{N(N-1)}{2} \quad (5)$$

$$N_{\text{pred}} = \frac{N_{\text{con}}}{N_{\text{tot}}} \quad (6)$$

$$\tau_K = \frac{N_{\text{con}} - N_{\text{non-con}}}{N_{\text{tot}}} = 2N_{\text{pred}} - 1 \quad (7)$$

While DTSS results obtained within GB, CAMM, and MMPBSA models, accounting for continuous solvent, correlate with experimentally observed trends (Spearman correlation coefficients 0.92, 0.80, or 0.79 and predictivity rates of 89.3%, 82.1%, or 78.6%), the TSS model anticorrelates (Spearman coefficients −0.65, −0.35, or −0.67 with predictivity rates of 25.0%, 35.7%, or 25%) confirming the essential role of neglected ground-state destabilization.[23] This was probably one of the reasons of limited success of initial theozyme design.[13,14]

Because of the approximations involved, the correspondence is not linear; however, this does not disqualify the overall approach, which is intended to determine the ranking or narrow the large set of possible mutations, so it would be manageable for more sophisticated methods. Thus, it seems to be sufficient to indicate a general qualitative trend, favoring the DTSS model over the TSS model.

■ NONADDITIVE EFFECTS OF MUTATIONS

Activation barrier changes (DTSS) originating from specific residue could sometimes vary as much as 1.25 kcal/mol, indicating possible side-chain conformation changes, resulting from interactions between other residues (see Table S1 in the Supporting Information). Such variations responsible for nonadditivity of mutations could be examined in detail using the multiscan technique presented here. Let us discuss in detail ASP7 catalytic activity in P3, P4, P5, P6, and P7 mutants. Figure 5a−e presents schematic side-chain orientations obtained using a multidimensional scaling technique, preserving Euclidean distances in the space of lower dimensionality. Mutation GLU146LYS in P4 forces change of ASP7 conformation in P3 and, as a result, a loss of its activity measured by DTSS from −1.12 kcal/mol to +0.02 kcal/mol. Then, mutations LYS19-GLU and LYS146XX (where XX rerpesents a neutral amino acid) introduced into P4 restore ASP7 activity in P5. In P6, the previous mutation is reversed, resulting in the same activity loss as that observed in P4. Finally, the LYS19XX mutation in P6 restores ASP7 activity in P7 again. All of these results indicate the essential role of molecular context, which can be explored by the multiscan technique presented in the previous chapter. Additional analysis of nonadditive effects of LYS146 and Arg202 is presented in section S4 of the Supporting Information.

■ DISCUSSION

The presented scheme correctly explains the effect of mutations leading from KE07 to R7−R10/11G, providing an atomistic insight into the possible role of each residue contribution to the activation barrier lowering, emerging from first principles.
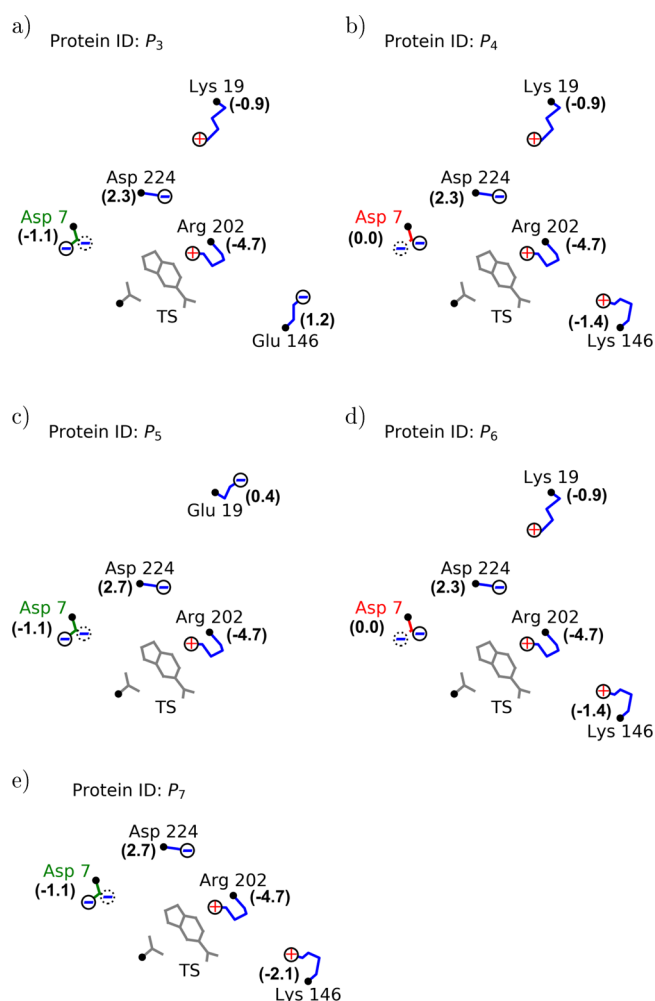
Figure 5. Optimal side-chain orientations in Kemp eliminase active site determined for P3−P7 mutants. DTSS values are given in units of kcal/mol.

The most significant result of this analysis is an effective approximate solution of the "inverse catalysis problem" confronted and confirmed for the first time by experimental results. We have shown that, based on the information provided by the catalytic field, the search space could be narrowed down to a limited number of computational steps. The underlying electrostatic differential transition state stabilization DTSS concept[25,26] has been confirmed later in recent independent studies.[18,23,31] Jindal and co-workers analyzed the differences between evolutionary pathways of two Kemp eliminase families: descendants of KE07 and HG3.[23] They noted that, in the first group, the improvement arises from ground-state destabilization, whereas, in the second one, te improvement results from an increase in transition-state stabilization (TSS). This is consistent with our more general DTSS and catalytic field approach, which covers both effects simultaneously.[25,26] Moreover, some elements of differential stabilization of the transition state approach appeared later in two other works,[18,22] which provides further reinforcement for this idea.

In summary, the role of residues 7, 19, 123, 146, 202, and 224 modified in laboratory-directed evolution experiments, which are not in direct contact with reactants, seems to be modulation of the catalytic field to minimize activation barrier, without changing the reactant binding pose determined primarily by the amino acids constituting the first coordination sphere.

The essential advantage of the present approach is the use of cumulative atomic multipole moments to describe molecular charge redistribution during the progress of the reaction, capturing the anisotropic character of molecular electrostatic potentials,[64] which may be more significant in neutral systems.[43] It may be still crucial in proper description of the catalytic field, since Δ_S has at least dipole character, because of the charge conservation principle. Another important feature of the catalytic field is that, despite special emphasis placed on charged residues in the present analysis, it can provide more general information; by considering its gradient $\bar{\Delta}_d$, one gains insight into the optimal distribution of catalytic dipoles in a way analogous to that presented above for charges (see Figures S1 and S2 in the Supporting Information). Finally, as mentioned earlier, whenever necessary, the DTSS energy can always be generalized to include remaining interaction energy terms defined in interaction energy decomposition schemes, such as SAPT[39] or HVPT.[40] For example, exchange repulsion and dispersion terms could be estimated from the corresponding nonempirical atom−atom functions approximating accurate SAPT results.[65,66] An important issue is a nonadditivity of mutations reported recently.[23] Within the presented framework, it could be included by exploration of the effect of mutation on rotamer distribution of other residues (and corresponding DTSS energy changes). An example of such analysis is presented in Figure 5, where we found that considering the mutual interactions of residues Asp7, Lys19, Lys146, Arg202, and Arg2246 in P3, P4, P5, P6, and P7 mutants changes the optimal DTSS contribution of Asp7 several times from −1.1 (P3), 0.0 (P4), −1.1 (P5), 0.0 (P6), and −1.1 (P7) kcal/mol.

Such an approach can be effectively used to obtain rapid estimates of nonadditive effects in multiple mutants. The general idea presented in this work may be easily introduced into other existing protocols, such as an ensemble of short MD[67] or Monte Carlo simulations.[36] Application of static catalytic field gradients (dynamic electrostatic catalytic fields DECF)[47] allows one to predict the catalytic role of proton transfer in enzyme hydrogen-bond networks.[48]

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00139.

> Discussion regarding dynamic catalytic fields, summary of KE07 mutant data, details of ONIOM and MMPBSA simulations, and additional tests on nonadditive mutation effects and execution times for exploration of rotamer space (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

W. Andrzej Sokalski − *Wroclaw University of Science and Technology, Wroclaw, Poland;* ⓞ orcid.org/0000-0001-5081-8175; Email: Andrzej.Sokalski@pwr.edu.pl

Wiktor Beker − *Wroclaw University of Science and Technology, Wroclaw, Poland;* Email: Wiktor.Beker@wp.eu

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00139

■ **REFERENCES**

(1) Świderek, K.; Tunon, I.; Moliner, V. Predicting enzymatic reactivity: from theory to design. *WIREs Comput. Mol. Sci.* **2014**, *4*, 407−421.

(2) Wijma, H. J.; Janssen, D. B. Computational design gains momentum in enzyme catalysis engineering. *FEBS J.* **2013**, *280*, 2948−2960.

(3) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320−327.

(4) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539−543.

(5) Dahiyat, B.; Mayo, S. De novo protein design: Fully automated sequence selection. *Science* **1997**, *278*, 82−87.

(6) Dahiyat, B.; Mayo, S. Protein design automation. *Protein Sci.* **1996**, *5*, 895−903.

(7) Bolon, D.; Mayo, S. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14274−14279.

(8) Röthlisberger, D.; Khersonsky, O.; Wollacott, A.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J.; Althoff, E.; Zanghellini, A.; Dym, O.; et al. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190−195.

(9) Khersonsky, O.; Rothlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **2010**, *396*, 1025−1042.

(10) Khare, S. D.; Kipnis, Y.; Greisen, P. J.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y. F.; Gallaher, J. L.; Silman, I.; Leader, H.; et al. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol.* **2012**, *8*, 294−300.

(11) Althoff, E. A.; Wang, L.; Jiang, L.; Giger, L.; Lassila, J. K.; Wang, Z. Z.; Smith, M.; Hari, S.; Kast, P.; Herschlag, D.; et al. Robust design and optimization of retroaldol enzymes. *Protein Sci.* **2012**, *21*, 717−726.

(12) Preiswerk, N.; Beck, T.; Schulz, J. D.; Milovnik, P.; Mayer, C.; Siegel, J. B.; Baker, D.; Hilvert, D. Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8013−8018.

(13) Arnaud, C. H. Proteins made to order. *Chem. Eng. News* **2012**, *90*, 8.

(14) Arnaud, C. H. Enzyme by design. *Chem. Eng. News.* **2013**, *91*, 26−27.

(15) Osuna, S.; Jimenez-Oses, G.; Noey, E. L.; Houk, K. N. Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc. Chem. Res.* **2015**, *48*, 1080−1089.

(16) Vaissier, V.; Sharma, S. C.; Schaettle, C.; Zhang, T.; Head-Gordon, T. Computational optimization of electric fields for improving catalysis of a designed Kemp eliminase. *ACS Catal.* **2018**, *8*, 219−227.

(17) Świderek, K.; Tunon, I.; Moliner, V.; Bertran, J. Computational strategies for the design of new enzymatic functions. *Arch. Biochem. Biophys.* **2015**, *582*, 68−79.

(18) Świderek, K.; Tunon, I.; Moliner, V.; Bertran, J. Protein flexibility and preorganization in the design of enzymes. The Kemp elimination catalyzed by HG3.17. *ACS Catal.* **2015**, *5*, 2587−2595.

(19) Alexandrova, A. N.; Rothlisberger, D.; Baker, D.; Jorgensen, W. L. Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J. Am. Chem. Soc.* **2008**, *130*, 15907−15915.

(20) Labas, A.; Szabo, E.; Mones, L.; Fuxreiter, M. Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 908−917.

(21) Frushicheva, M. P.; Cao, J.; Chu, Z.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16869−16874.

(22) Frushicheva, M. P.; Cao, J.; Warshel, A. Challenges and advances in validating enzyme design proposals: The case of Kemp eliminase catalysis. *Biochemistry* **2011**, *50*, 3849−3858.

(23) Jindal, G.; Ramachandran, B.; Bora, R. P.; Warshel, A. Exploring the development of ground-state destabilization and transition-state stabilization in two directed evolution paths of Kemp eliminases. *ACS Catal.* **2017**, *7*, 3301−3305.

(24) Sokalski, W. Theoretical model for exploration of catalytic activity of enzymes and design of new catalysts: CO_2 hydration reaction. *Int. J. Quantum Chem.* **1981**, *20*, 231−240.

(25) Sokalski, W. A. The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory - derivation of simplified models. *J. Mol. Catal.* **1985**, *30*, 395−410.

(26) Szefczyk, B.; Mulholland, A. J.; Ranaghan, K. E.; Sokalski, W. A. Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J. Am. Chem. Soc.* **2004**, *126*, 16148−16159.

(27) Dittner, M.; Hartke, B. Globally optimal catalytic fields - inverse design of abstract embeddings for maximum reaction rate acceleration. *J. Chem. Theory Comput.* **2018**, *14*, 3547−3564.

(28) Dittner, M.; Hartke, B. Globally optimal catalytic fields for a DielsAlder reaction. *J. Chem. Phys.* **2020**, *152*, 114106.

(29) Freeze, J.; Kelly, H. R.; Batista, V. S. Search for catalysts by inverse design: Artificial intelligence, mountain climbers, and alchemists. *Chem. Rev.* **2019**, *119*, 6595−6612.

(30) Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014**, *346*, 1510−1514.

(31) Bhowmick, A.; Sharma, S. C.; Head-Gordon, T. The importance of the scaffold for de novo enzymes: A case study with Kemp eliminase. *J. Am. Chem. Soc.* **2017**, *139*, 5793−5800.

(32) Chojnacka, M.; Feliks, M.; Beker, W.; Sokalski, W. Predicting substituent effects on activation energy changes by static catalytic fields. *J. Mol. Model.* **2018**, *24*, 28.

(33) Carter, P.; Wells, J. Engineering enzyme specificity by substrate-assisted catalysis. *Science* **1987**, *237*, 394−399.

(34) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 261−268.

(35) Gaudreault, F.; Chartier, M.; Najmanovich, R. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics* **2012**, *28*, i423−i430.

(36) Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T. The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19386.

(37) Kulkarni, Y. S.; Amyes, T. L.; Richard, J. P.; Kamerlin, S. C. Uncovering the role of key active-site side chains in catalysis: An extended Brønsted relationship for substrate deprotonation catalyzed by wild-type and variants of triosephosphate isomerase. *J. Am. Chem. Soc.* **2019**, *141*, 16139−16150.

(38) Dziekoński, P.; Sokalski, W. A.; Kassab, E.; Allavena, M. Electrostatic nature of catalytic effects resulting from Si > Al substitutions in ZSM-5 zeolite. *Chem. Phys. Lett.* **1998**, *288*, 538−544.

(39) Cwiok, T.; Jeziorski, B.; Kolos, W.; Moszyński, R.; Szalewicz, K. Symmetry-adapted perturbation-theory of potential-energy surfaces for

weakly-bound molecular-complexes. *J. Mol. Struct.: THEOCHEM* **1994**, *307*, 135−151.

(40) Sokalski, W. A.; Roszak, S.; Hariharan, P. C.; Kaufman, J. J. Improved SCF interaction energy decomposition scheme corrected for basis set superposition effect. *Int. J. Quantum Chem.* **1983**, *23*, 847−854.

(41) Szarek, P.; Dyguda-Kazimierowicz, E.; Tachibana, A.; Sokalski, W. A. Physical nature of intermolecular interactions within cAMP-dependent protein kinase active site: Differential transition state stabilization in phosphoryl transfer reaction. *J. Phys. Chem. B* **2008**, *112*, 11819−11826.

(42) Kędzierski, P.; Sokalski, W. A.; Krauss, M. Nonempirical analysis of nature of catalytic effects in ribonuclease A active site. *J. Comput. Chem.* **2000**, *21*, 432−445.

(43) Langner, K. M.; Beker, W. L.; Dyguda-Kazimierowicz, E.; Sokalski, W. A. Tracking molecular charge distribution along reaction paths with atomic multipole moments. *Struct. Chem.* **2016**, *27*, 429−438.

(44) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *106*, 3210−3225.

(45) Beker, W.; van der Kamp, M.; Mulholland, A. J.; Sokalski, W. A. Rapid estimation of catalytic efficiency by cumulative atomic multipole moments: application to ketosteroid isomerase mutants. *J. Chem. Theory Comput.* **2017**, *13*, 945−955.

(46) Sokalski, W. A.; Poirier, R. A. Cumulative atomic multipole representation of the molecular charge-distribution and its basis set dependence. *Chem. Phys. Lett.* **1983**, *98*, 86−92.

(47) Sokalski, W. A. Nonempirical modeling of the static and dynamic properties of the opitimum nvironment for chemical reactions. *J. Mol. Struct.: THEOCHEM* **1986**, *138*, 77−87.

(48) Kędzierski, P.; Zaczkowska, M.; Sokalski, W. A.Extreme Catalytic Power of Ketosteroid Isomerase Related to Reversal of Proton Dislocations in Hydrogen Bond Network. *J. Phys. Chem. B*, **2020**, DOI: 10.1021/acs.jpcb.0c01489.

(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, G.; Scalmani, J. R.; Barone, G. A.; Petersson, V.; Nakatsuji, X.; Li, H.; Caricato, M.; Marenich, A. et al. *Gaussian 09, Revision A.02*; Gaussian, Inc.: Wallingford, CT, 2016.

(50) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461−462*, 1−21.

(51) Kumari, R.; Kumar, R.; Lynn, A. g_mmpbsa - A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951−1962.

(52) Berendsen, H. J. C.; Van Der Spoel, D.; Van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(53) Besler, B.; Merz, K.; Kollman, P. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(54) Scouras, A. D.; Daggett, V. The dynameomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Sci.* **2011**, *20*, 341−352.

(55) Schmidt, M.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; et al. General atomic and molecular electronic-structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(56) Aguilar, B.; Shadrach, R.; Onufriev, A. V. Reducing the secondary structure bias in the generalized Born model via R6 effective radii. *J. Chem. Theory Comput.* **2010**, *6*, 3613−3630.

(57) Aguilar, B.; Onufriev, A. V. Efficient computation of the total solvation energy of small molecules via the R6 generalized Born model. *J. Chem. Theory Comput.* **2012**, *8*, 2404−2411.

(58) Goldstein, R. Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J.* **1994**, *66*, 1335−1340.

(59) Koehl, P.; Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **1994**, *239*, 249−275.

(60) Gasteiger, J.; Li, X.; Rudolph, C.; Sadowski, J.; Zupan, J. Representation of molecular electrostatic potentials by topological feature maps. *J. Am. Chem. Soc.* **1994**, *116*, 4608−4620.

(61) Schwans, J. P.; Hanoian, P.; Lengerich, B. J.; Sunden, F.; Gonzalez, A.; Tsai, Y.; Hammes-Schiffer, S.; Herschlag, D. Experimental and computational mutagenesis to investigate the positioning of a general base within an enzyme active site. *Biochemistry* **2014**, *53*, 2541−2555.

(62) Mukhopadhyay, A.; Aguilar, B. H.; Tolokh, I. S.; Onufriev, A. V. Introducing charge hydration asymmetry into the generalized Born model. *J. Chem. Theory Comput.* **2014**, *10*, 1788−1794.

(63) Beker, W.; Langner, K. M.; Dyguda-Kazimierowicz, E.; Feliks, M.; Sokalski, W. A. Low-Cost prediction of relative stabilities of hydrogen bonded complexes from atomic multipole moments for overly short intermolecular distances. *J. Comput. Chem.* **2013**, *34*, 1797−1799.

(64) Kramer, C.; Spinn, A.; Liedl, K. R. Charge anisotropy: Where atomic multipoles matter most. *J. Chem. Theory Comput.* **2014**, *10*, 4488−4496.

(65) Konieczny, J.; Sokalski, W. Universal short-range ab initio atom-atom potentials for interaction energy contributions with an optimal repulsion functional form. *J. Mol. Model.* **2015**, *21*, 197.

(66) Podeszwa, R.; Pernal, K.; Patkowski, K.; Szalewicz, K. Extension of the Hartree-Fock plus dispersion method by first-order correlation effects. *J. Phys. Chem. Lett.* **2010**, *1*, 550−555.

(67) Wijma, H. J.; Marrink, S. J.; Janssen, D. B. Computationally efficient and accurate enantioselectivity modeling by clusters of molecular dynamics simulations. *J. Chem. Inf. Model.* **2014**, *54*, 2079−2092.