

## Supporting Information for

### *Substrate Positioning Dynamics Involves a Non-Electrostatic Component to Mediate Catalysis*

Yaoyukun Jiang,<sup>1#</sup> Ning Ding,<sup>1#</sup> Qianzhen Shao,<sup>1#</sup> Sebastian L. Stull,<sup>1</sup> Zihao Cheng,<sup>1</sup> and  
Zhongyue J. Yang<sup>1-5,\*</sup>

<sup>1</sup>*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States*

<sup>2</sup>*Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States*

<sup>3</sup>*Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37235,*

*United States* <sup>4</sup>*Data Science Institute, Vanderbilt University, Nashville, Tennessee 37235, United*

*States,* <sup>5</sup>*Department of Computer Science, Vanderbilt University, Nashville, Tennessee 37235*

#Y. Jiang, N. Ding, and Q. Shao contributed equally.

#### Contents

<b>Text S1.</b> Detailed computational and experimental methods.	Page S3
<b>Table S1.</b> Folding free energy change upon mutation for 80 randomly generated single mutants of KE07-R7-2.	Page S7
<b>Figure S1.</b> Constraints applied in the MD simulation.	Page S8
<b>Table S2.</b> Values of molecular dynamics-derived descriptors for the 15 variants.	Page S8
<b>Figure S2.</b> Distribution of the snapshot electric field along the substrate breaking C–H bond for 14 KE07-R7-2 variants examined in this work.	Page S9
<b>Table S3.</b> Primer sequences used in this work.	Page S9
<b>Figure S3.</b> Sanger sequencing chromatograms depicting site-directed mutagenesis in KE07-R7-2 variants.	Page S11
<b>Figure S4.</b> SDS-PAGE analysis of the purified KE07-R7-2 variants.	Page S12
<b>Figure S5.</b> Scatter plots for the efficiency-enhancing free energy barrier changes upon mutation versus the substrate positioning index calculated using solvent-exclusive surface area.	Page S12
<b>Table S4.</b> Experimentally characterized kinetic parameters of the purified KE07-R7-2 variants.	Page S13
<b>Table S5.</b> Values of computed kinetic parameters for the 15 variants.	Page S13
<b>Text S2.</b> Decomposition of the change of transition state stabilization free energy upon mutation into $k_{\text{cat}}$ and $K_{\text{M}}$ components.	Page S14
<b>Figure S6.</b> Correlation between the $k_{\text{cat}}$ and $K_{\text{M}}$ components of the effective activation free energy change upon mutation versus the substrate positioning index.	Page S15
<b>Table S6.</b> Solvent-accessible surface area decomposition of WT, N247W, and R154W mutants.	Page S15

**Figure S7.** The molecular mechanism underlying the impact of N247W and R154W mutations on enzyme kinetics. Page S16

**Figure S8.** Scatter plots for the correlation between the change of activation free energy versus the root-mean-square deviation from the idealized transition state. Page S17

## Text S1. Detailed computational and experimental methods.

*EnzyHTP-based High-throughput modeling.* The high-throughput enzyme modeling package EnzyHTP is used to construct the high-throughput workflow in this work based on a Python main script. The workflow consists of 6 steps: protein structure preparation, mutant generation, folding stability calculation, molecular dynamics simulation, quantum mechanics calculations, and post-analysis. The detailed operations in these steps in the script are described in the separate sections below. A tutorial on building workflows like this using EnzyHTP can be found on our documentation website: [https://enzyhtp-doc.readthedocs.io/en/latest/qkst\\_general.html](https://enzyhtp-doc.readthedocs.io/en/latest/qkst_general.html). The tutorial also contains details of each involved function.

*Protein Structure Preparation.* The Protein Data Bank (PDB ID: 5D38)<sup>1</sup> provided the crystal structure of KE07-R7-2 (defined as the WT), from which all the crystallizing water molecules were subsequently removed. To match the original KE07 design,<sup>2</sup> the N-terminal alanine was changed manually to methionine, and the residues beyond Leu253 on the C-terminal were deleted. PyMOL<sup>3</sup> is used to align the substrate-bound KE07 design model<sup>2</sup> with the apo KE07-R7-2 crystal structure. This alignment was used to construct the WT-substrate complex, which was then used as the input structure of EnzyHTP. EnzyHTP prepares the structure for MD simulations. *rm\_wat()* is used to remove irrelevant water molecules, *get\_protonation()* is used to add the missing hydrogens and determine the protonation state of titratable residues in the enzyme (powered by the PDB2PQR<sup>4</sup> and PROPKA3<sup>5</sup> interface in EnzyHTP), *PDB2FF()* is used to solvate and parameterize the enzyme for MD (powered by the AMBER<sup>6</sup> *tleap* interface in EnzyHTP). In detail, the AMBER ff14SB force field is used for the protein part,<sup>7</sup> and the generalized AMBER force field<sup>8,9</sup> is used for the substrate, H5J (5-nitro-1,2-benzoxazole), using AM1-BCC model to determine the atomic charges.<sup>10</sup>

*Random Mutation Generation.* The *add\_MutaFlag()* function in EnzyHTP is used to generate 80 random mutations. The position of the mutation and the residue after the mutation are both randomly determined. The structure of the mutation is determined using the *PDB2PDBwLeap()* function<sup>11</sup> in EnzyHTP.

*Folding Stability Calculations.* Rosetta *cartesian\_ddg* application<sup>12, 13</sup> is interfaced in EnzyHTP to evaluate the folding free energy change for each of the 80 random single mutants. *get\_rosetta\_ddg()* is used to automate the process. In detail, the input structure was the crystal structure of the apoenzyme prepared in the previous step. Before performing the folding free energy calculations, the workflow conducted 10 iterations of conformational sampling using *relax\_with\_rosetta()*. The calculation was performed using a Cartesian coordinate representation of the protein structure, and the pairwise interactions were evaluated by the full-atom energy function within 9.0 Å. The workflow employed the *talaris2014\_cart* scoring function in the calculation. The free energy minimum was obtained from the 10 iterations, and the difference between the minima of the mutant and WT was used to calculate the folding free energy change upon mutation.

*Molecular Dynamics Simulations.* For the 61 random mutants (Table S1) that passed the folding stability test, the workflow performs MD simulations of the enzyme-substrate complexes using *PDBMD()* (powered by the AMBER<sup>6</sup> interface in EnzyHTP, a local install of AMBER 18 is used in this workflow specifically) The structures of enzyme mutants generated as described above are used as the input structure of each MD simulation. The SHAKE algorithm<sup>14</sup> was used to constrain all hydrogen-containing bonds. To sample near transition state conformations throughout

the simulations, the workflow applied geometric restraints between the substrate and key amino acid residues from minimization to production runs (Fig. S1). The enzyme complexes were then solvated in a periodic truncated octahedron box with a 10 Å buffer of TIP3P water and were neutralized with Na<sup>+</sup> counterions. For each mutant complex, the workflow first relaxed the whole solvent box using the steepest descent method for 10000 steps followed by the conjugate gradient method for another 10000 steps. After minimization, the workflow heated each box from 0 to 293.15 K within 36 ps with constant volume, equilibrated it for 4 ps under constant volume at 293.15 K, and further equilibrated at 293.15 K and 1 atm for 1 ns. In addition, the workflow restrained the backbone C<sub>α</sub>, C, and N of the amide group with a weight of 2 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> from minimization to equilibration. After equilibration, the workflow carried out production runs for 110 ns and output the trajectories every 100 ps. The snapshots from the last 100 ns of the production run were used for analyses, resulting in a total of 1000 snapshots for each production run. All simulations were performed with a time step of 2 fs. The Langevin thermostat<sup>15</sup> and Berendsen barostat<sup>16</sup> were used to control the temperature and pressure, respectively.

Apart from the variants mentioned above, we adopted the MD simulations for the WT and other 17 mutants originally reported in a previous benchmark work.<sup>17</sup> The 17 mutants are single mutants whose experimental kinetic data had been reported.<sup>18</sup> The 17 mutants are S48N, H201A, H201K, K222A, R16Q, N25S, I52A, M62A, H84Y, K132N, I199S, I199F, I199A, K132M, K162A, L170A, and E185A. The setups were the same as described in this section except that five parallel simulations were conducted, resulting in 5000 snapshots for these 18 variants.

*Quantum Mechanics Calculations.* To acquire the dipole moment of the breaking C–H bond, we conducted single-point electronic structure calculations using TeraChem<sup>19, 20</sup> for 500 snapshots sampled from WT MD trajectories with a 1 ns interval. Each snapshot was converted to an image in which the enzyme-substrate complex occupied the center of the box using the *autoimage* function of the AMBER *cpptraj* utility.<sup>21</sup> For each MD snapshot, the substrate and 14 surrounding amino acid residues were selected as the quantum mechanics (QM) cluster. The selected residues included Ala9, Leu10, Ile11, Ser48, Phe49, Trp50, Glu101, Asn103, Tyr128, Val169, His201, Arg202, Lys222, and Asp224. Our previous benchmark has shown that the ranking of EF contributions from these residues correlated well with that of the activation free energy across 18 KE variants.<sup>17</sup> We assumed that the dipole moment based on the same residues was also able to efficiently converge. The QM cluster boundaries cut the backbone amide C–N bond. To cap the unbonded atoms, explicit H atoms were placed along the dangling C–N bond vector, and the resulting N–H and C–H bond lengths were set to be 1.09 Å. The electronic structures were described using the range-separated exchange-correlation functional ωPBEh<sup>22</sup> (ω = 0.2 bohr<sup>-1</sup>) with 6-31G(d).<sup>23</sup> This combination of method and basis set has been validated in the study of large-scale electronic structure effects in catechol *O*-methyltransferase, cytochrome P450cam, lysozyme, DNA methyltransferase, and Kemp Eliminate.<sup>17, 24, 25</sup> The restrained electrostatic potential (RESP) point charges<sup>26</sup> of each snapshot were calculated for the following dipole analyses.

*Post Analyses.* The workflow calculated the substrate positioning index (SPI), the electric field of the breaking C–H bond (EF<sub>C–H</sub>, in MV/cm), and root-mean-square deviation from the idealized active site (RMSD<sub>ITS</sub>, in Å) as the average value from sampled MD snapshots. The average of EF<sub>C–H</sub> for each enzyme variant were then used to calculate the electric field change of the breaking C–H bond upon mutation (ΔEF<sub>C–H</sub>, in MV/cm) and electrostatic stabilization energy (ESE) change of the breaking C–H bond upon mutation (ΔG<sub>ele</sub>, in kcal/mol). The average values are summarized in Table S2.

SPI is the ratio of substrate SASA over the enzyme pocket SASA. For substrate SASA, all atoms were included in the calculation. The enzyme pocket was defined using Ala9, Ile11, Ser48, Trp50, Glu101, Tyr128, His201, Arg202, and Lys222. Our previous benchmark work on KE07-R7-2 has shown that including these residues can balance the accuracy and efficiency in SPI calculations.<sup>17</sup> SASA was calculated using the Shrake and Rupley algorithm<sup>27</sup> embedded in the Python library MDTraj.<sup>28</sup> The probe radius was 1.4 Å and the surface of each atom was represented by 5000 grid points.  $EF_{C-H}$  was the projected (C to H as the positive direction) electric field strength along the breaking C–H bond of the substrate. Within each MD snapshot, the  $EF_{C-H}$  was summed at the middle point of the breaking C–H bond from all atoms of the protein based on the RESP charges used in the classical force field. Then the relative electric field change,  $\Delta EF_{C-H}$ , was calculated as the difference of the average  $EF_{C-H}$  from all MD snapshots between a mutant and the WT:  $\Delta EF_{C-H} = \langle EF_{C-H}^{\text{mutant}} \rangle - \langle EF_{C-H}^{\text{WT}} \rangle$ . We universally used the WT dipole moment ( $p_{C-H}^{\text{WT}}$ ) of 0.285 a.u. to estimate the ESE for all the enzyme variants. This value was the average from 500 QM cluster calculations. In each QM cluster, the C–H bond dipole projection (C to H as the positive direction) along this bond was calculated based on the wave functions of the molecular orbitals using the wavefunction analyzer Multiwfn.<sup>29</sup> ESE change ( $\Delta G_{\text{ele}}$ ) of the breaking C–H bond was estimated as the negative product of the electric field change along this bond ( $\Delta EF_{C-H}$ ) and the WT dipole moment of this bond ( $p_{C-H}^{\text{WT}}$ ):  $\Delta G_{\text{ele}} = -\Delta EF_{C-H} \cdot p_{C-H}^{\text{WT}}$ . For RMSD<sub>ITS</sub>, we included all the heavy atoms of the substrate and C $_{\alpha}$  and side chain of the amino acid residues Trp50, Glu101, and Lys222. The rationally designed model reported by Röthlisberger et al. was used as the reference structure.<sup>2</sup> The mass-weighted RMSD was calculated using the *cpptraj* utility of AMBER.<sup>21</sup>

*Kernel Density Estimation.* The *kdeplot* function from the *Seaborn* python package<sup>30</sup> was used to perform the Kernel Density Estimation to generate the graphic. The default kernel (Gaussian) and band width selection method (Silverman's rule of thumb method) were applied.

*Standard error of the free energy difference.* The free energy difference is calculated as  $\Delta\Delta G^{\ddagger} = -RT \ln \frac{AVE(\text{mut})}{AVE(\text{WT})}$ . In this equation,  $R$  and  $T$  are the gas constant and temperature, respectively, and  $RT$  is taken to be 0.593 kcal/mol at 298 K.  $AVE()$  refers to the average value of the experimentally acquired kinetic parameters of  $k_{\text{cat}}$ ,  $K_M$ , or  $k_{\text{cat}}/K_M$ . The corresponding variance of the kinetic parameter is defined as  $VAR()$ . The variance of *ratio* =  $\frac{AVE(\text{mut})}{AVE(\text{WT})}$  is calculated as  $VAR(\text{ratio}) = \frac{AVE(\text{mut})^2 \cdot VAR(\text{WT}) + AVE(\text{WT})^2 \cdot VAR(\text{mut})}{AVE(\text{WT})^4}$ . The variance of *logarithm* =  $\ln(\text{ratio})$  is then calculated as  $VAR(\text{logarithm}) = \frac{VAR(\text{ratio})}{\text{ratio}^2}$ . The standard error of free energy difference is then calculated as  $RT \sqrt{\frac{VAR(\text{logarithm})}{n}}$  where  $n = 3$  is the sample size.

*Bacterial Strains, Plasmids, and Cloning.* KE07 was encoded by gene *ke07* with a GenBank accession of GQ414538.1. The gene encoding KE07-R7-2, a variant of KE07,<sup>2</sup> was derived from *ke07* and named as *ke07-r7-2*. *ke07-r7-2* was synthesized by Integrated DNA Technologies (Morrisville, NC) and amplified by polymerase chain reaction (PCR) using Q5 High-Fidelity Master Mix (New England Biolabs, Ipswich, MA) with primers listed in Table S3. Subsequently, the amplified *ke07-r7-2* gene and pET-29b(+) vector (Novagen, Madison, WI) were assembled by Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA). The resulting plasmid, *ke07-r7-2/pET-29b(+)*, was transformed into *Escherichia coli* DH5 $\alpha$ , extracted using the Monarch Plasmid Miniprep Kit (New England Biolabs, Ipswich, MA), and sequenced by Azenta Life Sciences (South Plainfield, NJ). To construct KE07-R7-2 variants, including WT, N247W,

K4M, R154W, D14F, E185A, and K37Q, PCR amplification was performed using *ke07-r7-2/pET-29b* as a template and primers containing the desired variants (Table S3). The amplified fragments were then cloned into the remaining backbone of *ke07-r7-2/pET-29b(+)* using Gibson Assembly Master Mix. The resulting plasmids were transformed into *E. coli* DH5 $\alpha$ , extracted using the Monarch Plasmid Miniprep Kit and sequenced. Finally, the sequenced plasmids were transformed into *E. coli* BL21 (DE3) for expression.

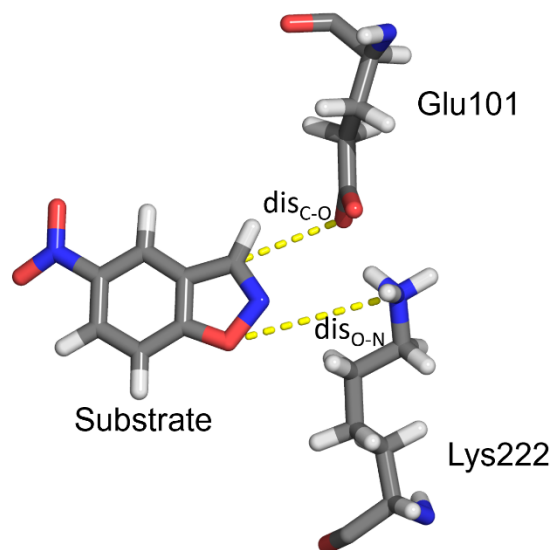
*Protein Expression and Purification.* A single colony from the transformed cells was inoculated into 20 mL of Luria-Bertani (LB) medium supplemented with kanamycin (50 mg mL<sup>-1</sup>) and grown at 37 °C overnight with agitation at 220 rpm. The resulting culture was then used to inoculate 1 L of kanamycin-supplemented LB medium and incubated at 37 °C with agitation at 220 rpm. Once the optical density of the culture at 600 nm reached 0.6, 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG, Fisher Scientific, Pittsburgh, PA) was added, and the culture was subsequently incubated at 16 °C with shaking at 160 rpm for 16 h to induce enzyme expression. The cells were harvested by centrifugation at 6,000  $\times$  g for 30 min at 4 °C.

To purify the enzymes, the cells were suspended in 25 mL of buffer A, which consisted of 25 mM HEPES (pH 7.5, Boston Bioproducts, Milford, MA), 100 mM NaCl, and 5 mM imidazole. The cells were then lysed through sonication on ice for 10 min. Cell debris was removed by centrifugation at 20,000  $\times$  g for 40 min at 4 °C. The supernatant was mixed with 5 mL of Ni-NTA resin (Invitrogen, Waltham, MA) and incubated under rotation overnight at 4 °C. The resin was then washed with buffer B containing 25 mM HEPES (pH 7.5), 100 mM NaCl, and 20 mM imidazole to eliminate non-target proteins. The target proteins were eluted with buffer C consisting of 25 mM HEPES (pH 7.5), 100 mM NaCl, and 250 mM imidazole. The purified proteins were dialyzed in buffer D, which contained 25 mM HEPES (pH 7.25, Teknova, Hollister, CA), 100 mM NaCl and 5% glycerol at 4 °C for 24 h. The purity of the proteins was determined using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

*Kinetic Analyses.* To determine the kinetic parameters ( $V_{\max}$ ,  $k_{\text{cat}}$ ,  $K_m$ , and  $k_{\text{cat}}/K_m$ ) of KE07-R7-2 variants, their activities were assayed using 5-nitro-1,2-benzoxazole as substrate. The substrate with concentrations ranging from 5 to 1500  $\mu$ M was prepared by diluting a 100 mM stock in acetonitrile. The reactions were performed at 25 °C in 25 mM HEPES (pH 7.25), 100 mM NaCl, 5% glycerol, and 1.5% (v/v) acetonitrile. Notably, to avoid distortion in the curvature of the Michaelis plots due to varying acetonitrile concentrations, the volume of acetonitrile was adjusted to maintain a consistent final concentration of 1.5% (v/v) in all reaction mixtures.<sup>31</sup> The reactions were initiated by adding 50  $\mu$ L of the enzyme (8  $\mu$ M final concentration) to 150  $\mu$ L of the substrate in a 96-well plate (Corning-Costar, Corning, NY). The formation of the product was monitored at 380 nm using a SpectraMax iD3 microplate reader (Molecular Device, San Jose, CA). An extinction coefficient of 15,800 M<sup>-1</sup> cm<sup>-1</sup> was used to calculate reaction rates.  $V_{\max}$  and  $K_m$  were determined by nonlinear regression with the Michaelis-Menten model using GraphPad Prism software version 8.  $k_{\text{cat}}/K_m$  was calculated to determine the catalytic efficiency of the enzymes. Three biologically independent replicates were used to calculate means and standard errors. Statistical significance was analyzed using GraphPad Prism software version 8 by performing analysis of variance (ANOVA) and Tukey's *post hoc* multiple comparisons tests.

**Table S1.** Folding free energy change upon mutation (stability) for 80 randomly generated single mutants of KE07-R7-2. The stability is predicted by the *cartesian\_ddg* application of the Rosetta software suite. The unit of stability is the Rosetta energy unit (REU). The stability values of the enzyme variants are arranged in ascending order. The 5 mutants that are experimentally tested in this work are shown in bold and the 19 variants with stability values larger than 10 REU are shaded.

Index	Variant	Stability (REU)	Index	Variant	Stability (REU)	Index	Variant	Stability (REU)
1	S90V	-4.957	28	E57R	0.574	55	G217S	6.443
2	S229Q	-3.678	<b>29</b>	<b>K37Q</b>	<b>0.603</b>	56	V157Q	6.448
3	K4W	-3.554	30	D224F	0.647	57	A89M	6.995
4	A223V	-3.434	31	E46Y	0.754	58	M62S	8.089
5	K242C	-3.292	<b>32</b>	<b>N247W</b>	<b>0.977</b>	59	G205E	8.441
<b>6</b>	<b>K4M</b>	<b>-2.960</b>	33	G245E	1.022	60	G80N	9.326
7	E212C	-2.691	34	S144T	1.127	61	T194D	9.869
8	A9M	-2.191	35	D85F	1.428	62	V157D	10.443
9	E24C	-1.642	36	I42V	1.506	63	G30P	10.568
10	E24Y	-1.618	37	T114R	1.510	64	F120N	11.375
11	S229G	-1.469	38	K13I	1.715	65	L112G	11.609
12	H244K	-1.464	39	Y240M	1.738	66	L170R	11.899
13	R59V	-1.351	40	V18L	1.743	67	V234S	11.928
14	N109I	-1.255	41	E91R	1.972	68	L112D	12.216
15	E71H	-1.155	42	K4Q	2.285	69	V190W	12.581
16	K147F	-1.093	43	T194M	2.351	70	F102A	13.082
17	G15W	-0.894	44	T195D	2.354	71	I6G	14.200
18	S111T	-0.796	45	L152T	2.549	72	V160D	14.419
19	S180H	-0.466	46	R175E	3.603	73	L237W	15.490
20	E87Y	-0.377	47	Y39M	3.832	74	A221P	16.773
21	N109D	-0.347	48	R133C	4.152	75	D45P	18.094
22	F38L	-0.344	<b>49</b>	<b>R154W</b>	<b>4.301</b>	76	G36K	20.679
23	T114W	0.004	50	R5Q	4.358	77	F49P	20.730
24	K99F	0.067	51	A223M	4.688	78	A97K	20.954
25	L193I	0.268	52	I11E	5.284	79	L241R	21.003
<b>26</b>	<b>D14F</b>	<b>0.273</b>	53	D174E	5.590	80	K238P	23.081
27	E185T	0.279	54	I75Y	5.954			



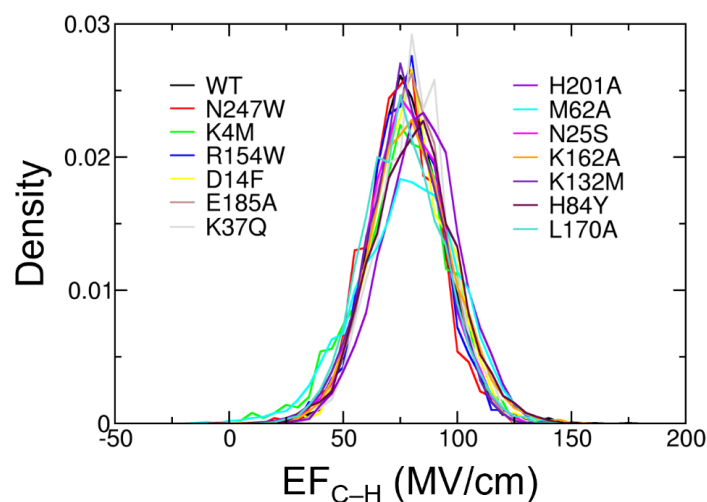
**Figure S1.** Constraints applied in the MD simulation. The cutoff distance between Glu101 carbonyl O and substrate deprotonated C,  $dis_{C-O}$ , is 2.93 Å. The cutoff distance between Lys222 N $\epsilon$  and substrate O in the ring,  $dis_{O-N}$ , is 4.00 Å. A weight of 100 kcal mol<sup>-1</sup> Å<sup>-2</sup> is applied if the distance is larger than the corresponding cutoff.

**Table S2.** Values of molecular dynamics-derived descriptors for the 15 variants investigated in this work. The descriptors include substrate positioning index (SPI), electric field along the breaking C–H bond ( $EF_{C-H}$ ), electric field difference along the breaking C–H bond upon mutation ( $\Delta EF_{C-H}$ ), electrostatic stabilization energy difference upon mutation ( $\Delta G_{ele}$ ), and root-mean-square deviation from the idealized transition state ( $RMSD_{ITS}$ ). SPI,  $EF_{C-H}$ , and  $RMSD_{ITS}$  are averages from the corresponding MD snapshots. Those values are the means  $\pm$  standard deviations. Averages with different superscript letters within the same column are significantly different. The variants that do not satisfy the electric field selection criteria ( $\Delta EF_{C-H}$  between  $\pm 2.88$  MV/cm) are shaded. Mutants from random mutations (N247W, K4M, R154W, D14F, and K37Q) are averaged from 1000 snapshots, and other variants including the WT are averaged from 5000 snapshots.

Index	Variant	SPI	$EF_{C-H}$ (MV/cm)	$\Delta EF_{C-H}$ (MV/cm)	$\Delta G_{ele}$ (kcal/mol)	$RMSD_{ITS}$ (Å)
1	WT	1.42 $\pm$ 0.18	75.75 $\pm$ 16.38	0	0	0.94 $\pm$ 0.16
2	N247W	1.29 $\pm$ 0.08	73.26 $\pm$ 15.87	-2.49	0.086	0.98 $\pm$ 0.15
3	K4M	1.43 $\pm$ 0.15	73.00 $\pm$ 19.84	-2.75	0.095	0.94 $\pm$ 0.14
4	R154W	1.56 $\pm$ 0.23	74.79 $\pm$ 15.76	-0.96	0.033	0.89 $\pm$ 0.15
5	D14F	1.61 $\pm$ 0.15	76.85 $\pm$ 16.95	1.11	-0.038	1.03 $\pm$ 0.14
6	E185A	1.65 $\pm$ 0.20	76.31 $\pm$ 15.59	0.56	-0.020	0.95 $\pm$ 0.15
7	K37Q	1.68 $\pm$ 0.18	78.28 $\pm$ 15.57	2.54	-0.088	0.98 $\pm$ 0.16
8	S48N	1.23 $\pm$ 0.12	81.42 $\pm$ 17.53	5.68	-0.197	1.14 $\pm$ 0.21
9	H201A	1.17 $\pm$ 0.16	75.45 $\pm$ 22.18	-0.30	0.010	1.14 $\pm$ 0.21
10	M62A	1.30 $\pm$ 0.14	75.18 $\pm$ 16.86	-0.56	0.020	1.01 $\pm$ 0.19
11	N25S	1.35 $\pm$ 0.15	77.94 $\pm$ 19.62	2.19	-0.076	0.93 $\pm$ 0.18
12	K162A	1.38 $\pm$ 0.17	77.34 $\pm$ 18.39	1.59	-0.055	0.95 $\pm$ 0.21
13	K132M	1.40 $\pm$ 0.19	74.39 $\pm$ 16.84	-1.36	0.047	0.88 $\pm$ 0.14



14	H84Y	1.42±0.21	78.28±18.46	2.53	-0.088	0.95±0.15
15	L170A	1.68±0.36	73.78±17.60	-1.97	0.068	0.95±0.16



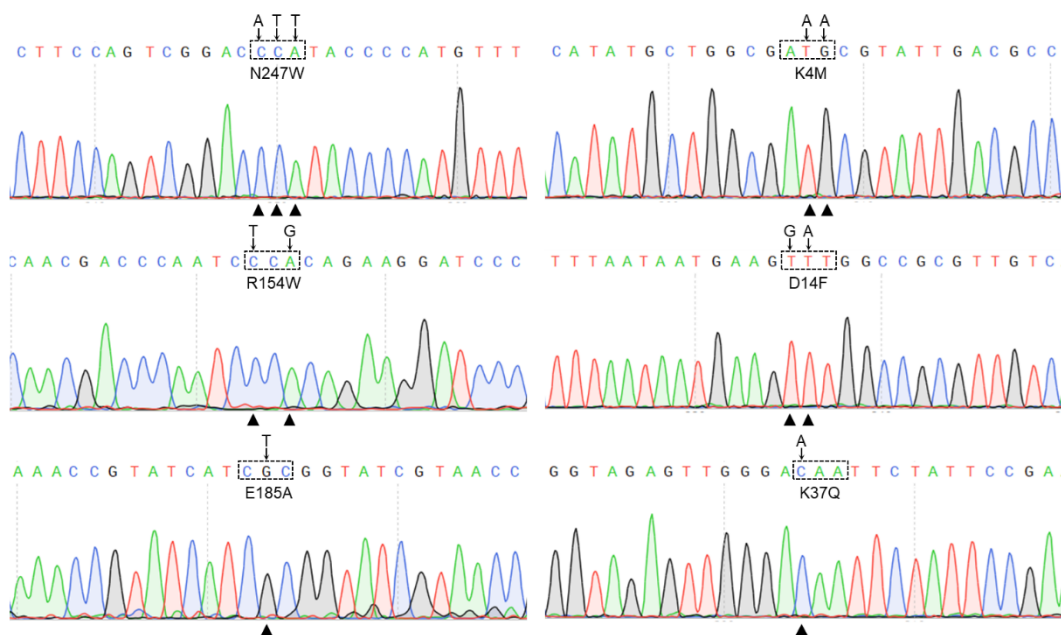
**Figure S2.** Distribution of the snapshot electric field along the substrate breaking C–H bond ( $EF_{C-H}$ ) for 14 KE07-R7-2 variants examined in this work. The bin size is 5 MV/cm.

**Table S3.** Primer sequences used in this work.

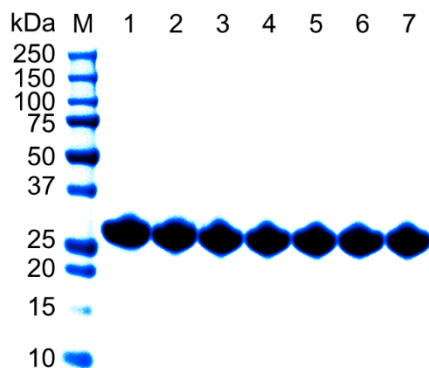
Construct	Primer	Template	Sequence 5'-3'	T <sub>m</sub> (°C)
	For_1	<i>ke07-r7-2</i>	ttaagaaggagatatacatATGCTGGCGAAAC GTATTGA	64
<i>ke07-r7-2/pET-29b</i>	Rev_1	<i>ke07-r7-2</i>	tctctcacatcgatttcgCGAGTGAAAAACC GAATCGG	64
	For_2	pET-29b	CGCGAAATCGATGTGAGAGA	60
	Rev_2	pET-29b	ATGTATATCTCCTTCTTAAAGTTAAA CAAA	60
	For_3	<i>ke07-r7-2/pET-29b</i>	gaaacatgggggatgggtccGACTGGAAGGCT TGGGTA	62
<i>n247w/pET-29b</i>	Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAAGAA CGTGGACTCC	62
	For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCTTAAAGTTAAA CAAA	62
	Rev_4	<i>ke07-r7-2/pET-29b</i>	ggaccataCCCATGTTTCTTAAATAC TCTTTG	62

	For_5	<i>ke07-r7-2/pET-29b</i>	atatacatatgctggcgatgCGTATTGACGCCG CATTAAAT	62
<i>k4m/pET-29b</i>	Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAAGAA CGTGGACTCC	62
	For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCAAACACTGGAAC	62
	Rev_6	<i>ke07-r7-2/pET-29b</i>	catCGCCAGCATATGTATATCTCCT	62
	For_7	<i>ke07-r7-2/pET-29b</i>	tggGATTGGGTCGTTGAGGTAGA	62
<i>r154w/pE T-29b</i>	Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAAGAA CGTGGACTCC	62
	For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCAAACACTGGAAC	62
	Rev_8	<i>ke07-r7-2/pET-29b</i>	acctcaacgaccaatcccaCAGAAGGATCCC GGTGTT	62
	For_9	<i>ke07-r7-2/pET-29b</i>	attaataatgaagtttggccGCGTTGTCAAAGG TAGCAAT	62
<i>d14f/pET-29b</i>	Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAAGAA CGTGGACTCC	62
	For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCAAACACTGGAAC	62
	Rev_10	<i>ke07-r7-2/pET-29b</i>	ggccaaacttcATTATTAATGCGGCGTCA ATACG	62
	For_11	<i>ke07-r7-2/pET-29b</i>	gcgATGATACGGTTTGTCCGTCC	62
<i>e185a/pE T-29b</i>	Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAAGAA CGTGGACTCC	62
	For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCAAACACTGGAAC	62

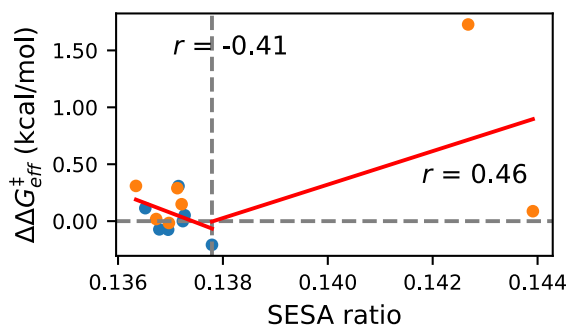
Rev_12	<i>ke07-r7-2/pET-29b</i>	cggacaaaccgtatcatcgcGGTATCGTAACC GGATTTCG	62
For_13	<i>ke07-r7-2/pET-29b</i>	caaTTCTATTCCGAAATTGGCATCG	62
Rev_3	<i>ke07-r7-2/pET-29b</i>	tccagtttgaacaagagtcCACTATTAAGAA CGTGGACTCC	62
For_4	<i>ke07-r7-2/pET-29b</i>	GACTCTTGTTCCAAACCTGGAAC	62
Rev_14	<i>ke07-r7-2/pET-29b</i>	ccaatttcggaatagaattgTCCCAACTCTACC GGATC	62



**Figure S3.** Sanger sequencing chromatograms depicting site-directed mutagenesis in KE07-R7-2 variants. DNA sequences carrying mutations K4M, K14F, and K37Q are read in 5' to 3' (forward) direction, while those carrying N247W, R154W, and E185A are in 3' to 5' (reverse) direction.



**Figure S4.** SDS-PAGE analysis of the purified KE07-R7-2 variants. M, protein molecular weight marker; lane 1, KE07-R7-2; lane 2, N247W; lane 3, K4M; lane 4, R154W; lane 5, D14F; lane 6, E185A; lane 7, K37Q.



**Figure S5.** Scatter plots for the efficiency-enhancing free energy barrier changes upon mutation,  $\Delta\Delta G_{\text{eff}}^{\ddagger}$ , vs. the substrate positioning index calculated using solvent-exclusive surface area (SESA) in place of solvent-accessible surface area (SASA). The plot contains 14 KE07-R7-2 variants shown in Fig. 4b except S48N. The horizontal dashed line indicates the position of  $\Delta\Delta G_{\text{eff}}^{\ddagger} = 0$  where catalysis efficiency does not change; The vertical dashed line indicates the position of SESA ratio = 0.138 where the most beneficial mutant, i.e., R154W locates. The vertical dashed line is also the boundary of the piecewise linear fitting. The fitting lines are shown in red and labeled with the corresponding Pearson correlation coefficient ( $r$ ). The data of R154W is included in both fitting lines. The SPI values calculated using SESA, i.e., SESA ratio, showed significantly less variation, indicating that it was poor to reflect changes in substrate positioning upon mutation. Although both methods consider all the atoms of the selected residue, they also include surface areas that are far from the binding pocket. In this regard, SASA contains a smaller portion of substrate-positioning-irrelevant surface areas than SESA. Consequently, SASA exhibited better descriptive performance than SESA.

**Table S4.** Experimentally characterized kinetic parameters of the purified KE07-R7-2 variants. The measurements were recorded in 20 mM HEPES at pH 7.25 containing 100 mM NaCl, and 5% glycerol at 25°C. The substrate (5-nitroxybenzoxazole) was dissolved in acetonitrile and the enzymatic assay contained a final concentration of acetonitrile at 1.5% (v/v).  $K_M$  and  $V_{max}$  were determined by nonlinear regression with the Michaelis-Menten model using Prism 8. The superscripts of the kinetic data are the results of Tukey's *post hoc* multiple comparisons tests. Values are the means  $\pm$  standard errors ( $n = 3$ ). Averages with different superscript letters within the same column are significantly different.

KE07-R7-2 variant	$k_{cat}$ (s <sup>-1</sup> )	$K_M$ (mM)	$k_{cat}/K_M$ (M <sup>-1</sup> s <sup>-1</sup> )
WT	0.93 $\pm$ 0.07 <sup>ab</sup>	0.47 $\pm$ 0.04 <sup>a</sup>	1983.72 $\pm$ 33.25 <sup>bc</sup>
N247W	1.83 $\pm$ 0.01 <sup>d</sup>	1.14 $\pm$ 0.11 <sup>b</sup>	1636.91 $\pm$ 159.92 <sup>ab</sup>
K4M	1.35 $\pm$ 0.06 <sup>bc</sup>	0.60 $\pm$ 0.02 <sup>a</sup>	2241.25 $\pm$ 99.75 <sup>c</sup>
R154W	1.72 $\pm$ 0.17 <sup>cd</sup>	0.61 $\pm$ 0.05 <sup>a</sup>	2818.00 $\pm$ 54.11 <sup>d</sup>
D14F	0.98 $\pm$ 0.11 <sup>ab</sup>	0.43 $\pm$ 0.03 <sup>a</sup>	2252.38 $\pm$ 98.68 <sup>c</sup>
E185A	0.83 $\pm$ 0.06 <sup>a</sup>	0.46 $\pm$ 0.01 <sup>a</sup>	1817.66 $\pm$ 155.06 <sup>bc</sup>
K37Q	0.66 $\pm$ 0.07 <sup>a</sup>	0.56 $\pm$ 0.03 <sup>a</sup>	1180.39 $\pm$ 50.00 <sup>a</sup>

**Table S5.** Values of computed kinetic parameters for the 15 variants investigated in this work. Computed kinetic parameters are based on the experimentally characterized data shown in Table S4 and previously reported ones.<sup>18</sup> These data include the percentages of  $k_{cat}$ ,  $K_M$ , and  $k_{cat}/K_M$  relative to the wildtype (%WT) and the corresponding contributions to the free energy changes upon mutation, i.e.,  $\Delta\Delta G_{k_{cat}}^\ddagger$ ,  $\Delta\Delta G_{K_M}^\ddagger$ , and  $\Delta\Delta G_{eff}^\ddagger$ .

Index	Variant	$k_{cat}$ (%WT)*	$\Delta\Delta G_{k_{cat}}^\ddagger$ (kcal/mol)	$K_M$ (%WT)*	$\Delta\Delta G_{K_M}^\ddagger$ (kcal/mol)	$k_{cat}/K_M$ (%WT)*	$\Delta\Delta G_{eff}^\ddagger$ (kcal/mol)
1	WT	100	0	100	0	100	0
2	N247W	196.77	-0.40	242.55	-0.53	82.52	0.11
3	K4M	145.16	-0.22	127.66	-0.14	112.98	-0.07
4	R154W	184.95	-0.36	129.79	-0.15	142.06	-0.21
5	D14F	105.38	-0.03	91.49	0.05	113.54	-0.08
6	E185A	89.25	0.07	97.87	0.01	91.63	0.05
7	K37Q	70.97	0.20	119.15	-0.10	59.50	0.31
8	S48N	12.35	1.24	169.46	-0.31	7.29	1.55
9	H201A	N/A	N/A	N/A	N/A	5.43	1.73
10	M62A	79.01	0.14	133.17	-0.17	59.35	0.31
11	N25S	71.60	0.20	117.69	-0.10	61.36	0.29
12	K162A	88.89	0.07	102.95	-0.02	86.33	0.09
13	K132M	88.89	0.07	86.49	0.09	102.76	-0.02
14	H84Y	95.06	0.03	122.11	-0.12	77.84	0.15
15	L170A	80.25	0.13	83.05	0.11	96.93	0.02

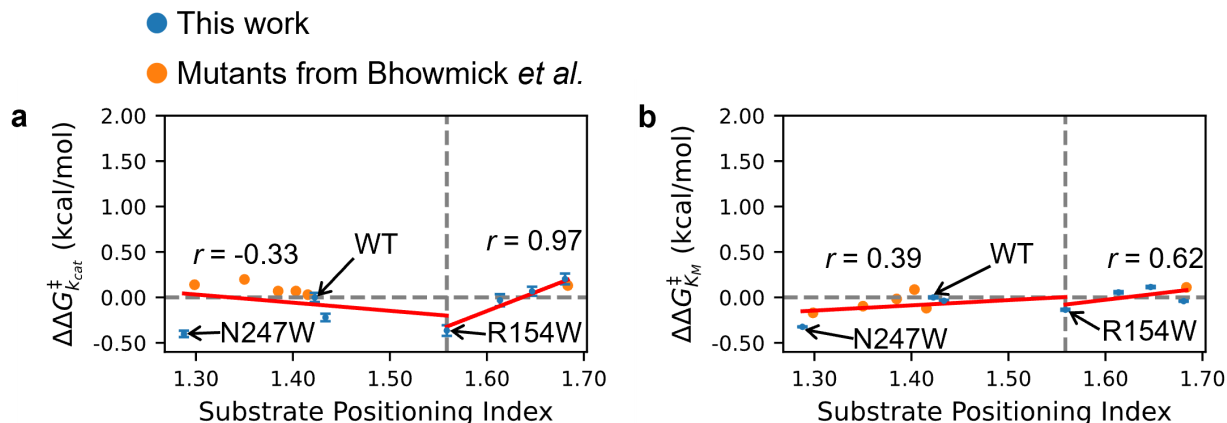
\*For the calculation of percentage compared to the WT, the denominator used the values measured in this work ( $k_{cat} = 0.93$  s<sup>-1</sup>,  $K_M = 0.47$  mM, and  $k_{cat}/K_M = 1983.72$  M<sup>-1</sup> s<sup>-1</sup>) for variants 1 to 7 and the result reported by Bhowmick *et al.*<sup>18</sup> ( $k_{cat} = 0.81$  s<sup>-1</sup>,  $K_M = 0.407$  mM, and  $k_{cat}/K_M = 1990$  M<sup>-1</sup> s<sup>-1</sup>) for variants 8 to 15.

**Text S2.** Decomposition of the change of transition state stabilization free energy upon mutation

$\Delta\Delta G_{\text{stab}}^{\text{TS}}$  (i.e.,  $\Delta\Delta G_{\text{eff}}^{\ddagger}$ ) into  $k_{\text{cat}}$  ( $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$ ) and  $K_{\text{M}}$  ( $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$ ) components.

*Correlations between  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$  and SPI and between  $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$  and SPI.* Pearson correlation coefficients between  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$  and SPI are -0.33 and 0.97 for  $\text{SPI} \leq 1.56$  and  $\geq 1.56$ , respectively (Figure S6a). N247W is the main reason of the low correlation under  $\text{SPI} \leq 1.56$ . The correlation is improved to -0.89 without N247W's  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$ .  $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$  has a poor correlation to SPI (Figure S6b). Compared to  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$ , the inclusion of  $K_{\text{M}}$  in the calculation of activation free energy difference, i.e.,  $\Delta\Delta G_{\text{eff}}^{\ddagger}$ , significantly improves the correlation in the range of  $\text{SPI} \leq 1.56$ : The Pearson correlation coefficient changes from -0.33 ( $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$  against SPI, Figure S6a) to -0.82 ( $\Delta\Delta G_{\text{stab}}^{\text{TS}}$  or  $\Delta\Delta G_{\text{eff}}^{\ddagger}$  against SPI excluding H201A, Figure 3a).

*Influences of mutation N247W on  $k_{\text{cat}}$  and  $K_{\text{M}}$ .* The inconsistent observation of the enhanced reactivity (negative  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$ ) and reduced binding (negative  $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$ ) of N247W is closely related to the configuration change of the active site. To identify the source of change, we decomposed the total pocket SASA to residue contributions. As shown in Table S6, 84% of the SASA change comes from Arg202. Further analysis of the N247W trajectory reveals a much more frequent side chain flip of Arg202 compared to the WT (Figure S7b *Top*). The flip causes a partial block of substrate's entering window (Figure S7a), weakening the binding affinity of N247W compared to the WT. As a result, the corresponding  $K_{\text{M}}$  increases a lot and the  $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$  has a large negative value. As a support, the calculated binding free energy of N247W using the MM/PBSA method is higher than the WT: The results of WT and N247W are -7.15 and -4.34 kcal/mol, respectively. However, once the substrate manages to squeeze into the pocket, the flipped guanidino group of Arg202 forms strong electrostatic interaction with the nitro group of the substrate, leading to a more rigid positioning of the active conformations. This contributes significantly to the increase of  $k_{\text{cat}}$  compared to the WT and the  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$  has a large negative value. The dual function of Arg202 in N247W provides a representative case where the SPI can reflect both the substrate binding and the chemical transformation. In addition, we observed a similar  $\theta_{\text{CA-CB-CG-CD}}$  distribution of Arg202 of R154W to that of N247W (Figure S7b). This indicates that the blocking effect of Arg202 to impair the substrate binding also exists in R154W, which is consistent with the negative  $\Delta\Delta G_{K_{\text{M}}}^{\ddagger}$  value shown in Figure S6b.

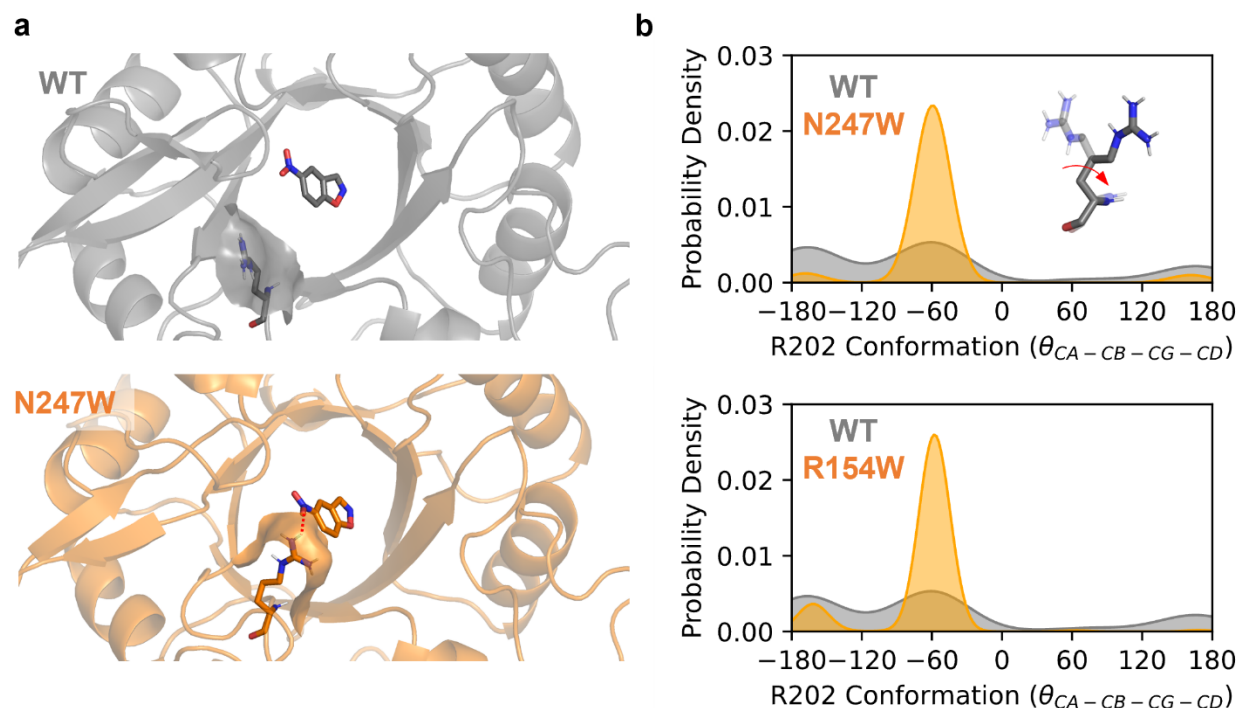


**Figure S6.** Correlation between the  $k_{\text{cat}}$  ( $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger}$ , a) and  $K_M$  ( $\Delta\Delta G_{K_M}^{\ddagger}$ , b) components of the effective activation free energy change upon mutation ( $\Delta\Delta G_{\text{eff}}^{\ddagger}$ ) versus the substrate positioning index (SPI) for KE07-R7-2 variants that are experimentally tested in this work (blue) and reported by Bhowmick *et al.*<sup>18</sup> (orange). The  $k_{\text{cat}}$  and  $K_M$  components of  $\Delta\Delta G_{\text{eff}}^{\ddagger}$  are calculated as  $\Delta\Delta G_{k_{\text{cat}}}^{\ddagger} = -RT \ln \frac{k_{\text{cat}}^{\text{mutant}}}{k_{\text{cat}}^{\text{WT}}}$  and  $\Delta\Delta G_{K_M}^{\ddagger} = -RT \ln \frac{K_M^{\text{mutant}}}{K_M^{\text{WT}}}$ , respectively. For each data point tested in this work, the mean and standard error (shown as the vertical error bar) are derived from three independently repeated kinetic measurements. In each plot, the horizontal dashed line indicates the position of  $\Delta\Delta G^{\ddagger} = 0$ . The vertical dashed line indicates the position of SPI = 1.56 where the beneficial mutant R154W is located. The vertical dashed line is also the boundary of the two-segment piecewise linear fitting. The fitting lines are shown in red and labeled with the corresponding Pearson correlation coefficient ( $r$ ). The data point of R154W is included in both fitting lines. The detail of the standard error calculation is described in Text S1.

**Table S6.** Solvent-accessible surface area (SASA) decomposition of WT, N247W, and R154W mutants. The values are in  $\text{\AA}^2$  and are based on the averages from MD snapshots. The difference is calculated as the value of the mutant subtracts from that of the WT. The SASA differences contribute most to the SASA change, i.e., Arg202 in N247W and Trp50 and Ser144 in R154W, are highlighted in bold.

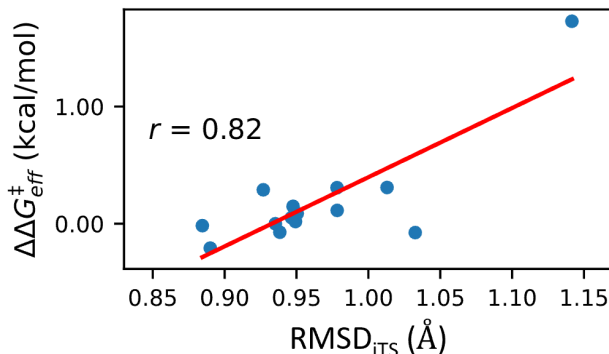
	Residue Index	WT	N247W	Diff <sub>N247W</sub>	R154W	Diff <sub>R154W</sub>
Pocket	9	4.62	3.99	-0.63	5.40	0.78
	11	47.43	49.21	1.78	49.05	1.61
	48	1.51	0.81	-0.70	1.33	-0.18
	50	55.58	57.31	1.72	40.17	<b>-15.42</b>
	101	3.05	3.75	0.70	2.23	-0.82
	128	21.30	23.50	2.20	19.35	-1.95
	201	18.41	16.40	-2.01	12.86	-5.55
	202	67.49	84.64	<b>17.15</b>	71.19	3.70
	222	7.12	7.80	0.69	6.65	-0.46
Pocket sum		226.52	247.42	20.91	208.22	-18.30
Ser144 loop	141	20.50			20.31	-0.20

	142	3.29			2.72	-0.56
	143	119.74			118.89	-0.85
	144	112.86			76.10	<b>-36.76</b>
	145	13.89			7.41	-6.48
	146	111.71			110.48	-1.23
	147	132.92			137.32	4.39



**Figure S7.** The molecular mechanism underlying the impact of N247W and R154W mutations on enzyme kinetics. (a) Representative conformation of WT (gray) and N247W (orange) with residue Arg202 and substrate shown in stick. The surfaces of Arg202 are also shown. The conformation of WT is the same as Figure 4 in the main text. The conformation of R154W was chosen because its Arg202 SASA is closest to the average shown in Table S6. In the two specific conformations shown in (a), the  $\theta_{CA-CB-CG-CD}$  of Arg202 are  $-174.3^\circ$  and  $-64.6^\circ$  for WT and N247W, respectively. The highlighted close distance between Arg202 guanidino H and substrate nitro O is  $2.9 \text{ \AA}$ . (b) Distributions of dihedral angle  $\theta_{CA-CB-CG-CD}$  of Arg202 in WT, N247W, and R154W. The distribution is derived from the corresponding MD snapshots. An illustration of the rotation axis CB-CG is shown in the top panel.





**Figure S8.** Scatter plots for the correlation between the change of activation free energy ( $\Delta\Delta G_{\text{eff}}^{\ddagger}$ ) versus the root-mean-square deviation from the idealized transition state ( $\text{RMSD}_{\text{ITS}}$ ) of the KE variants within the electric field window between  $\pm 2.88$  MV/cm. The fitted line is shown in red and labeled with the corresponding Pearson correlation coefficient.

## References

- (1) Hong, N. S.; Petrovic, D.; Lee, R.; Gryn'ova, G.; Purg, M.; Saunders, J.; Bauer, P.; Carr, P. D.; Lin, C. Y.; Mabbitt, P. D.; et al. The Evolution of Multiple Active Site Configurations in a Designed Enzyme. *Nat. Commun.* **2018**, *9*, 3900, DOI: 10.1038/s41467-018-06305-y From NLM Medline.
- (2) Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190-195, DOI: 10.1038/nature06879.
- (3) *The Pymol Molecular Graphics System, Version 2.4*; 2015.
- (4) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. Pdb2pqr: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522-W525, DOI: 10.1093/nar/gkm276.
- (5) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. Propka3: Consistent Treatment of Internal and Surface Residues in Empirical Pk(a) Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525-537, DOI: 10.1021/ct100578z.
- (6) *Amber 2018*; University of California, San Francisco, 2018.
- (7) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713, DOI: 10.1021/acs.jctc.5b00255.
- (8) Wang, N. X.; Wilson, A. K. The Behavior of Density Functionals with Respect to Basis Set. I. The Correlation Consistent Basis Sets. *J. Chem. Phys.* **2004**, *121*, 7632-7646, DOI: 10.1063/1.1792071.
- (9) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247-260, DOI: 10.1016/j.jmgm.2005.12.005.
- (10) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641, DOI: 10.1002/jcc.10128 From NLM PubMed-not-MEDLINE.

- (11) Shao, Q.; Jiang, Y.; Yang, Z. J. EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. *J. Chem. Inf. Model.* **2022**, *62*, 647-655, DOI: 10.1021/acs.jcim.1c01424.
- (12) Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; DiMaio, F. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12*, 6201-6212, DOI: 10.1021/acs.jctc.6b00819.
- (13) Frenz, B.; Lewis, S. M.; King, I.; DiMaio, F.; Park, H.; Song, Y. F. Prediction of Protein Mutational Free Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. *Front. Bioeng. Biotech.* **2020**, *8*, DOI: 10.3389/fbioe.2020.558247.
- (14) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341, DOI: 10.1016/0021-9991(77)90098-5.
- (15) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32*, 523-535, DOI: 10.1002/bip.360320508 From NLM Medline.
- (16) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690, DOI: 10.1063/1.448118.
- (17) Jiang, Y.; Stull, S. L.; Shao, Q.; Yang, Z. J. Convergence in Determining Enzyme Functional Descriptors across Kemp Eliminase Variants. *Electron. Struct.* **2022**, *4*, DOI: 10.1088/2516-1075/acad51.
- (18) 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*, 19386-19396, DOI: 10.1039/c6cp03622h From NLM Medline.
- (19) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619-2628, DOI: 10.1021/ct9003004 From NLM PubMed-not-MEDLINE.
- (20) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213-221, DOI: 10.1021/ct300321a From NLM PubMed-not-MEDLINE.
- (21) Roe, D. R.; Cheatham, T. E., 3rd. Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095, DOI: 10.1021/ct400341p From NLM PubMed-not-MEDLINE.
- (22) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *J. Chem. Phys.* **2009**, *130*, 054112, DOI: 10.1063/1.3073302.
- (23) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta.* **1973**, *28*, 213-222, DOI: 10.1007/BF00533485.
- (24) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563-576, DOI: 10.1021/acs.jctc.6b01049 From NLM Medline.

- (25) Mehmood, R.; Kulik, H. J. Both Configuration and Qm Region Size Matter: Zinc Stability in QM/MM Models of DNA Methyltransferase. *J. Chem. Theory Comput.* **2020**, *16*, 3121-3134, DOI: 10.1021/acs.jctc.0c00153 From NLM Medline.
- (26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of Resp Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631, DOI: 10.1021/ja00074a030.
- (27) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. *J. Mol. Biol.* **1973**, *79*, 351-371, DOI: 10.1016/0022-2836(73)90011-9 From NLM Medline.
- (28) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. Mdtraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528-1532, DOI: 10.1016/j.bpj.2015.08.015 From NLM Medline.
- (29) Lu, T.; Chen, F. W. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580-592, DOI: 10.1002/jcc.22885.
- (30) Waskom, M. L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021, DOI: 10.21105/joss.03021.
- (31) Gutierrez-Rus, L. I.; Alcalde, M.; Risso, V. A.; Sanchez-Ruiz, J. M. Efficient Base-Catalyzed Kemp Elimination in an Engineered Ancestral Enzyme. *Int. J. Mol. Sci.* **2022**, *23*, DOI: 10.3390/ijms23168934.