JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

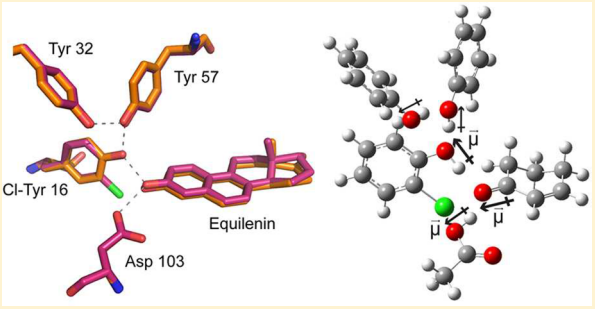
A Critical Test of the Electrostatic Contribution to Catalysis with Noncanonical Amino Acids in Ketosteroid Isomerase

Yufan Wu and Steven G. Boxer*

Department of Chemistry, Stanford University, Stanford, California 94305-5012, United States

Supporting Information

ABSTRACT: The vibrational Stark effect (VSE) has been used to measure the electric field in the active site of ketosteroid isomerase (KSI). These measured fields correlate with $\Delta G^{\ddagger}$ in a series of conventional mutants, yielding an estimate for the electrostatic contribution to catalysis (Fried et al. *Science* **2014**, *346*, 1510−1513). In this work we test this result with much more conservative variants in which individual Tyr residues in the active site are replaced by 3-chlorotyrosine via amber suppression. The electric fields sensed at the position of the carbonyl bond involved in charge displacement during catalysis were characterized using the VSE, where the field sensitivity has been calibrated by vibrational Stark spectroscopy, solvatochromism, and MD simulations. A linear relationship is observed between the electric field and $\Delta G^{\ddagger}$ that interpolates between wild-type and more drastic conventional mutations, reinforcing the evaluation of the electrostatic contribution to catalysis in KSI. A simplified model and calculation are developed to estimate changes in the electric field accompanying changes in the extended hydrogen-bond network in the active site. The results are consistent with a model in which the O−H group of a key active site tyrosine functions by imposing a static electrostatic potential onto the carbonyl bond. The model suggests that the contribution to catalysis from the active site hydrogen bonds is of similar weight to the distal interactions from the rest of the protein. A similar linear correlation was also observed between the proton affinity of KSI's active site and the catalytic rate, suggesting a direct connection between the strength of the H-bond and the electric field it exerts.



INTRODUCTION

Electrostatic stabilization has been widely discussed as contributing to enzymes' high catalytic proficiency.[1] The essential concept is that enzymes create a preorganized electrostatic environment in the active site that preferentially stabilizes the charge distribution of the transition state more than the substrate to accelerate the reaction.[1,2] Recent work from our lab using the vibrational Stark effect (VSE) to probe the electric field exerted at the active site of the model enzyme ketosteroid isomerase (KSI) provided an experimental approach to quantifying the contribution of electrostatic stabilization to catalysis in this enzyme.[3] Ketosteroid isomerase catalyzes the isomerization of its steroid substrate via formation of a dienolate intermediate with a reaction rate approximately $10^{11.5}$-fold faster than that of the uncatalyzed reaction (Figure. 1A).[4,5] An extended hydrogen-bonding network is present in the active site composed of the tyrosine triad (Y^{16}, Y^{32}, and Y^{57}) and aspartic acid 103 (Asp^{103}). Fried et al.[3] utilized a product analogue, 19-nortestosterone (19NT, Figure 1C), whose C═O group is located in KSI's active site in the same manner as the carbonyl bond of the substrate that undergoes a charge rearrangement in KSI's rate-limiting step (Figure 1B), to probe the electric field sensed by this bond. The IR spectral shifts for this carbonyl group were evaluated using vibrational Stark spectroscopy, solvatochromic data, and molecular dynamics simulations[6] and demonstrated that a very large electric field is exerted onto the C═O group of 19NT. The magnitude of the electric field was found to be linearly correlated with the activation free energy in wild-type and a series of conventional mutants, and from this the electrostatic contribution to catalysis could be obtained.[3]

While site-directed mutagenesis is widely used, a legitimate criticism of this approach, particularly for active site residues in enzymes, is that the changes in chemical functionality using natural amino acids are intrinsically large (e.g., Tyr to Phe or Asp to Asn) and can lead to unexpected structural rearrangements or water binding pockets which can confound the interpretation of the results. More specifically in the case of KSI (and a number of other enzymes), the removal of specific H-bond interactions prevents a direct comparison of electrostatic stabilization with a widely discussed alternative proposal that KSI's active site preferentially forms a strong, short H-bond with the intermediate relative to the substrate to provide additional stabilization energy.[7,8] Therefore, in this work we introduce more subtle and conservative changes by site-specifically replacing each of the key Tyr residues with 3-Cl-Tyr (Cl-Y)[9] in the KSI active site[10] to provide a critical test of the earlier analysis on electric field/

Received: July 2, 2016
Published: August 20, 2016

ACS Publications © 2016 American Chemical Society

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

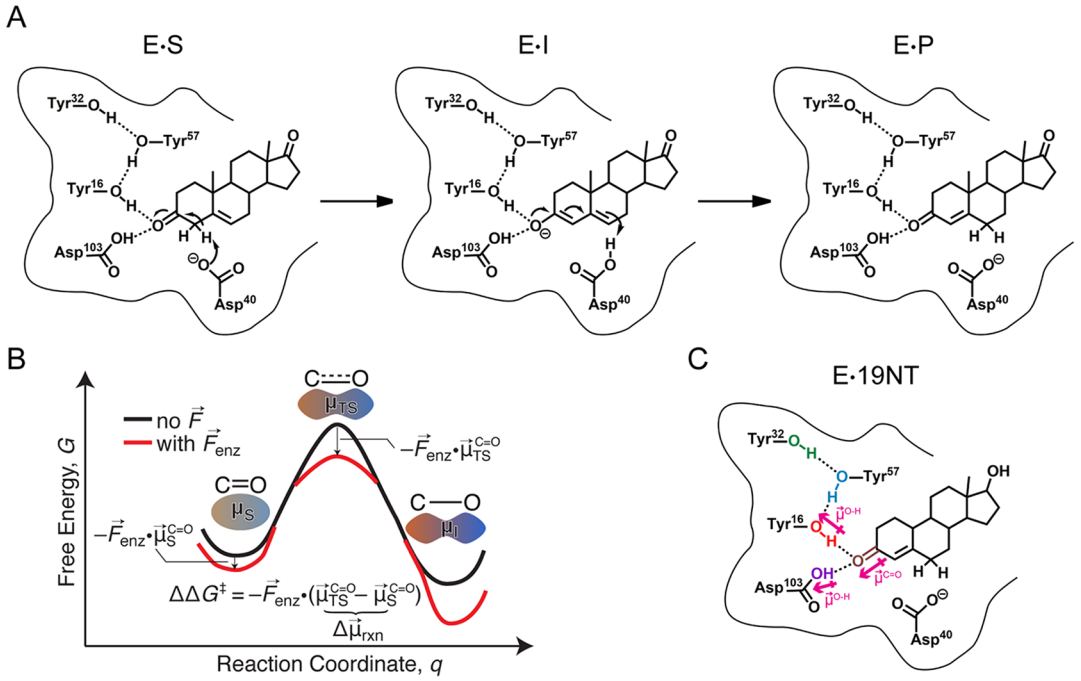


Figure 1. Catalysis by ketosteroid isomerase (KSI) and product-like inhibitor, 19-nortestosterone (19NT). (A) Reaction mechanism for KSI-catalyzed isomerization of 5-androstene-3,17-dione (substrate) to 4-androstene-3,17-dione (product). In the intermediate state, the H-bond network formed by the tyrosine triad and Asp^{103} (oxyanion hole) stabilizes the negative charge of the enolate. The dipole moment along the C═O bond increases as the reaction proceeds from substrate to intermediate, and the possible effect of the electric field, $\vec{F}_{enz}$, from the protein on the free energy of the transition state compared with the starting state is illustrated in the simplified reaction coordinate diagram in B. (C) Complex between KSI and inhibitor 19NT that closely mimics the natural substrate.

Table 1. Kinetic Parameters for Wild-Type (WT) and Cl-Y-KSI Variants (mutant) with 5-AND

	k_{cat} (10^3 s^{-1})	K_M (μM)	k_{cat}/K_M (10^7 M^{-1} s^{-1})	k_{cat} ratio (mutant/WT)	k_{cat}/K_M ratio (mutant/WT)	$\Delta G^{\ddagger a}$ (kcal mol^{-1})
WT	17.3 ± 0.5	78 ± 5	22.2 ± 2.1	1	1	11.68 ± 0.02
Cl-Y^{57}	4.08 ± 0.26	41 ± 6	9.9 ± 2.0	0.24 ± 0.02	0.45 ± 0.13	12.54 ± 0.04
Cl-Y^{32}	13.8 ± 1.1	79 ± 12	17.5 ± 4.0	0.80 ± 0.09	0.79 ± 0.25	11.82 ± 0.05
Cl-Y^{16}	4.51 ± 0.30	48 ± 7	9.3 ± 1.9	0.26 ± 0.02	0.42 ± 0.12	12.48 ± 0.04
Cl-Y^{57} D103N	0.336 ± 0.012	70 ± 5	0.56 ± 0.06	0.019 ± 0.001	0.025 ± 0.005	14.02 ± 0.02

$^a\Delta G^{\ddagger} = -RT \ln[k_{cat}/(k_B T/h)]$; T = 293 K; R = gas constant; k_B = Bolzmann's constant; $\hbar$ = Planck's constant.

activation free energy correlations. We then probe the physical basis of the large electric field and demonstrate the functional connection between the preferential H-bond interactions and electrostatic stabilization.

RESULTS AND DISCUSSION

Electrostatic Contribution to Catalysis in Conservative KSI-Cl-Y Mutants. The enzymatic activities of the wild-type and the KSI-Cl-Y variants on the natural substrate 5-androstenedione (5-AND) were determined with high accuracy (Table 1). Similar to the effect on the catalysis of the slow substrate 5(10)-estrene-3,17-dione (Table S1),[10] small but systematic changes were observed in these conservative variants, making them ideal candidates to critically test the electric field/activation free energy correlations mapped by vibrational Stark spectroscopy.[3]

The crystal structures of KSI-Cl-Y^{16} and KSI-Cl-Y^{57} bound with the transition state analogue equilenin (PDB ID: 5KP1, 5KP3) have an overall RMS deviation of 0.20 and 0.23 Å when superimposed onto the structure of wild-type KSI (PDB: 1OH0), suggesting little change in the overall enzyme architecture within the error of the structures (Figure 2, statistics in Table S2). Equilenin was chosen as the bound substrate for crystallization because of its tight binding and a binding mode similar to that of 19NT (Figure S1, PDB ID: 5KP4). Focusing on the H-bond network in the active site, in KSI-Cl-Y^{57}, the plane of the benzene ring of Cl-Y^{57} rotates about 27° from that of Y^{57} in wild-type and the O−O distance between Cl-Y^{57} and Y^{16} is shortened by 0.15 Å. The rest of the active site residues are unperturbed within the uncertainty of the X-ray structures (Table S3). In KSI-Cl-Y^{16}, the plane of the benzene ring of Cl-Y^{16} rotates about 10° from that of Y^{16} in the wild-type, the O−O distance between Y^{57} and Cl-Y^{16} is shortened by 0.07 Å on average, and that between Asp^{103} and the ligand is lengthened by 0.15 Å with the rest unperturbed. We note that although the benzene rings of the Cl-Y residues are not coplanar with their tyrosine counterparts in WT, the hydroxyl group of the residue and the carbonyl of the substrate (the chemical group directly responsible for the H-bond) retain the same relative orientation as in WT. Also, the position of the general base Asp^{40} is preserved for both variants (Figure S2), minimizing the entropic perturbation to catalysis from misplacing the residue directly involved in the reaction (i.e., the breaking and formation of covalent bonds, Figure 1A).[11,12]

The preservation of the active site H-bond network in these conservative variants allows us to relate the activation free energy

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

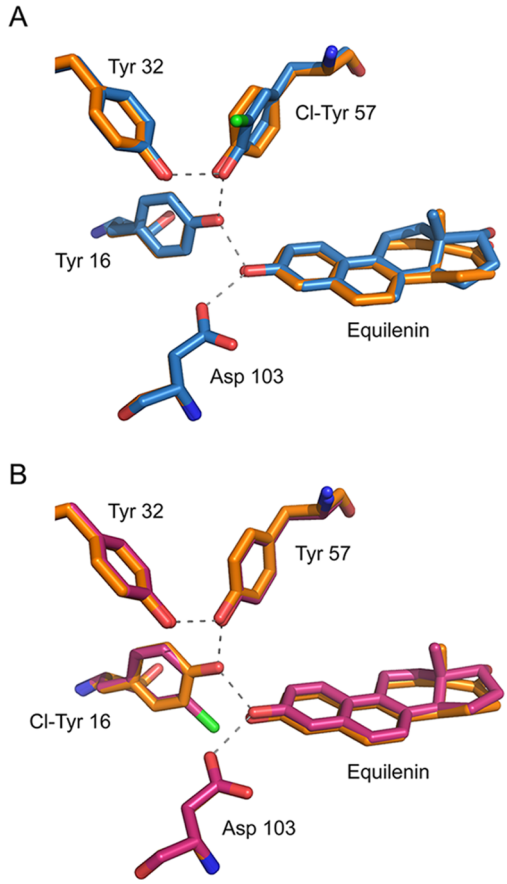


Figure 2. Structures of the tyrosine clusters in KSI-Cl-Y variants bound with transition state analogue equilenin, superimposed onto a high-resolution structure of KSI bound with equilenin (1OH0), shown in orange. (A) KSI-Cl-Y^{57} (chain a) aligned with wild-type (chain a) with an overall RMS of 0.23. (B) KSI-Cl-Y^{16} (chain a) aligned with wild-type (chain a) with an overall RMS of 0.20.

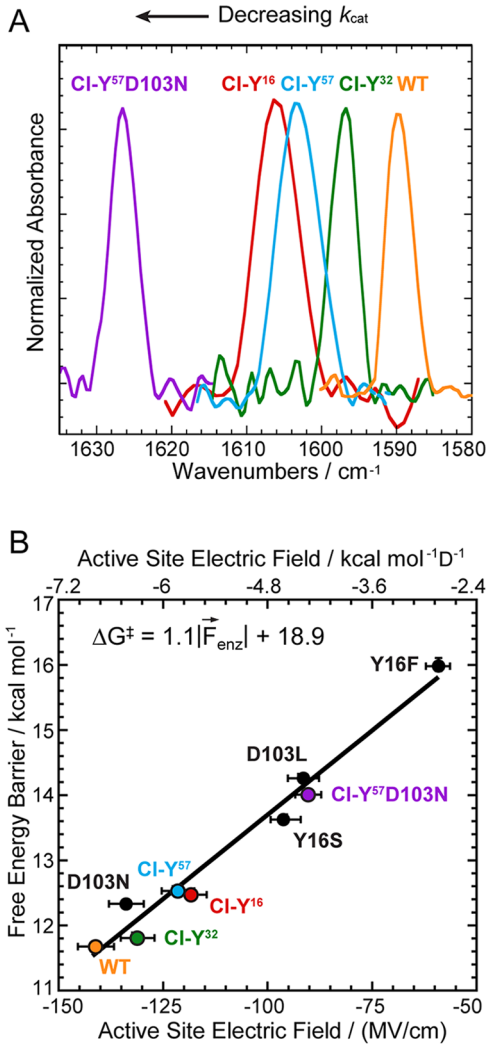


Figure 3. Preservation of the linear correlation between electric field and activation free energy in the conservative variant series using Cl-Y substitutions. (A) Infrared spectra of 19NT sulfate bound to the active site of wild-type and KSI-Cl-Y variants blue shift as k_{cat} decreases. (B) Plot of enzymatic unimolecular free energy barrier, $\Delta G^{\ddagger}$, against the electric field $|\vec{F}_{enz}|$ experienced by 19NT sulfate's C═O group in the active site of each KSI variant obtained from the IR frequency/field correlation for 19NT.

to the frequency shift of the IR stretch of the carbonyl group of bound 19NT, and from this to the electric field felt by the carbonyl group. As shown in Figure 3A, there is a systematic shift to the blue of the C═O stretch band as Y is replaced by Cl-Y at different positions. The electric field corresponding to these frequency shifts in the KSI variants was assigned to an absolute scale from 19NT's solvatochromism calibration curve (Table S4).[3,13,14] When the apparent activation energies of these conservative variants and their corresponding electric fields were plotted together with the conventional mutants, a robust linear correlation was observed (Figure 3B; R^2 = 0.98). The correlation between the activation free energy and the electric field is highly similar to the one derived from the conventional mutants (Figure S3),[3] demonstrating that changes in electrostatics associated with small perturbations on the H-bond interactions affect catalysis by the same physical mechanism as changes in electrostatics caused by large perturbations.

A Simple Model To Relate Changes in the Electric Field to Changes in the Extended Hydrogen-Bond Network. The KSI-Cl-Y variants offer a relatively simple system to examine the origins of changes in the electric field and their connection with the H-bond interactions. The most straightforward model to relate the resulting electric field to the H-bond network is the dominant dipole moments of the O−H group of tyrosine and Asp^{103} imposing a static electrostatic potential onto the carbonyl of 19NT (Figure 1C). The carbonyl group would experience a smaller electric field when H-bonded with Cl-Tyr than Tyr, as the inductive effect of Cl decreases its O−H dipole moment. Such an effect was indeed observed in a simple solution mixture mimicking this particular aspect of KSI's active site interaction: placing acetophenone in an excess amount of 2-chlorophenol from phenol blue-shifts the vibrational frequency of its carbonyl group by 3 cm^{-1}, corresponding to a ∼ 3 MV/cm decrease in electric field (Figure S4). While conceptually similar, the actual effect upon Cl-Tyr substitution at each site in the Tyr H-bond network of KSI is more profound and complex, and a proper treatment would require a full quantum simulation of the cluster as described previously for the apoenzyme;[15] here we provide a semiquantitative estimate for the change in electric field in each mutant based on a simplified computational model of the active site. The model uses phenol, 2-chlorophenol, and acetic acid to represent the active site residues Y^{32}/Y^{57}, Cl-Y^{16}, and Asp^{103}, respectively, and 2-cyclohexen-1-one to mimic the A-ring of 19NT that is directly involved in the H-bond network (Figure 4).

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

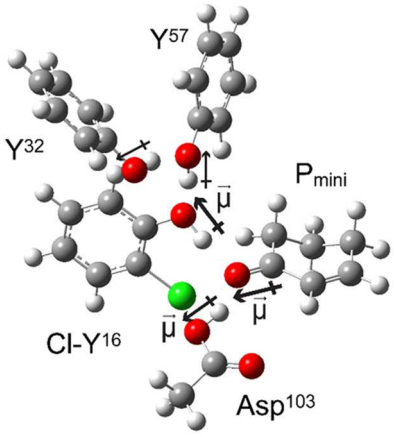


Figure 4. A simplified molecular complex mimicking the active site geometry of KSI used for DFT calculations. To construct the complex for KSI-Cl-Y^{16}, phenol, 2-chlorophenol, and acetic acid are used to represent the active site residues Y^{32}/Y^{57}, Cl-Y^{16}, and Asp^{103}, respectively, and 2-cyclohexen-1-one (P_{mini}) mimics the A-ring of 19NT that is directly involved in the H-bond network.[16] The heavy atom coordinates were taken from the crystal structures and *fixed*. The hydrogen atoms were added, and their positions were optimized by energy minimization at the B3LYP/6-31G(d,p) level. The same strategy is applied to KSI-Cl-Y^{32} and KSI-Cl-Y^{57}, as shown in Figure S5.

The change in the electric field on the carbonyl of 19NT can be the result of a smaller O−H dipole moment of the modified Tyr or/and a shift in the relative distances and orientations of the active site residues' hydroxyl groups and the carbonyl. Therefore, we first performed electric field and partial charges calculations on the simplified active site complex with fixed geometry, where the heavy atom coordinates were directly taken from the crystal structures, H atoms were added, and their coordinates were optimized for the entire complex by energy minimization at the B3LYP/6-31G(d,p) level. Next, we investigated the sensitivity of the calculated electric field to the O−O distance of a given geometry, which allowed us to dissect the calculated electric field shift into the changes in the magnitude or/and the relative positioning of the dipole moments, and to understand the bias in our modeling scheme.

The final optimized structures possess the extended H-bond network in the active site of KSI (Figure S5). The electric fields were calculated with several methodologies and compared with the experimental data (Tables 2 and S6). The estimated total field from these specific H-bond interactions is about 65 MV/cm smaller than that observed from VSE for all KSI variants, indicating that the rest of the protein architecture contributes approximately half of the total electric field experienced by the C═O bond of the substrate, as suggested previously.[3] The predicted *change* in the electric field of KSI-Cl-Y^{16} relative to WT (21 MV/cm, Table 2) is consistent with the observed shift from the VSE data (23 ± 6 MV/cm). A smaller magnitude of the O−H dipole moment in Cl-Y^{16} is predicted and estimated to account for 70% of the change, while the lengthening of the O−O distance between residue 16 and the ligand and the perturbations to the O−H dipole of Asp^{103} contribute the rest (Table 2 and SI discussion 1). A dominant contribution to the electric field from the dipole−dipole interaction between the O−H of Tyr^{16} and the C═O of the ligand is also consistent with the kinetic data where the Tyr^{16} to Phe mutation has a much more detrimental effect on the catalytic rate than mutating Asp^{103} to Asn.[17,18] On the other hand, the simple model did not reproduce the observed changes in the electric field from the more remote residues Cl-Y^{32} and Cl-Y^{57} (Tables 2 and S6). Experimentally, substitution of Cl-Y at position 57 exhibits similar electrostatic effects on 19NT as that of Cl-Y^{16}, indicating a strong interaction within the H-bond network of KSI that is capable of propagating secondary perturbations to the very site of the reaction center.[10] The highly simplified model is clearly insufficient to capture such subtly coupled interactions, as it predicted little or even opposite changes in the magnitude of the electric field upon distal substitutions. The discrepancy can be partially accounted for by the possible differences between the X-ray geometry modeled and the actual equilibrium structure of each variant in solution. An uncertainty of 0.10 Å in the O−O distance between Tyr^{16} and 19NT from the structural data could introduce an error of 8 MV/cm in the calculated electric field (Table S5), limiting the model's ability to predict an observed shift of similar magnitude (for instance, 10 MV/cm for KSI-Cl-Y^{32}, Table 2). Much more advanced theoretical methods, such as those developed in ref 15, should provide further insight.

Direct Comparison between Electrostatic Stabilization and Preferential H-Bond Interaction. There has been considerable discussion of a catalytic contribution of KSI's H-bond network due to the formation of a strong, short H-bond (SSHB) between Tyr^{16} and the intermediate along the reaction coordinate.[7,8] In this formulation, differential stabilization of the intermediate (and transition state) by KSI happens because the intermediate is a better H-bond acceptor to the active site, rather than because it has a greater dipole moment. This putative increase in H-bond strength is proposed to occur due to the matched proton affinity between the oxyanion hole[19] and the intermediate and the hydrophobic nature of the active site.[18] The primary difference between such an SSHB and electrostatic stabilization is the additional energetic benefit that might arise from the hypothesized covalent (delocalized) nature of a strong H-bond.[1,20,21] However, the consistent linear correlation between the electric field and the activation free energy across the full spectrum of KSI mutants, regardless of the presence of specific H-bonds, strongly suggests that the specific H-bond interactions within the active site of KSI play the *same* functional role in catalysis as the distal interactions from the enzyme's

Table 2. Electric Field Projected along the Carbonyl of 19NT from DFT Calculations

	WT	KSI-Cl-Y^{16}	KSI-Cl-Y^{57}	KSI-Cl-Y^{32}
$O-H_{16}$ dipole (D)[a]	2.74	2.62	2.89	2.71
$O_{16}-O_L$ (Å)	2.48	2.58	2.49	2.49
electric field calculated (MV/cm)[b]	−74.5 (−50.8, −23.7)	−53.4 (−33.6, −19.8)	−80.9 (−55.0, −25.9)	−75.5 (−51.8, −23.7)
observed from VSE (MV/cm)	−141.3 ± 4.3	−118.5 ± 3.7	−121.6 ± 3.9	−131.3 ± 4.0

[a]The dipole moments are calculated from ESP charges derived using Merz−Kollman scheme. [b]The electric field is calculated via Coulomb's law from the O−H dipoles of the two residues (16, 103) that are directly H-bonded with 19NT. The first number in the parentheses refers to the contribution to the electric field from residue 16 and the second to residue 103.

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

overall architecture. If additional (covalent) stabilization did occur, we would observe a different slope for KSI-Cl-Y mutants because the barrier would be lowered more for the same measured electric field (see SI discussion 2).

While a matched proton affinity between the active site of wild-type KSI and equilenin, a model for the intermediate, has been demonstrated previously,[19] the conservative KSI-Cl-Y variants allowed us to show next that the enzyme's preferential H-bond interaction with the intermediate over the substrate provides similar energetic benefits as that proposed from the electrostatic model. We note that the pK_a perturbations introduced by Cl-Y incorporation in these mutants are sufficient to impact the formation energy of a presumably strong H-bond interaction between the ligand and the enzyme, as suggested by a systematic change in their rate constants.[23,24] We first rationalize our strategy to quantify the catalytic benefit from the differential proton affinity match. The proton affinity of each KSI variant's active site (pK_α^E) is defined in terms of the solution pK_a of a ligand that upon binding forms a 50:50 mixture of protonated and deprotonates states (this is denoted differently from the active site's pK_a because the active site cannot exchange protons with solvent when a ligand is bound).[25] By deconvolving the electronic absorption spectrum of equilenin (or other naphthols) bound to the enzyme into protonated and deprotonated contributions, the fraction ionization of the bound substrate can be established (this can also be probed in the IR as shown in ref 26). This is shown for WT and the Cl-Y variants in Figure 5 and Figure S7, and the extracted values of the proton affinities are listed in Table 3. Empirically, the free energy of H-bond formation (Δ_fG°, which defines the strength of the H-bond) varies linearly with the degree of proton affinity match (ΔpK_α) between the H-bond partners.[20,27] Therefore, the difference in the H-bond strength between the intermediate and the enzyme ($\Delta_fG^{\circ(E\cdot I)}$) vs the substrate and the enzyme ($\Delta_fG^{\circ(E\cdot S)}$) is proportional to the difference in the ΔpK_α between the two pairs ($|pK_a^I - pK_\alpha^E| - |pK_a^S - pK_\alpha^E|$). The p$K_a$ of the intermediate was determined to be 10.0 ± 0.1 and that of the substrate's conjugate acid is approximately −2;[19,22] therefore, this model anticipates that the enzyme's active site H-bond network provides maximal activation barrier reduction when its pK_α^E matches the intermediate, and no energetic benefit when its pK_α^E equals 4 (where the H-bonds with the intermediate and the substrate have the same strength).

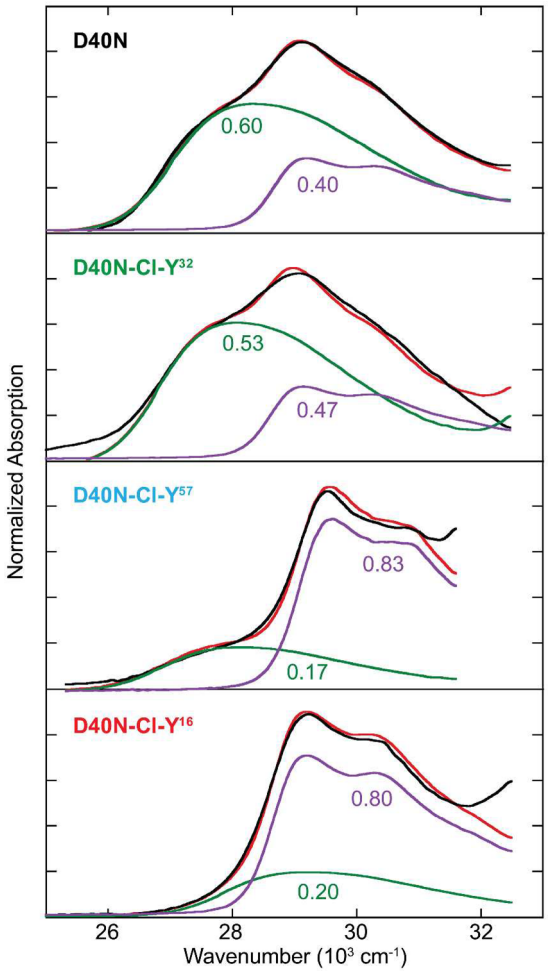


Figure 5. Proton affinities of the oxyanion hole of the KSI-Cl-Y variants. Absorption spectra of equilenin when bound to the active site of WT and KSI-Cl-Y variants fitted to a linear combination (red) of the protonated (purple) and deprotonated (green) species.

Table 3. Proton Affinity of KSI-Cl-Y Variants[a]

	proton affinity (pK_α^E)	ΔpK_α $^{(E\cdot I)}$ $\lvert pK_a^I - pK_\alpha^E\rvert$	ΔpK_α $^{(E\cdot S)}$ $\lvert pK_a^S - pK_\alpha^E\rvert$
WT	9.80 ± 0.03	0.20 ± 0.10	11.80 ± 0.10
Cl-Y^{32}	9.64 ± 0.04	0.36 ± 0.11	11.64 ± 0.11
Cl-Y^{57}	8.92 ± 0.03	1.08 ± 0.10	10.92 ± 0.10
Cl-Y^{16}	9.09 ± 0.06	0.91 ± 0.12	11.09 ± 0.12

[a]The estimated pK_a of the substrate (S) and the intermediate (I) are −2.0 and 10.0, respectively, with an uncertainty of 0.1.[19,22] The calculation details of the proton affinity can be found in Figure S7.

As shown in Figure 6, a linear correlation is observed between the free energy barrier and the proton affinity of KSI. The linear trend indicates that changes in the enzyme's additional stabilization of the intermediate by a stronger H-bond ($\Delta_fG^{\circ(E\cdot I)} - \Delta_fG^{\circ(E\cdot S)}$) can explain changes in barrier reduction. Extrapolating the correlation line to $pK_\alpha^E = 4$ yields the activation barrier without the preferential H-bonding toward the intermediate (17.7 kcal/mol). Compared to wild-type KSI, the preferential H-bond interaction is estimated to provide a stabilization energy of 6.1 ± 0.2 kcal/mol, which accounts for 60% of KSI's total rate acceleration and agrees well with the

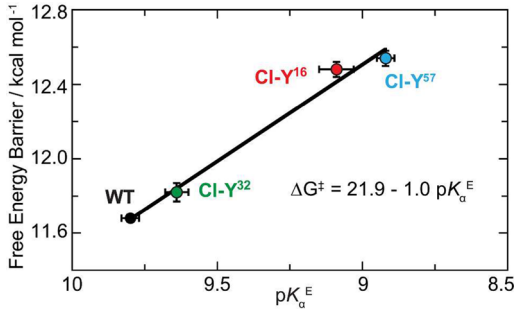


Figure 6. Linear correlation between the proton affinity of KSI's active site and the catalytic proficiency. The catalytic proficiency improves linearly as the proton affinity of the enzyme (pK_α^E) approaches that of the intermediate ($pK_a^I = 10$) ($R^2 = 0.99$). The extra stabilization energy for barrier reduction is presumably due to the increased H-bond strength between the enzyme and the intermediate, which is estimated to be 6.3 ± 0.2 kcal/mol.

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

partial contribution to barrier reduction from intermediate stabilization by computational studies.[28] The value is also comparable to the estimated barrier reduction due to the electric field in the active site (7.3 ± 0.4 kcal/mol),[3] providing an independent piece of evidence that the energetic contribution from the H-bond network in KSI is electrostatic and not due to special covalency.[29] In addition, the slope of the correlation suggests that the stabilization energy becomes 1.0 kcal/mol smaller for every pK_a unit mismatch between the enzyme and the intermediate, suggesting a noticeable sensitivity of KSI's catalytic rate to preferential H-bond interactions. The linear slope is highly reminiscent to the Brønsted β coefficient which describes the dependence of the strength of the H-bond upon the basicity of the H-bond acceptor (in our case, the proton affinity of the protein) and is larger for more nonpolar environment.[27,30] Analogous analysis yields β = 0.37 for the H-bond interaction between KSI and the ligand, reflecting a rather hydrophobic nature of KSI's active site (calculation details in SI discussion 3).

CONCLUSIONS

In summary, we test and extend the correlation between the electric field in the active site of KSI and the activation free energy for catalysis by utilizing amber suppression to add small but significant changes to the electric field and catalytic rate. The preserved linear trend demonstrates that the functional relevance of the specific H-bond interactions in KSI's extended H-bond network is of electrostatic origin, stabilizing the intermediate to about the same extent as the rest of the protein skeleton. A similar linear correlation was also observed between the proton affinity of KSI's active site and the catalytic rate, suggesting a direct connection between the strength of the H-bond and the electric field it exerts. The set of conservative variants whose H-bond network is preserved while the electrostatic properties are incrementally perturbed provides a well-controlled platform for high-level simulations to further address the functional role of H-bonds in an enzyme's active site.[15]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b06843.

Experimental details including protein preparation, crystallization, enzymatic assay, IR and electric field characterization, and DFT calculations. Further discussion on the catalytic contributions of H-bonds and modeling bias are also presented (PDF)

AUTHOR INFORMATION

Corresponding Author

*sboxer@stanford.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by grants from the NIH (GM27738 and GM118044 to S.G.B.). We thank Professor Jiangyun Wang from the Institute of Biophysics, Chinese Academy of Sciences, for making available plasmids pEVOL-CLYRS. We thank staff at SSRL and Dr. Marc Deller for support with X-ray diffraction data collection and analysis. We also thank Dr. Stephen Fried, Dr. Lu Wang, and Sam Schneider for useful discussions about this work.

REFERENCES

(1) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210−3235.
(2) Warshel, A. *J. Biol. Chem.* **1998**, *273*, 27035−27038.
(3) Fried, S. D.; Bagchi, S.; Boxer, S. G. *Science* **2014**, *346*, 1510−1513.
(4) Pollack, R. M. *Bioorg. Chem.* **2004**, *32*, 341−353.
(5) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90−93.
(6) Fried, S. D.; Bagchi, S.; Boxer, S. G. *J. Am. Chem. Soc.* **2013**, *135*, 11181−11192.
(7) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529−25532.
(8) Gerlt, J. A.; Gassman, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 11552−11568.
(9) Liu, X.; Jiang, L.; Li, J.; Wang, L.; Yu, Y.; Zhou, Q.; Lv, X.; Gong, W.; Lu, Y.; Wang, J. *J. Am. Chem. Soc.* **2014**, *136*, 13094−13097.
(10) Wu, Y.; Fried, S. D.; Boxer, S. G. *Biochemistry* **2015**, *54*, 7110−7119.
(11) Hawkinson, D. C.; Pollack, R. M.; Ambulos, N. P. *Biochemistry* **1994**, *33*, 12172−12183.
(12) Schwans, J. P.; Hanoian, P.; Lengerich, B. J.; Sunden, F.; Gonzalez, A.; Tsai, Y.; Hammes-Schiffer, S.; Herschlag, D. *Biochemistry* **2014**, *53*, 2541−2555.
(13) Fried, S. D.; Boxer, S. G. *Acc. Chem. Res.* **2015**, *48*, 998−1006.
(14) Bagchi, S.; Fried, S. D.; Boxer, S. G. *J. Am. Chem. Soc.* **2012**, *134*, 10373−10376.
(15) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18454−18459.
(16) Schwans, J. P.; Kraut, D. A.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14271−14275.
(17) Choi, G.; Ha, N.; Kim, M.; Hong, B.; Oh, B.; Choi, K. Y. *Biochemistry* **2001**, *40*, 6828−6835.
(18) Kim, S. W.; Choi, K. Y. *J. Bacteriol.* **1995**, *177*, 2602−2605.
(19) Childs, W.; Boxer, S. G. *Biochemistry* **2010**, *49*, 2725−2731.
(20) Shan, S. O.; Loh, S.; Herschlag, D. *Science* **1996**, *272*, 97−101.
(21) Oltrogge, L. M.; Boxer, S. G. *ACS Cent. Sci.* **2015**, *1*, 148−156.
(22) Zeng, B.; Pollack, R. M. *J. Am. Chem. Soc.* **1991**, *113*, 3838−3842.
(23) Mckenzie, R. H. http://condensedconcepts.blogspot.com/2014/11/strong-hydrogen-bonds-can-be.html, 2014.
(24) Natarajan, A.; Schwans, J. P.; Herschlag, D. *J. Am. Chem. Soc.* **2014**, *136*, 7643−7654.
(25) Fried, S. D.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 12271−12276.
(26) Fried, S. D.; Boxer, S. G. *J. Phys. Chem. B* **2012**, *116*, 690−697.
(27) Shan, S. O.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 14474−14479.
(28) Feierberg, I.; Åqvist, J. *Biochemistry* **2002**, *41*, 15728−15735.
(29) We note that energetic benefit estimated from the extrapolation to pK_a^E = 4 is not a reflection of the sole contribution from the active site H-bond, as the electrostatics from the protein architecture certainly affect the proton affinity and thus are encompassed in the measurements (see SI discussion 3).
(30) Stahl, N.; Jencks, W. P. *J. Am. Chem. Soc.* **1986**, *108*, 4196−4205.

DOI: 10.1021/jacs.6b06843
J. Am. Chem. Soc. 2016, 138, 11890−11895

Supporting information for:

"A Critical Test of the Electrostatic Contribution to Catalysis with Non-canonical Amino Acids in Ketosteroid Isomerase"

Yufan Wu and Steven G. Boxer

Department of Chemistry, Stanford University, Stanford, CA, 94305-5012, United States

Contents

Supplementary Figures

1. Binding modes of various ligands
2. Active site electron density maps of KSI variants
3. Linear correlation between the electric field and the catalytic efficiency of conventional mutants.
4. Simple solution model for KSI's active site interactions
5. Optimized simple molecular cluster structures mimicking the H-bond network in KSI's active site.
6. Dipole moments and orientations in WT and KSI-Cl-Y^{16} from DFT calculations.
7. Electronic spectra of naphthols bound to KSI variants to determine the proton affinity

Experimental section

Materials. 5-androstene-3,17-dione (5-AND), 19-Nortestosterone sulfate sodium salt (19NT sulfate) and equilenin were purchased from Steraloids (Newport, RI). Anhydrous solvents and deuterium oxide (99.5% D or higher) were purchased from Acros Organics and Cambridge Isotopes.

Expression and purification of KSI variants. Wild-type KSI and KSI-Cl-Y variants were overexpressed in BL21(DE3) cells (Invitrogen), purified with Ni-NTA affinity chromatography and anion exchange chromatography (GE Healthcare) as described previously. [1] The samples for protein crystallization were taken through an additional purification step with size-exclusion chromatography (Superdex 75 10/300 GL; GE Healthcare).[1] All KSI variants were characterized by LC/ESI Mass Spectrometry using maximum entropy deconvolution.

X-ray crystallography. Crystals of two D40N-Cl-Y variants (Cl-Y^{16} and Cl-Y^{57}) bound with equilenin and wild type KSI bound with 19-nortestosterone were obtained using hanging drop vapor diffusion by mixing 1 μL of 15-20 mg/mL KSI variant with 1 μL of reservoir solution (1.1-1.4 M ammonium sulfate, 40 mM KPi (pH 7.2), 3-6% isopropanol). Cubic-shaped crystals appeared after incubation for 3 days at 20°C. Cryoprotection was achieved by directly soaking crystals in 2.0 M sucrose prior to flash freezing in liquid nitrogen. Single crystal diffraction data for each KSI variant were collected at the Stanford Synchrotron Radiation Laboratory BL7-1, BL11-1 and BL12-2. Data were integrated and scaled using mosflm[2] and CCP4[3], respectively. Data collection and refinement statistics are summarized in Table S2. An initial protein model for each structure was obtained by molecular replacement with Phaser[4] using the coordinates from the previously published structure of KSI-D40N with bound 3-fluoro-4-nitrophenol (PDB 3VGN) and apo KSI-D40N-ClY^{57} (PDB 5D81) . Simulated annealing refinement was carried out using a maximum-likelihood amplitude-based target function as implemented in Phenix[5]. Further refinement was carried out with Phenix, interspersed with manual model building in Coot[6]. All structural figures were prepared using PyMOL[7]. D40N-Cl-Y^{16} was processed in space group P1 due to significant crystal slippage and the refined coordinate file was post-analyzed using Labelit[8] to determine the highest possible space group symmetry.

Kinetic assays of KSI variants. Reactions were carried out at 20°C in a solution containing 40 mM potassium phosphate (KPi), pH 7.2, 2.5 mM EDTA, 3.3% (v/v) methanol and appropriate amounts of 5-AND and enzyme. Because of the fast reaction rate, homogenous and consistent mixing of the substrate and the enzyme were achieved using the stop-flow accessory SFA-20 (TgK Scientific) coupled to a Cary 6000i UV/Vis spectrometer. For each measurement, 200 μL of enzyme solution (40 mM KPi and 0.01% - 0.1% (m/v) bovine serum albumin) and 200 μL of substrate solution (40 mM KPi, 5 mM EDTA, and 6.6% (v/v) methanol) were simultaneously injected into a 90 μL observation cell through a T-mixer, and data acquisition was started with a dead time of ~8 ms. Formation of the reaction product was monitored for 2 minutes by measuring the absorbance at 248 nm. The initial rates were determined from the first 30s of the data and fitted to both the Michaelis-Menten equation and Lineweaver-Burk reciprocal plots to determine the values of k_{cat}, K_M, and k_{cat}/K_M. [9,10] The final substrate concentrations for the kinetic assays were 11.7 μM, 23.3 μM, 40.8 μM, 58.3 μM, 87.5 μM and 151.7μM. For each KSI variant, three independent experiments with enzyme concentrations over five folds variations were performed. The presence of bovine serum albumin (BSA) in the enzyme solution was crucial for the accurate measurements as it greatly alleviated enzyme adhesion at sub-micro molar concentration. A BSA concentration up to 0.05% (m/v) in the final reaction mixture was found to have no effect on the reaction rate.[11]

FTIR spectroscopy. FTIR spectra were obtained on a Bruker Vertex 70 spectrometer with a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector using methods very similar to those described previously[12]. A liquid demountable cell was prepared with two windows (CaF_2, 0.750 in thick, Red Optronics) separated using two semicircular mylar spacers (50 μm and 75 μm). 18 μL of liquid sample was injected into the cell. For each KSI variant, the infrared spectra were obtained using samples matched as perfectly as possible using a ligand editing method introduced earlier[13] to deal with the complex background arising from the amide region of the protein that overlaps the C=O band of 19-NT sulfate. KSI was concentrated to around 4 mM, then buffer exchanged into 40 mM KPi in D_2O, pD = 7.4. The sample and reference for IR spectroscopy were always prepared in unison: first, 20μL of KSI was portioned twice into two separate tubes, one for the sample and the other reference. Ligand stocks of 19-NT sulfate, [^{18}O]19NT sulfate and

equilenin were prepared at 40 mM in DMSO. An appropriate amount of the ligand stock was then added to the protein aliquot, 19-NT sulfate for the sample and [^{18}O]19-NT sulfate or equilenin for the reference. The final liquid solution contains around 4 mM protein, 3 mM ligand and 6% (v/v) DMSO in KPi buffer. Spectra of the sample and the reference were acquired with 10 minutes nitrogen purging and averaging over 512 scans. Clear C=O features were identified in the difference spectra where the background intensity of the amide I feature was reduced to 1-3 mOD without further baseline modification. The C=O feature of each KSI mutant was observed from at least three independently prepared samples/measurements.

Determination of absolute electric field from solvatochromism model. The measured IR peak of C=O bond of 19NT sulfate in each KSI variant is converted to an absolute electric field scale using a solvatochromism model from previous work.[12] In brief: the vibrational frequencies of the carbonyl bond of 19-NT dissolved in various organic solvents and water were measured. Next, the electric field exerted onto the carbonyl bond by each solvation environment was calculated from MD simulations in GROMACS with the generalized AMBER force field (GAFF).[13] By plotting the calculated electric field in the simple solvents (solvent field) against the IR peak positions of 19NT, a linear correlation between the absolute electric field and the carbonyl's vibrational frequency was established to be $\bar{\nu}_{C=O} = 0.702\ |\vec{F}_{solv}| + 1689$. The correspondence from the solvatochromism data was then used to map the measured IR peaks of the C=O bond of 19-NT in different KSI variants to the electric field the enzyme projects on the C=O bond axis. Note that the slope of the linear correlation represents the vibrational probe's sensitivity to electric field (the Stark tuning rate), which can be independently measured by vibrational Stark spectroscopy.[14]

DFT calculations of electric field. All *ab initio* electronic structure calculations described here were carried out in Gaussian09.[15] A model molecule P_{mini}[16] for 19-nortestosterone (19NT) consisting only of the A ring of the steroid (terminating the ring with hydrogens where the fused B ring would connect) was built in Gaussview and geometry optimized by density functional theory (DFT) at the B3LYP/6-311G++(d,p) level.[17–19] Phenol, 2-chlorophenol and acetic acid were used to model the tyrosine, 3-chlorotyrosine, and aspartic acid residues, respectively, and were individually optimized at the same theory level as P_{mini}. For each variant, a simplified molecular complex mimicking the active site geometry was constructed by aligning each of the optimized

molecules onto the corresponding residue in the crystal structure in PyMol (Figure S5). The heavy atom coordinates obtained from the alignment were fixed while the hydrogen atom positions were further optimized within the entire complex at the B3LYP/6-31G(d,p) level. For all single point calculations of electric field and partial charges, B3LYP with the 6-31G(d,p) functional was used for the final reported values. Optimizations and calculations in combination with higher-level functionals (e.g. B3LYP with D3 correction or 6-311G+(d,p)) were also tested and did not yield qualitative difference in the results. Two approaches were used to estimate the effect of Cl-Tyr substitution on the electric field experienced by the carbonyl group of 19NT: (1) the electric fields were calculated directly (prop = efg) at the position of the C=O probe when P_{mini} is in the molecular complex and when it is in vacuum – the net electric field on the C=O probe due to the active site residues was then determined by subtracting the two values; (2) the electric field at the C=O probe was estimated by calculating the partial charges of the O and H atoms in residues 16 and 103 and using them with Coulomb's law (effectively the O-H dipole – C=O dipole interaction). The atomic charges of O and H were assigned by fitting against the electrostatic potential with the Merz-Kollman scheme[20] and the CHelpG scheme of Breneman[21]. The two fitting schemes were found to give similar values; the values from the Merz-Kollman scheme were used in the discussion in the main text. The full dataset can be found in Table S6 and Figure S6.

Determination of the proton affinities of KSI variants. Electronic absorption measurements were conducted with a PerkinElmer Lambda 25 spectrophotometer. Absorption spectra of KSI-naphthol complexes were recorded with 1 mM protein and 0.5 mM naphthol in 40 mM potassium phosphate buffer (pH 7) and 1 % DMSO, with protein-alone solutions serving as a blank. High-pH basis spectra were recorded in a 0.5 M NaOH / 20 % DMSO solution, and a 0.5 M HCl / 20 % DMSO solution was used for low-pH basis spectra for all naphthols except equilenin, which required 50 % DMSO for solubility. All spectra were recorded in a 1 mm quartz cuvette. The percentage ionization of each naphthol bound at the active site of D40N-Cl-Y was derived using a previously established method involving the deconvolution of the naphthol's bound spectrum into contributions from basis spectra of the fully protonated and the fully deprotonated forms.[22]

Supplementary Discussions

1. Sensitivity test and modeling bias in the DFT calculations

To test how sensitive the calculated electric field is to a given geometry, a set of structures of KSI with different O-O distance between residue 16 and ligand was prepared. For each structure, Y^{32}, Y^{57}, D^{103} and the ligand were fixed in space, while residue 16 was translated along the $O_{ligand} - O_{16}$ vector for a certain displacement (0.05-0.2 Å). The electric field of each structure projected onto the C=O probe was then calculated via Coulomb's law with the ESP charges (Merz-Kollman scheme) of the hydroxyl group of Y^{16} and D^{103} (Table S5).

We note that on average, the electric field in KSI is 7 MV/cm smaller as the O-O distance increases by 0.1 Å. Since a distance difference of 0.1 Å is within the uncertainty of the crystal structures, 7 MV/cm sets the lower limit of the error in our simple modeling scheme. This value could partially explain the model's inability to predict the relatively small changes in the electric field of KSI-Cl-Y^{57} and KSI-Cl-Y^{32} which are more remote from the probe. Nevertheless, if we take this value into consideration in our analysis of the shift between KSI-Cl-Y^{16} and WT, it seems reasonable to estimate that a lengthening of O-O distance from 2.48 Å to 2.58 Å contributes ~30% of the reduction in electric field. Since the relative orientations between the O-H dipoles and the C=O dipole of the ligand remain unperturbed in the two variants (Figure S6), the smaller partial charges (thus smaller dipole moment) on the O-H group due to chlorine addition dominates the rest of the effect.

A subtle but interesting observation emerged from the calculation. While Asp^{103} remains unchanged in its relative position to the ligand in both variants, it possesses a slightly smaller dipole (Figure S6) thus imposing a weaker electric field in KSI-Cl-Y^{16} than in WT. In the sensitivity test, we also see a systematic small decrease in its electric field as the O-O distance gets smaller. It seems to suggest that in our modeling scheme, the fitted charges of Asp^{103} encompasses the influence from the aromatic structure of the nearby residue 16 (which is not unexpected since the electrostatic potential surface is used to fit the charges)[23]. Therefore, it is not hard to imagine that Asp 103 would experience a somewhat different effect from Y^{16} or Cl-Y^{16} in the two variants, possibly due to the rotation of the benzene ring (Figure S6), leading to the small alternation in the calculated electric field.

2. Further analysis on the correlation between electric field and activation energy

The correlation between the electric field and the catalytic efficiency has been extensively examined for nine KSI variants. We see a systematic shift in the electric fields projected on the 19NT C=O bond for the three conservative mutants (Cl-Y^{32}, Cl-Y^{57}, Cl-Y^{16}). We attribute these changes in the electric field to the change in the H-bond strengths between residue 16 and the ligand; this simple model is valid if the H-bond strength derives mainly from an electrostatic interaction.[24] The double mutant Cl-Y^{57}/D103N is an interesting case because its catalytic efficiency (50 fold reduction in k_{cat}) suggests the non-additive effect in protein mutations (Cl-Y^{57} or D103N mutation independently decreases the rate by 4 fold). Moreover, Cl-Y^{57}D103N lies on the same correlation line as the conservative variants, which suggests that a change in the H-bond type from O-H to N-H has a similar functional effect as tuning the strength of the O-H ••• O interaction. Taking one step further, we see that the correlation is extended to the conventional mutants whose active site's specific H-bonds are replaced with nonspecific interactions (e.g. Y16F, D103L). These conventional mutant residues interact with the ligand through dipole-dipole interactions, yet the effect these interactions have on KSI's catalytic efficiency follows the same trend as the conservative variants' H-bond effects. The preservation of the linear correlation strongly indicates that the H-bond network in KSI's active site plays the *same* functional role as the rest of the protein, namely, contributing to the large overall electric field. If the substrate experiences significant energetic stabilization from a strong, short H-bond's covalent character, we would have expected a deviation in the correlation for conservative variants as the measured shift in electric field would only reflect a fraction of the total effect on catalysis.

3. Further interpretations on proton affinity extrapolation and Bronsted relation analysis

By investigating the correlation between the free energy barrier and proton affinity of the active site of different KSI-Cl-Y mutants, we aim to explore a second framework to explain catalytic role of H-bonds, apart from the electric field analysis discussed above. There are uncertainties in the estimation of stabilization energy from extrapolating the proton affinity to $pK_\alpha^E = 4$, for instance, the pK_a of the intermediate is used to approximate that of the transition state, which is not really measurable (though it has been widely discussed[23–29]) and is presumably somewhere between the substrate and the intermediate. This uncertainty could lead to the overestimation of the energetic benefit of H-bonds to catalysis. Moreover, since the distant electrostatic interactions from the protein architecture certainly affect the proton affinity of the enzyme, we do not regard the extrapolation as the measurement of the sole effect that would arise if the active site H-bonds were deleted. Instead, this extrapolation is more to the case when the electric field from the entire protein is zero, where the transition state does not gain preferential stabilization over the substrate.

To further interpret the slope of the correlation between the free energy barrier and the proton affinity of the variants, we performed an analysis analogous to the Bronsted relation. The Bronsted coefficient β is typically defined as the slope of the linear dependence of $\log K^{AB}$ (where K^{AB} is the equilibrium constant for the H-bond formation between a H-bond donor A and an acceptor B) on the pK_a value of the H-bond acceptor, while the pK_a of the donor remains constant.[30] In the context of KSI, we use the β coefficient to represent the sensitivity of $\log K^{AB}$ to the proton affinity difference between the enzyme and the ligand ($\Delta pK_\alpha = |pK_a^L - pK_\alpha^E|$). Because the Bronsted slopes have small dependences on the acidity of the H-bond donor[30,31], we can further approximate the β coefficient to be identical for the interactions between the intermediate and the enzyme *vs.* the substrate and the enzyme, which yields:

$$\Delta\mathbf{G}^{\ddagger} = \Delta G_o^{\ddagger} + [\Delta_f G^{\circ(E\bullet I)} - \Delta_f G^{\circ(E\bullet S)}] = \Delta G_o^{\ddagger} - 2.303\cdot R\cdot T\cdot\beta\cdot[|pK_a^I - pK_\alpha^E| - |pK_a^S - pK_\alpha^E|] =$$
$$\Delta G_o^{\ddagger} - 2.303\cdot R\cdot T\cdot\beta\cdot[pK_a^I + pK_a^S - 2\,pK_\alpha^E] = \mathbf{\Delta G_o^{\ddagger} + 2.303\cdot R\cdot T\cdot\beta\cdot[2\cdot pK_\alpha^E - 8]}$$

where $\Delta G_o^{\ddagger}$ is the free energy barrier without preferential H-bond interaction, T = 293K, R = gas constant. Plugging in the slope value of 1.0 kcal/mol, we see that the H-bond strength's sensitivity to pK_a match in KSI is similar to a relatively nonpolar environment whose Bronsted slope is ~0.37,[32] reflecting a hydrophobic nature of KSI's active site.

Supplementary Tables

Table S1 | Kinetic parameters of KSI variants with the natural substrate 5-Androstene-3,17-dione (5-AND) and a slow substrate 5(10)-Estrene-3,17-dione (5,10-EST)

	k_{cat} (s^{-1}) (5-AND)	k_{cat} (s^{-1}) (5,10-EST)	k_{cat} ratio (5-AND)	k_{cat} ratio (5,10-EST)
WT	17300 ± 500	9.0 ± 0.2	1	1
Cl-Y^{57}	4080 ± 260	1.8 ± 0.1	0.24 ± 0.02	0.20 ± 0.02
Cl-Y^{32}	13800 ± 1100	5.1 ± 0.1	0.80 ± 0.09	0.57 ± 0.02
Cl-Y^{16}	4510 ± 300	1.9 ± 0.1	0.26 ± 0.02	0.21 ± 0.02
Cl-Y^{57} D103N	336 ± 12	0.08 ± 0.01	0.019 ± 0.001	0.009 ± 0.001

Table S2 | X-ray data collection and structure refinement statistics

Dataset	D40N-Cl-Y^{57} bound with equilenin	D40N-Cl-Y^{16} bound with equilenin [a]	WT bound with 19-nortestosterone
Resolution range (Å)	47.41-1.70	24.07-1.22	36.29-1.71
Space group	$P2_12_12_1$	P1	$P2_12_12_1$
a, Å	35.18	50.42	33.85
b, Å	73.16	55.59	72.59
c, Å	94.81	55.57	94.82
α,°	90.00	79.73	90.00
β,°	90.00	89.77	90.00
γ,°	90.00	89.79	90.00
No. unique reflections	27002	158414	26203
Completeness, %	97.3	89.2	99.8
Multiplicity	4.2	5.2	6.6
R_{merge}, %	3.6	5.2	9.6
$I/\sigma_{overall}$	18.4(5.7)	17.5(4.5)	9.6(2.1)
Refinement statistics			
No. residues	255	501	254
No. waters	136	702	138
Rwork, %	22.5	16.7	20.3
Rfree, %	25.4	18.7	23.9
rmsd bond, Å	0.003	0.013	0.006
rmsd angle, °	0.60	1.33	0.85

[a] Data was reduced and refined in space group P1 due to significant crystal slippage. Post-analysis of the coordinate file using Labelit[8] suggested C2 symmetry.

Table S3 | O-O distances between the active site residues and the ligands equilenin (EQU) and 19-nortestosterone (19NT)

	Chain #	$O_{57} - O_{32}$ (Å)	$O_{57} - O_{16}$ (Å)	$O_{16} - O_{ligand}$ (Å)	$O_{103} - O_{ligand}$ (Å)
D40N : EQU	a	2.7(1)	2.5(1)	2.5(4)	2.4(8)
	b	2.6(9)	2.5(3)	2.5(5)	2.5(6)
D40N-Cl-Y^{57} : EQU	a	2.6(9)	2.4(8)	2.5(4)	2.5(2)
	b	2.7(6)	2.2(8)	2.5(0)	2.6(0)
D40N-Cl-Y^{16} : EQU	a	2.7(3)	2.4(7)	2.5(3)	2.6(6)
	b	2.7(3)	2.4(7)	2.5(8)	2.7(2)
	c	2.7(4)	2.4(6)	2.5(4)	2.6(9)
	d	2.7(3)	2.4(6)	2.5(2)	2.6(7)
WT : 19NT [a]	b	2.4(1)	2.7(6)	2.5(7)	2.6(5)

[a] co-crystallization of apoWT (chain a) and WT bound with 19NT (chain b) was observed.

Table S4 | IR Peaks and Active Site Electric Fields in the active site of KSI variants

	Peak Position ν (cm^{-1})	FWHM (cm^{-1})	Electric Field [a] $\lvert\vec{F}\rvert$ (MV/cm)	Electric Field [b] $\lvert\vec{F}\rvert$ (kcal mol^{-1} D^{-1})
WT	1590.0 ± 0.4	3.8 ± 0.3	-141.31 ± 4.31	-6.78 ± 0.21
Cl-Y^{57}	1603.8 ± 0.8	6.0 ± 0.5	-121.65 ± 3.92	-5.84 ± 0.19
Cl-Y^{32}	1597.0 ± 0.3	4.2 ± 0.4	-131.34 ± 4.03	-6.30 ± 0.19
Cl-Y^{16}	1606.0 ± 0.3	5.5 ± 0.8	-118.52 ± 3.69	-5.69 ± 0.18
Cl-Y^{57}D103N	1625.7 ± 0.7	4.4 ± 0.5	-90.46 ± 3.11	-4.34 ± 0.15

[a] $\lvert\vec{F}\rvert$ (MV/cm) = (ν -1689.2) / 0.702 [12]

[b] $\lvert\vec{F}\rvert$ (kcal mol^{-1} D^{-1}) = 0.048 × $\lvert\vec{F}\rvert$ (MV/cm) [12]

Table S5 | Sensitivity of the calculated electric field sensed at the 19NT C=O to O-O distance [a]

$O_{16} - O_L$ (Å)	Electric field in WT (MV/cm)			Electric field in Cl-Y^{16} (MV/cm)		
	Y^{16}	D^{103}	Total	Cl-Y^{16}	D^{103}	Total
2.48	-50.8	-23.7	-74.5	-40.3 (-6.7)	-19.1 (+0.7)	-59.4 (-6.0)
2.58	-41.5 (+9.3)	-24.3 (-0.6)	-65.8 (+8.7)	-33.6	-19.8	-53.4
2.63	/	/	/	-31.1 (+2.5)	-19.8 (-0)	-50.9 (+2.5)
2.68	-34.8 (+16.0)	-25.0 (-1.3)	-59.8 (+14.7)	/	/	/
Sensitivity	+ 8 MV/cm / 0.1 Å			+ 6 MV/cm / 0.1 Å		

[a] the O-O distance found in the original crystal structures of the two variants are used as reference states. The relative shifts in electric field are highlighted in red and in parenthesis.

Table S6 | Electric field calculated using different methodologies [a]

	Ab initio [b] (MV/cm)	ESP charge (Merz-kollman) (MV/cm)	ESP charge (ChelpG) (MV/cm)	ESP charge (NBO) (MV/cm)	Observed from VSE (MV/cm)
WT	-78.2	-74.5	-78.5	-98.1	-141.3
Cl-Y^{32}	-77.9	-75.5	-74.3	-97.7	-131.3
Cl-Y^{57}	-78.0	-80.9	-79.4	-97.8	-121.6
Cl-Y^{16}	-72.6	-53.4	-56.9	-93.0	-118.5

[a] All methods consistently predict a smaller electric field in KSI-Cl-Y^{16} than in WT, in agreement with the observed shift. The changes in KSI-Cl-Y^{57} and KSI-Cl-Y^{32} cannot be captured without modeling in the subtle changes in the O-O distance between Cl-Y^{16} and the ligand. [b] The electric field is derived from the electrostatic potential of the entire model molecular complex (see Fig. S5); the other three methods calculate the electric field via Coulomb's law from the O and H atoms of the two residues (Y16, D103) that are directly H-bonded with the ligand.

Table S7 | Shifts in absorption band maxima of naphthols bound to KSI variants and the fractions of neutral species [a]

ligand	protein	neutral shift (cm^{-1})	anionic shift (cm^{-1})	fractional neutral
equilenin	D40N-Cl-Y^{57}	205 ± 15	770 ± 80	0.83 ± 0.01
equilenin	D40N-Cl-Y^{16}	-155 ± 5	1650 ± 100	0.80 ± 0.03
equilenin	D40N-Cl-Y^{32}	-180 ± 20	1280 ± 50	0.38 ± 0.02
6-bromo-2-naphthol	D40N-Cl-Y^{57}	-44 ± 30	-580 ± 330	0.79 ± 0.01
6-bromo-2-naphthol	D40N-Cl-Y^{32}	-585 ± 25	195 ± 25	0.34 ±0.02
6-cyano-2-naphthol	D40N-Cl-Y^{57}	160 ± 90	-105 ± 10	0.26 ± 0.03
6-cyano-2-naphthol	D40N-Cl-Y^{32}	-1520 ± 440	100 ± 10	0.07 ± 0.03

[a] The electronic absorption of the naphthols bound to KSI exhibits additional spectral shifts in the absorption band for both the neutral and anionic forms of the ligand. The fractional ionization of the ligand was determined by fitting the absorption spectrum to a linear combinations of the two aqueous basis spectra, where the basis spectra were allowed to shift in energy in 10 wavenumber steps (but retaining the line shape) till the fitting converges. [22]

Supplementary Figures

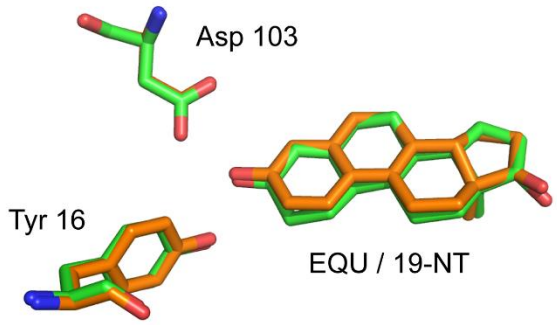


Figure S1 | 19NT and equilenin (EQU) share similar binding modes to the active site of KSI

The crystal structures of WT bound to equilenin (1OH0, orange) and the WT bound to 19NT (5KP4, green). Global alignment of the two structures gives an overall RMS of 0.23. The O-O distances between the two residues and the ligand in KSI structures bound with equilenin or 19NT are highly similar (Table S3). While EQU and 19NT have different conjugations in the four carbon-rings, KSI's active site orientated the two molecules in the binding pocket such that the C=O group participating in the H-bond interactions sit at approximately the same position. The convergence of the carbonyl orientation towards that of the intermediate seems universal across different ligands, a possible reflection of a strong, *pre-oriented* electric field in KSI's active site. The orientation of KSI's electric field is further investigated in a separate study.

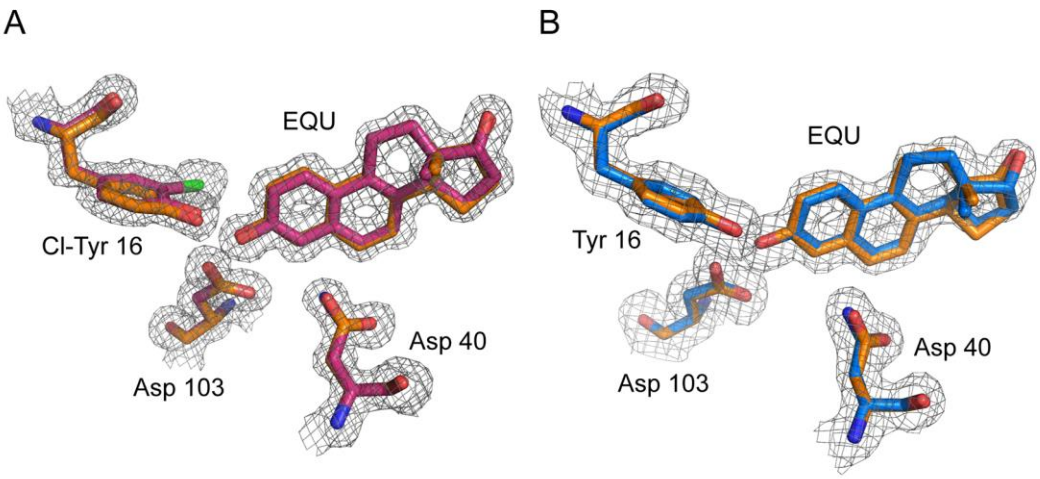


Figure S2 | Electron density maps showing the preserved geometry of active site residues (Y16, D103 and D40) upon introducing Cl-Tyr at position 57 or 16.

(A) KSI-Cl-Y^{16} (red) aligned with WT (orange); (B) KSI-Cl-Y^{57} (blue) aligned with WT (1OH0, in orange). The position of the general base aspartic acid 40 is unperturbed in both mutants.

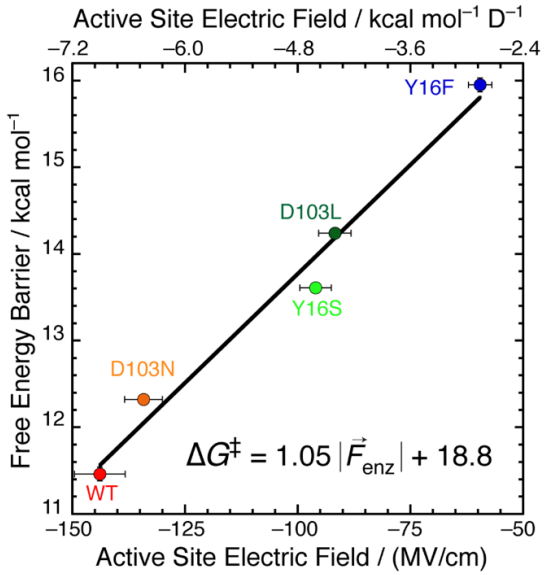
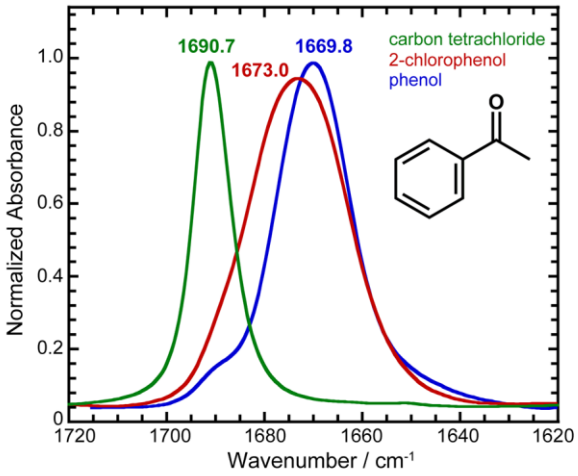


Figure S3 | Linear correlation between the apparent activation energies of conventional KSI mutants and their corresponding electric field evaluated from C=O frequency shifts, solvatochromism and MD simulations.[12]

The linear correlation between the catalytic efficiency and the electric field experienced by the carbonyl group of the substrate for conventional mutants (where H-bond interaction is replaced by non-specific interaction) is consistent with that for conserved mutants whose H-bond interactions are modulated (Figure 3). Figure reproduced from Fried *et al. Science,* **2014,** *346*, 1510–1513.



	Peak Position ν (cm^{-1})	FWHM (cm^{-1})	Electric Field Shift [a] $\lvert F \rvert$ (MV/cm)
phenol	1669.8	17.7	+ 3.02
2-chlorophenol	1673.0	23.6	
WT (Y^{16})	1590.0	3.8	+ 11.51
KSI-Cl-Y^{16}	1606.0	5.5	

[a] Based on the Stark tuning rate $|\Delta\mu|/f = 1.06 \pm 0.08$ $cm^{-1}/(MV/cm)$[10] for acetophenone and 1.39 ± 0.05 $cm^{-1}/(MV/cm)$ for 19NT[4]

Figure S4 | Effect of phenol vs. 2-Cl-phenol on the IR spectrum of acetophenone's carbonyl group as a simple solution model for the active site interaction.

6) For a similar degree of O-H dipole moment change between Cl-tyrosine and tyrosine in KSI's active site, the enzyme caused a decrease in the electric field experienced by the substrate 3-4 times larger. Although the sign is the same, the considerably smaller magnitude observed in solution is presumably because the distances in fluid solution reflect competing interactions between solvent and solute *vs.* solvent and solvent. This is a reflection of the much more constrained and pre-organized geometry found in enzyme's active site in contrast to the solute-solvent interactions in solution.

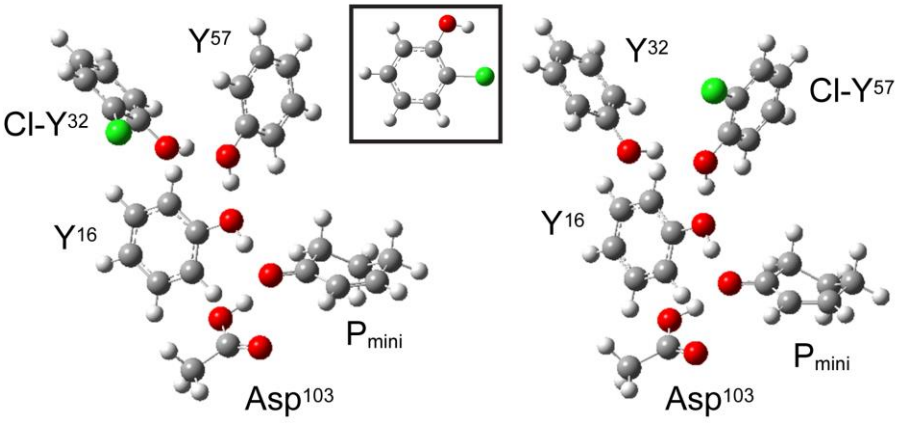
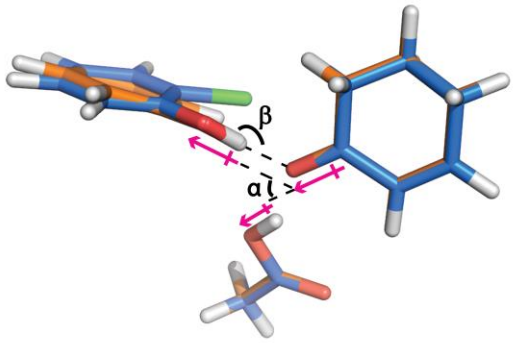


Figure S5 | The simplified molecular complex used in DFT calculations reproducing the active site geometry of KSI-Cl-Y^{32} and KSI-Cl-Y^{57}.

In the optimized structure of isolated chlorophenol, the hydrogen atom in the hydroxyl group points towards chlorine (inset). On the contrary, in the optimized geometry of the cluster, the orientation of the hydrogen atom is biased away from the chlorine to form an H-bond with the nearby tyrosine, reproducing the extended H-bond network in the active site of KSI.



	WT	Cl-Y^{16}
α (°)	73.6	71.3
β (°)	170	173
ESP charge (O_{16}) (a.u)	-0.682	-0.659
ESP charge (H_{16}) (a.u)	0.466	0.409
ESP charge (O_{103}) (a.u)	-0.586	-0.555
ESP charge (H_{103}) (a.u)	0.372	0.336

Figure S6 | Comparison of the dipole orientation and dipole moments in WT and KSI-Cl-Y^{16}

The panel on the left shows a close-up view of the optimized geometry of WT and KSI-Cl-Y^{16} simplified model complex used for the electric field calculations. The heavy atom coordinates of each complex were taken from the X-ray structure of the corresponding variant (globally aligned with each other) as described in the main text and method section. The angles between the O-H and C=O groups were measured in PyMOL. The partial charges are taken from the ESP charges fitted using the Merz-kollman scheme. Little change is observed in the relative orientation of the O-H and C=O dipoles (α) between WT and Cl-Y^{16}. The O-H ••• O interaction is linear (β close to 180º). The partial charges on the O and H atoms of the hydroxyl group of residue 16 and Asp 103 possess larger magnitudes in WT than in KSI-Cl-Y^{16}, explaining the larger electric field calculated.

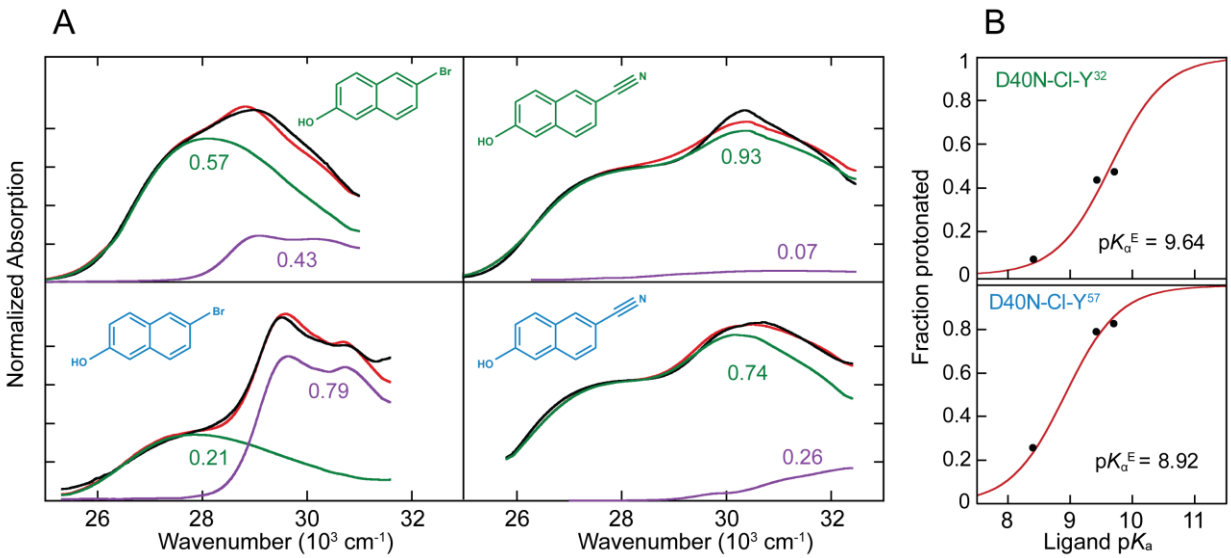


Figure S7 | Absorption spectra of KSI bound with 6-bromo-2-naphthol and 6-cyano-2-naphthol and the proton affinity of D40N-Cl-Y variants.

(A) Absorption spectra of substituted naphthols fit to a linear combination (red) of the protonated (purple) and deprotonated (green) species[22]: the top two correspond to 6-bromo-2-naphthol and 6-cyano-2-naphthol bound to D40N-Cl-Y^{32}, respectively; the bottom two correspond to 6-bromo-2-naphthol and 6-cyano-2-naphthol bound to D40N-Cl-Y^{57}, respectively. The details of the fitting results are listed in Table S7. (B) Fraction of ligand bound as the protonated species *vs.* the ligand pK_a (pK_a^L). From low to high pK_a^L, the black dots correspond to that of 6-cyano-2-naphthol, 6-bromo-2-naphthol and equilenin (Figure 4). The fractional protonation of each ligand (%HL) bound to the active site is related to its pK_a with equation (1):

$$\%HL = \frac{1}{1+\exp(-2.3\,(\mathrm{p}K_a^L - \mathrm{p}K_\alpha^E))} \qquad (1)$$

where pK_α^E is the proton affinity of the active site.[22]

Note that the end points (fraction protonated 0 and 1) are fixed; only a single point is needed to obtain pK_α^E. The proton affinities of the three variants were determined as follows: for D40N-Cl-Y^{16}, we used the pK_a (9.7) and %HL (0.8) of the bound equilenin into equation (1) to obtain its proton affinity (Figure 4A); for D40N-Cl-Y^{32} and D40N-Cl-Y^{57}, since the two naphthols also bind to the protein tightly enough to allow an estimation of their fractional protonation, as shown in (A), their proton affinities were more accurately derived by simultaneous fitting of equation (1) to the data from the three ligands (red line in (B)).

References

(1) Wu, Y.; Fried, S. D.; Boxer, S. G. *Biochemistry* **2015**, *54*, 7110–7119.
(2) Leslie, A. G. W. *Jt. CCP4+ESF-EAMCB Newsl. Protein Crystallogr.* **1992**, *26*.
(3) Collaborative Computational Project, Number 4. *Acta Crystallogr. D. Biol. Crystallogr.* **1994**, *50*, 760–763.
(4) McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. *Acta Cryst.* **2005**, *D61*, 458–464.
(5) Adams, P. D.; Grosse-Kunsleve, R. W.; Hung, L.-W.; Loerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Cryst.* **2002**, *D58*, 1949–1954.
(6) Emsley, P.; Cowtan, K. *Acta Cryst.* **2004**, *D60*, 2126–2132.
(7) DeLano, W. L. http://www.pymol.org.
(8) Poon, B. K.; Grosse-Kunstleve, R. W.; Zwart, P. H.; Sauter, N. K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 503–513.
(9) Kraut, D. A.; Sigala, P. A.; Fenn, T. D.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 1960–1965.
(10) Choi, G.; Ha, N.-C.; Kim, S. W.; Kim, D.-H.; Park, S.; Oh, B.-H.; Choi, K. Y. *Biochemistry* **2000**, *39*, 903–909.
(11) Xue, L.; Talalay, P.; Mildvan, A. S. *Biochemistry* **1990**, *29*, 7491–7500.
(12) Fried, S. D.; Bagchi, S.; Boxer, S. G. *Science* **2014**, *346*, 1510–1513.
(13) Fried, S. D.; Bagchi, S.; Boxer, S. G. *J. Am. Chem. Soc.* **2013**, *135*, 11181–11192.
(14) Fried, S. D.; Boxer, S. G. *Acc. Chem. Res.* **2015**, *48*, 998–1006.
(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; J. W. Ochterski; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian, Inc.*, Wallingford CT **2009**.
(16) Schwans, J. P.; Kraut, D. A.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14271–14275.
(17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(18) Devlin, F. J.; Finley, J. W.; Stephens, P. J.; Frisch, M. J. *J. Phys. Chem.* **1995**, *99*, 16883–16902.
(19) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(20) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
(21) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
(22) Childs, W.; Boxer, S. G. *Biochemistry* **2010**, *49*, 2725–2731.
(23) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377–395.
(24) Feierberg, I.; Åqvist, J. *Biochemistry* **2002**, *41*, 15728–15735.
(25) Cleland, W. W.; Kreevoy, M. M. *Science* **1994**, *264*, 1887–1890.

(26) Gerlt, J. A.; Kreevoy, M. M.; Cleland, W. W.; Frey, P. A. *Chem. Biol.* **1997**, *4*, 259–267.
(27) Zeng, B.; Pollack, R. M. *J. Am. Chem. Soc.* **1991**, *113*, 3838–3842.
(28) Cleland, W. W.; Frey, P. A.; Gerlt, J. A. *J. Biol. Chem.* **1998**, *273*, 25529–25532.
(29) Natarajan, A.; Schwans, J. P.; Herschlag, D. *J. Am. Chem. Soc.* **2014**, *136*, 7643–7654.
(30) Stahl, N.; Jencks, W. P. *J. Am. Chem. Soc.* **1986**, *108*, 4196–4205.
(31) Shan, S. O.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 14474–14479.
(32) Shan, S. O.; Loh, S.; Herschlag, D. *Science* **1996**, *272*, 97–101.