Supporting information for:

"Dissecting Proton Delocalization in an Enzyme's Hydrogen Bond Network with Unnatural Amino Acids"

Yufan Wu, Stephen D. Fried, Steven G. Boxer

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

Contents

Supplementary Discussions

Discussions on the preparation of KSI variants

A. Difference in purification strategies

In this study, a His-tag was engineered to the C-terminal of KSI with a thrombin cleavable linker (Table S1), and the expressed proteins were subsequently purified with Ni-NTA, cleaved and then purified by anion exchange chromatography. Compared to the earlier methods in which the protein was purified using a deoxycholate-sepharose affinity column[1,2], two main differences stand out: first, the final protein has four extra residues, LVPR, at the C-terminus; second, the protein remains folded as it was expressed in cell without unfolding and refolding *in vitro*. Careful characterizations on these newly prepared proteins are described below to ensure valid connections between this and the past work.

B. The two populations of KSI and comparison with previous constructs

Two populations of KSI are observed on anion exchange chromatography with close elution time (150 and 170mM NaCl) for all variants prepared, with the first population usually the major peak (Fig. S3). The two populations remained separate upon rerunning, showed identical and correct primary sequence as verified by ESI-MS (Fig. S3) and Tandem MS analysis, and appeared as a single band on a native gel (both are dimers). They both preserve the high binding affinity to the transition state analog equilenin as the previous construct. The ^{13}C-NMR chemical shifts of $^{13}C_{\zeta}$-Tyr show differences < 0.7 ppm for the two populations of D40N-ClY[16] (Fig. 4A and 4B). This suggests only a subtle difference in the folds of the two populations that are non-interconvertible in the native conditions. Note that only single ^{13}C-NMR peaks for each Tyr were observed for the previous KSI constructs[3]. It is possible that the *in vitro* denaturing and refolding of the protein might explore a different folding energy landscape from that *in vivo* and limit the folding heterogeneity. Because the two populations were observed for all proteins with the LVPR C-terminus prepared this way, including D40N with no Cl-Y substitution, we do not think this has anything to do with the two orientations of Cl-Y observed in D40N-Cl-Y[57]. In fact, the crystal structure of D40N-Cl-Y[57] was obtained from a purified protein sample that contained >80% of the first population, while the two Cl-Y

orientations were observed in similar amount (59% vs. 41%, Table 2). We expect that the barrier to rotation (ring-flipping) of Tyr is such that these populations would interconvert at room temperature and not be stably separable by ion-exchange chromatography.

The only noticeable difference between the two populations was observed when the pK_a of the two populations of D40N were measured (Fig. S10). The first population displayed a conventional single site titration curve with a pK_a of 5.8, while the second population appeared denatured below pH 5 and the resulting incomplete titration curve gave a pK_a of 6.9. Further characterization on the first population verified that it exhibits the same isotope effect and ^{13}C-NMR shifts as the previous construct, suggesting that its active site has the same properties as what was observed using the earlier purification protocol (Fig. S6, also see main text). Therefore, all data discussed in the main text were collected on the first protein population from the anion exchange chromatography. The only exception is the ^{13}C-NMR spectra of D40N-Cl-Y^{16}, where the mixed populations were initially measured, with subsequent separation of the two to elucidate the corresponding peaks of the first population.

Supplementary Tables

Table S1 | DNA and primary sequence of *p*KSI (D40N)[a]

Gene sequence:
ATGAACCTACCGACTGCGCAGGAAGTCCAGGGCCTGATGGCCCGTTACATCGAGCTGGT CGATGTCGGGGATATCGAGGCGATCGTGCAGATGTACGCCGATGACGCCACAGTCGAAA ACCCGTTTGGCCAGCCGCCGATCCACGGCCGCGAGCAGATTGCCGCGTTCTACCGCCAG GGTTTGGGCGGAGGCAAGGTCCGCGCCTGCCTGACCGGGCCGGTACGGGCCAGCCATA ACGGCTGCGGGGCGATGCCGTTTCGCGTCGAGATGGTCTGGAACGGCCAGCCCTGTGCA CTGGATGTCATCGATGTGATGCGCTTTGATGAGCACGGCCGGATCCAGACGATGCAAGCC TACTGGAGCGAGGTCAACCTCAGCGTGCGCGAGCCGCAG**CTGGTTCCGCGTGGTTCT CATCATCACCATCACCACTAA**
Primary sequence:
MNLPTAQEVQGLMARYIELVDVGDIEAIVQMYADDATVENPFGQPPIHGREQIAAFYRQGL GGGKVRACLTGPVRASHNGCGAMPFRVEMVWNGQPCALDVIDVMRFDEHGRIQTMQAY WSEVNLSVREPQ**LVPR▼GSHHHHHH**

[a] highlighted region corresponds to the extra amino acids from the previous construct. His-tag is eliminated by thrombin digestion in the final protein sample.

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

Dataset	D40N-Cl-Y^{57}	D40N-Cl-Y^{32}	D40N-Cl-Y^{16}
Resolution range (Å)	47.41-1.39	32.76-1.70	47.89-1.37
Space group	$C222_1$	$P2_12_12_1$	$P2_12_12_1$
a, Å	35.65	34.90	35.68
b, Å	94.83	72.76	73.28
c, Å	72.95	94.95	95.79
α,°	90.00	90.00	90.00
β,°	90.00	90.00	90.00
γ,°	90.00	90.00	90.00
No. unique reflections	25007	27352	51987
Completeness, %	98.8	99.7	97.2
Multiplicity	4.2	6.8	5.2
R_{merge}, %	3.6	5.7	5.2
$I/\sigma_{overall}$	18.4(5.7)	17.6(4.2)	17.5(4.5)
Refinement statistics			
No. residues	127	253	252
No. waters	109	112	247
Rwork, %	18.8	19.6	20.1
Rfree, %	22.2	24.8	22.3
rmsd bond, Å	0.007	0.006	0.007
rmsd angle, °	1.06	1.00	1.02

Table S3 | Fractional ionizations of each tyrosine residue in D40N-Cl-Y^{16} *a*

	D40N (first population)			D40N-Cl-Y^{16}		
	δ [ppm]	%ionized (raw)	%ionized (norm)	δ [ppm]	%ionized *b* (raw)	%ionized *b* (norm)
Tyr16	160.4	45.4	20.0	158.0/157.3	45.7	27.1
Tyr32	159.6	38.0	12.7	158.7/159.3	29.6	11.0
Tyr57	165.5	92.6	67.3	164.2/164.8	80.5	61.9
total		176	100		155.8	100

a. Fractional ionizations of tyrosine residues are calculated from the equation:

$$\%I_{raw} = (\delta_{apo} - \delta_{AA}) / (\delta_{AA}^{\ominus} - \delta_{AA}),$$

in which δ_{apo} is the chemical shift observed for the active site residue, $\delta_{AA}^{\ominus}$ and δ_{AA} are the chemical shifts of ionized and neutral amino acids (Y or Cl-Y) in Fig. S2 and reference 3. The values were further normalized using a uniform arithmetic correction so that the sum of the fractional ionizations of the three residues equals to unity:

$$\%I_{norm} = \%I_{raw} - (\sum\%I_{raw} - 100) / 3$$

The calculation neglects the effect of local protein environment on the inherent chemical shift of the three residues, assuming that their chemical shifts reflect the changes in the fractional ionization in the same way as the amino acids in the aqueous solutions. We regard the absolute value of %ionizations to be estimates and only attempt to see the trends among different KSI variants.

b. fractional ionizations are calculated for the first population of D40N-Cl-Y^{16}.

Supplementary Figures

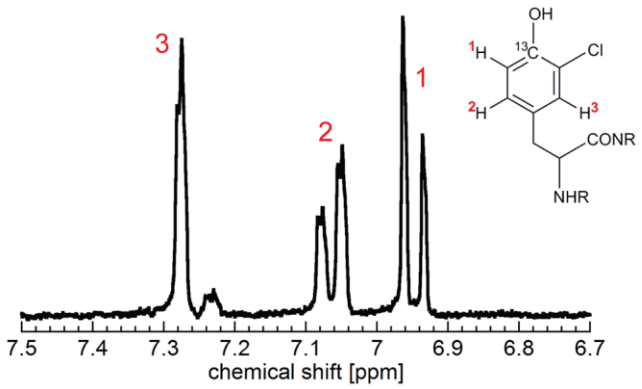


Figure S1 | ^{13}C decoupled ^{1}H NMR of ^{13}Cζ-3-chlorotyrosine.

The spectrum demonstrates the purity of the ^{13}Cζ-3-chlorotyrosine synthesized. ^{13}C decoupled ^{1}H-NMR of the aromatic protons show identical chemical shifts as those with unlabeled 3-chlorotyrosine.

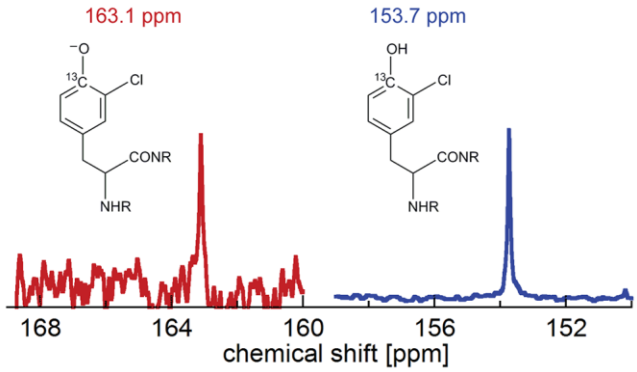


Figure S2 | ^{13}C NMR of protonated and deprotonated ^{13}Cζ-3-chlorotyrosine.

The spectrum demonstrates the purity of the ^{13}Cζ-3-chlorotyrosine synthesized and the large chemical shift difference (9.4 ppm) between its protonated (pH 4, acetic acid) and deprotonated forms (pH 13, sodium hydroxide).

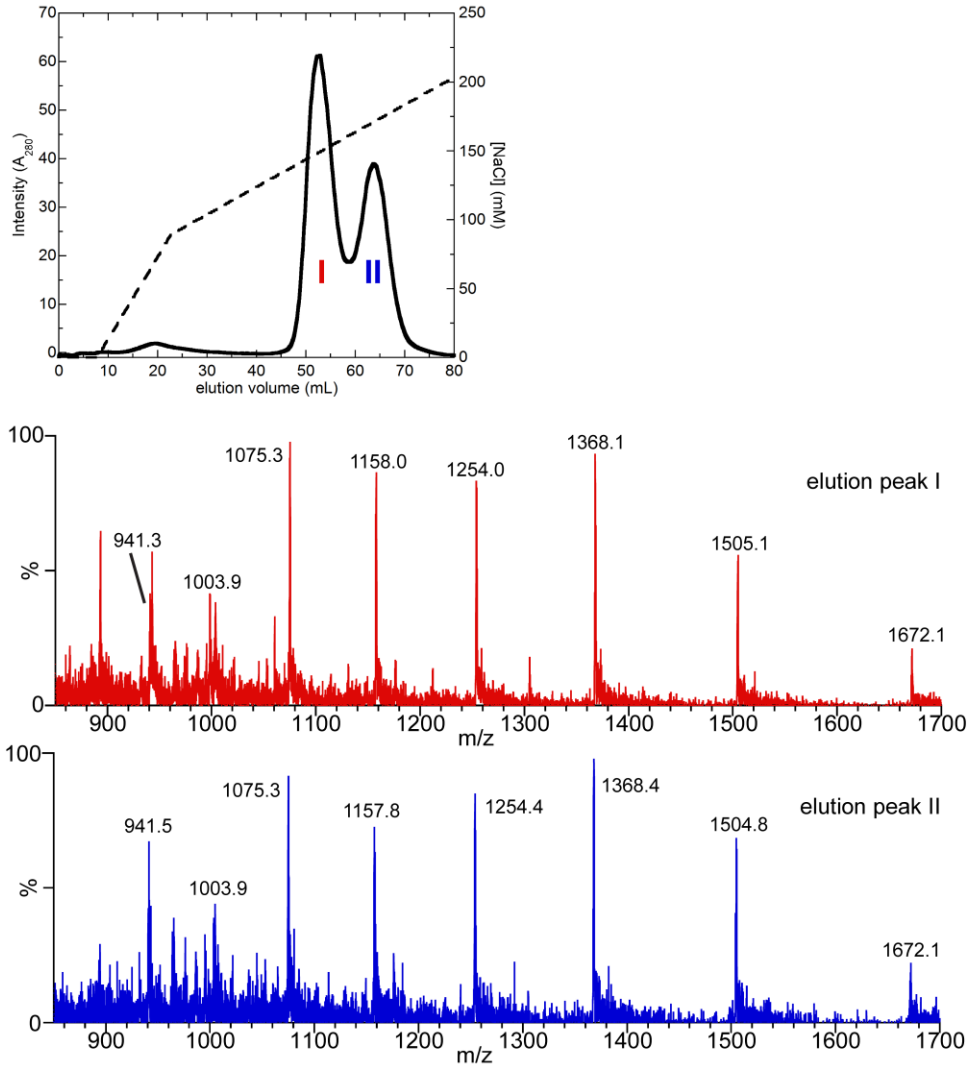


Figure S3 | Elution profile of KSI variants on anion exchange column and the corresponding ESI-MS spectra of the two populations.

Two populations of D40N-ClY[16] (with $^{13}C_{\zeta}$-ClY[16]) were separated and eluted on an anion exchange column with pH 8, 20 mM Tris buffer, 150 mM and 170 mM NaCl concentrations, respectively. The ESI-MS spectra of the two peaks have the identical fragmentation peaks, giving a molecular weight of 15039±3 Da (theoretical MW=15035.6).

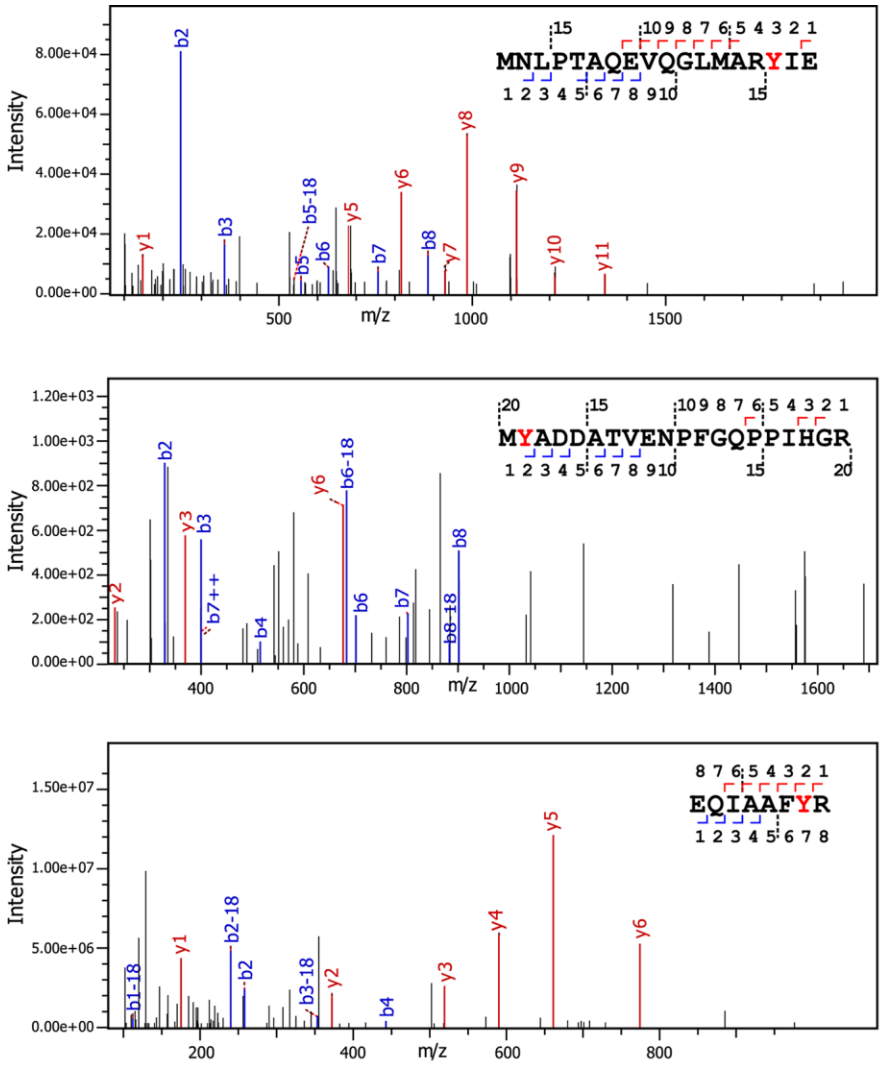


Figure S4 | Tandem-mass data on the first population of KSI variants confirming the incorporation of Cl-Y.

From top to bottom, the peptide fragments contain Cl-Y^{16}, Cl-Y^{32} and Cl-Y^{57} are identified, respectively, to confirm the unnatural amino acid incorporation. The position of Cl-Y incorporation is highlighted in red.

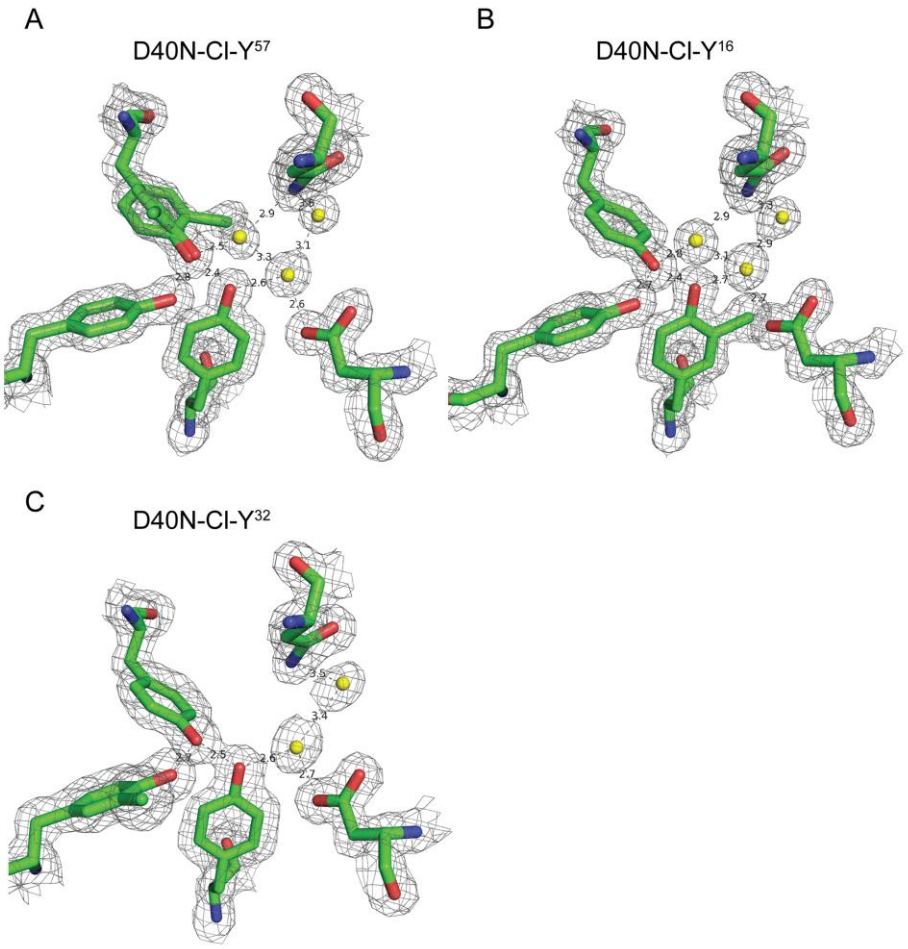


Figure S5 | 2Fo-Fc electron density map (contoured at 1.0 σ) for D40N-Cl-Y variants.

(A) The 1.4-Å structure of D40N-Cl-Y^{57}: two ordered water molecules are present in the active site, and further H-bonded to another water molecule. (B) The 1.4-Å structure of D40N-Cl-Y^{16}: two ordered water molecules are present in the active site, and further H-bonded to another water molecule. (C) The 1.7-Å structure of D40N-Cl-Y^{32}: one ordered water molecule is present in the active site, and further H-bonded to another water molecule.

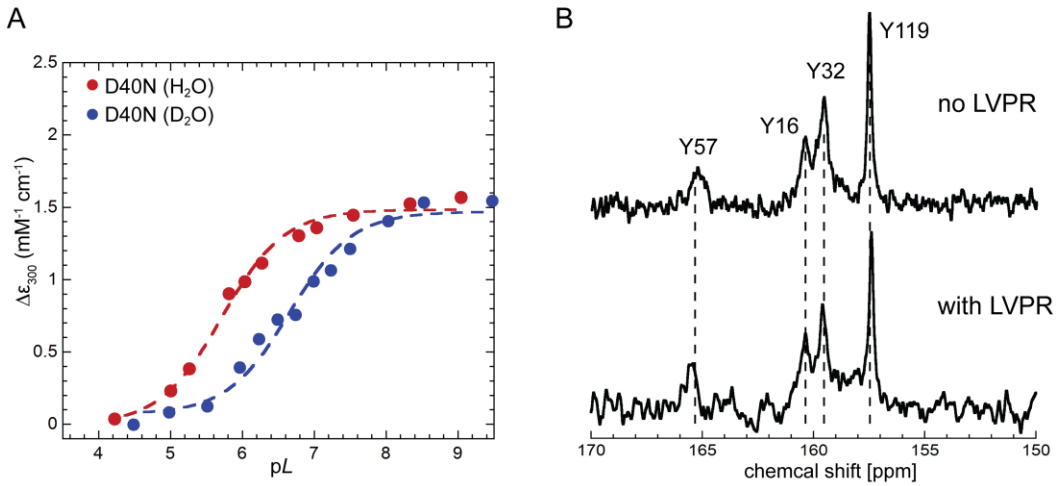


Figure S6 | The first population of D40N with C-terminal LVPR exhibits same isotope effect and ^{13}C-NMR of labeled tyrosines as the previous construct (no C-terminal LVPR).

(A) p*L* titration on D40N (first population with LVPR) yielded a pK_a (H) of 5.8 and pK_a (D) of 6.7. The large isotope effect of 0.9 ± 0.15 pK_a unit is similar to the previous construct (1.1 ± 0.14 pK_a unit). (B) ^{13}C-NMR spectrum for $^{13}C_\zeta$-Tyr-labeled D40N (lower spectrum) shows 4 peaks with chemical shift identical to the previously report values (upper spectrum): 157.4 ppm (Y119), 159.6 ppm (Y32), 160.4 ppm (Y16) and 165.5 ppm (Y57). The upper spectrum is reproduced from Fried, S.D. On the Origins of Catalysis by Ketosteroid Isomerase. Ph.D. Dissertation, Stanford University, Stanford, CA, June 2014. Both pieces of evidence support the preservation of the active site proton delocalization despite the addition of four C-terminal residues.

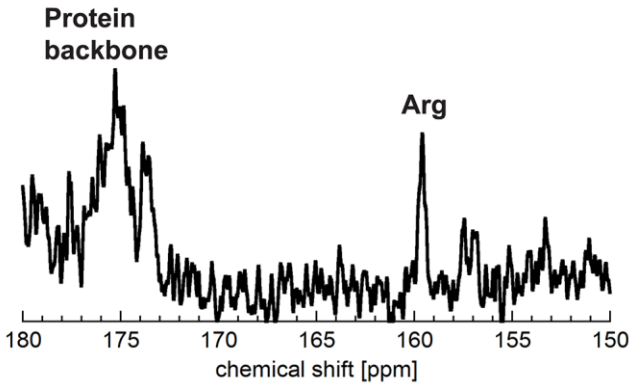
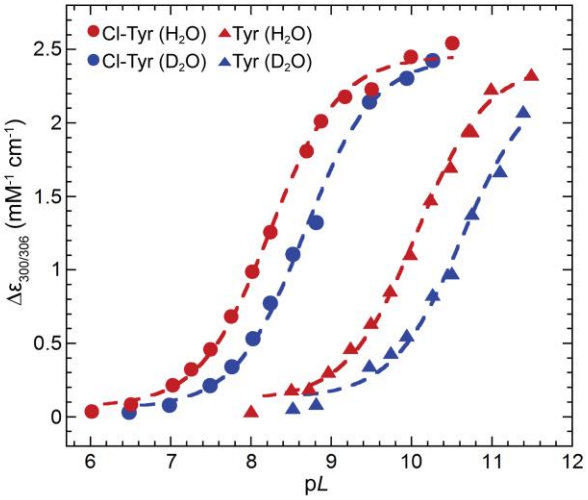


Figure S7 | ^{13}C-NMR of unlabeled wild type KSI showing signals from protein backbone and a peak at 159.6 ppm assigned to Arg.

^{13}C-NMR of unlabeled wild type KSI (1.6mM) showed signals at 170-180 ppm from the protein backbone carbonyls at natural abundance and a sharp peak at 159.6 ppm. This is likely from the fast rotating arginine side chain at the C-terminus of the protein and is absent in KSI lacking this C-terminal Arg[3,4]. This peak is observed in some spectra of labeled KSI with very good signal-to-noise (c.f. Fig. 4).



	Tyrosine (Tyr)	3-Chloro-tyrosine (Cl-Tyr)
pK_a (H)	10.08 ± 0.05	8.21 ± 0.06
pK_a (D)	10.63 ± 0.08	8.67 ± 0.05
pK_a shift (Cl-Tyr - Tyr)	--	-1.87
isotope shift	0.55 ± 0.09	0.46 ± 0.08

Figure S8 | UV-vis titration curves of 3-chlorotyrosine (Cl-Tyr) and tyrosine (Tyr) used to obtain the pK_a difference due to addition of chlorine.

Titration curves of Cl-Tyr (circles) and Tyr (triangles) in buffered H_2O (red) and D_2O (blue). Fractional ionizations of Cl-Tyr and Tyr are monitored by measuring the change in absorption at 306 nm and 300 nm, respectively. The addition of chlorine is calculated to shift the pK_a downwards by 1.87 pK_a unit.

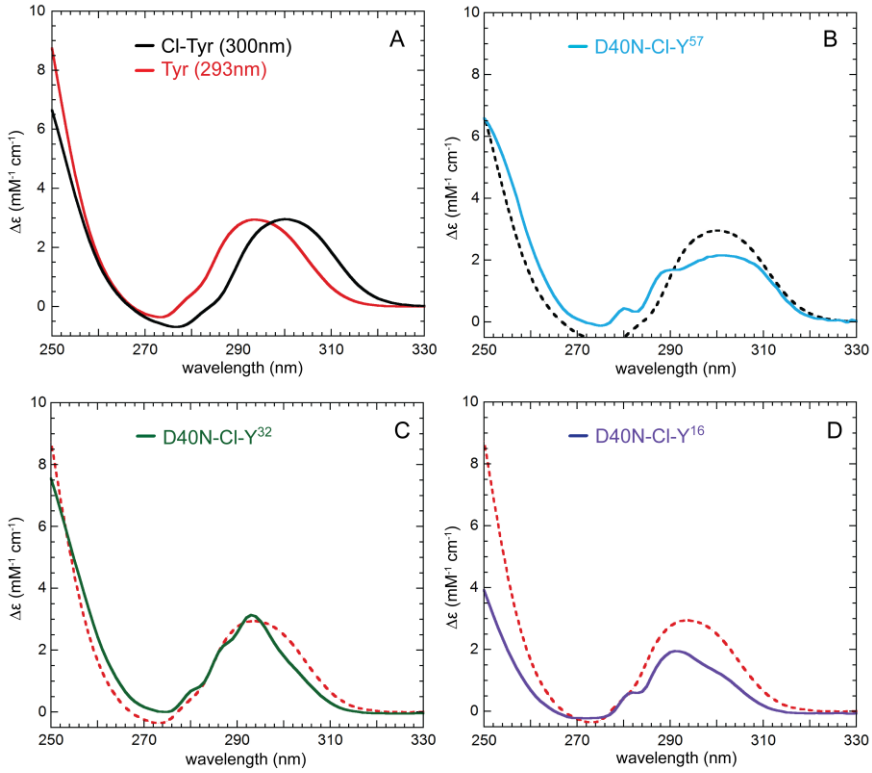


Figure S9 | Superposition of UV-vis absorption difference spectra of D40N-Cl-Y variants with amino acids in aqueous solution.

(A) Difference spectra of aqueous 3-chloro-tyrosine (Cl-Tyr, black) and tyrosine (Tyr, red), calculated from pH 10 minus pH 6, and pH 11.5 and pH 7.5 spectra, respectively. Cl-Tyr exhibits maximum absorption at 300 nm and Tyr at 293 nm. (B) Difference spectrum of D40N-Cl-Y^{57} calculated from pH 9 and pH 4, overlaid with aqueous Cl-Tyr. (C) Difference spectrum of D40N-Cl-Y^{32} calculated from pH 9.5 and pH 4, overlaid with aqueous Tyr. (D) Difference spectrum of D40N-Cl-Y^{16} calculated from pH 9 and pH 4, overlaid with aqueous Tyr.

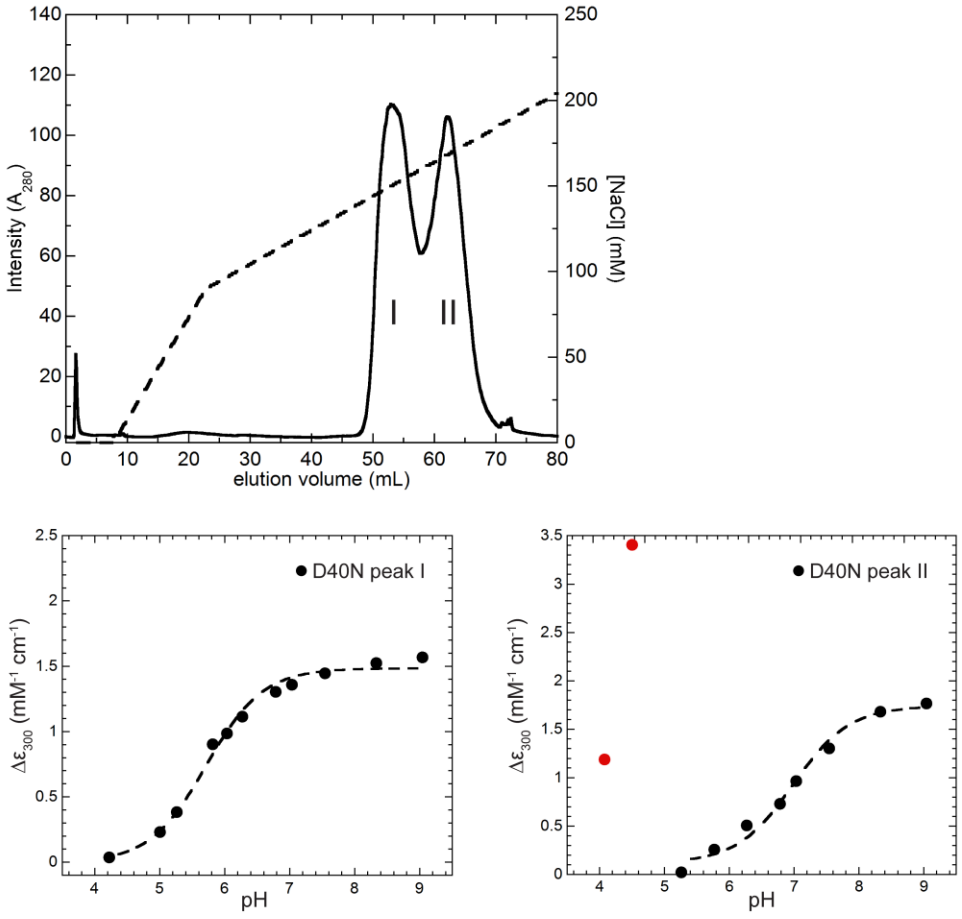


Figure S10 | Different pK_a's were observed for the two populations of D40N.

Two populations of D40N were separated and eluted on an anion exchange column with pH 8, 20 mM Tris buffer, 150 mM and 170 mM NaCl concentrations, respectively. The first population exhibited a pK_a of 5.8 with a normal shaped single site titration curve. The second population showed disrupted absorptions at pH < 5 (red dots), possibly due to the protein denaturation. The partial titration curve gave a pK_a of ~7.

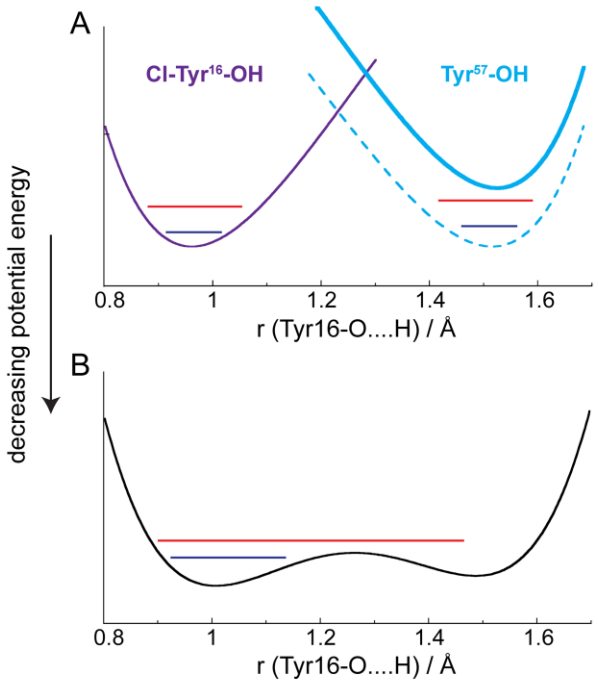


Figure S11 | Simplified two-site model for the potential energy curves of the H-bond interaction in the tyrosine triad using the interaction between Tyr^{57} and Cl-Tyr^{16} in D40N-Cl-Y^{16} as an example. [5]

(A) The diabatic potentials of Cl-Tyr^{16} (purple) and Tyr^{57} (blue, solid line) reflect the intrinsic proton affinity of the two residues. The additional coupling of Tyr^{57} to Tyr^{32} lowers (tunes) the potential of site 57 (blue, dashed line), bringing its *in situ* proton affinity closer to that of Cl-Tyr^{16}. The red and dark blue lines in each potential energy well indicate the zero point energy (ZPE) of H and D, respectively. The ZPE of H defines the proton affinity of the residue. The values on the r axis are based on the O-O distances found experimentally. The arrow next to the y-axis indicates the direction of lower potential energy. (B) The adiabatic energy potential curve from coupling of the diabatic potentials of the two H-bonded residues. The proton in the H-bonded system possesses lower energy. The energy barrier for proton transfer becomes comparable to the ZPE of H, increasing the degree of proton delocalization to provide extra stabilization energy. The difference between the ZPE of H and D becomes smaller in the shallower adiabatic potential (as compare to the diabatic states), causing a change in the pK_a isotope effect.

Reference

(1) Kim, S. W., Kim, C. Y., Benisek, W. F., and Choi, K. Y. (1994) Cloning, nucleotide sequence, and overexpression of the gene coding for delta 5-3-ketosteroid isomerase from Pseudomonas putida biotype B. *J. Bacteriol. 176*, 6672–6676.

(2) Kraut, D. A., Sigala, P. A., Pybus, B., Liu, C. W., Ringe, D., Petsko, G. A., and Herschlag, D. (2006) Testing electrostatic complementarity in enzyme catalysis: hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol. 4*, e99.

(3) Fafarman, A. T., Sigala, P. A., Schwans, J. P., Fenn, T. D., Herschlag, D., and Boxer, S. G. (2012) Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase. *Proc. Natl. Acad. Sci. U. S. A. 109*, E299–308.

(4) Sigala, P. A., Fafarman, A. T., Schwans, J. P., Fried, S. D., Fenn, T. D., Caaveiro, J. M. M., Pybus, B., Ringe, D., Petsko, G. A., Boxer, S. G., and Herschlag, D. (2013) Quantitative dissection of hydrogen bond-mediated proton transfer in the ketosteroid isomerase active site. *Proc. Natl. Acad. Sci. U. S. A. 110*, E2552–2561.

(5) McKenzie, R. H. (2012) A diabatic state model for donor-hydrogen vibrational frequency shifts in hydrogen bonded complexes. *Chem. Phys. Lett. 535*, 196–200.