# Supporting Information

## Fafarman et al. 10.1073/pnas.1111566109

### I Materials and Methods

**KSI Expression and Purification.** Thiocyanate labeling and purification of KSI-CN variants were performed as previously described (1). The $^{13}$C-labeled KSI-CN variants were obtained by substituting KCN with K$^{13}$CN (Cambridge Isotope Labs) during cyanylation, as described previously (2). For uniform $^{13}$C-tyrosine labeling of wild-type and D40N *P. putida* KSI, protein expression was carried out in M9 minimal medium (Sigma-Aldrich), augmented with 1 mM $MgSO_4$, 0.2% glucose, MEM Vitamin Solution (GIBCO), 100 μg/mL $^{13}C_\zeta$-labeled tyrosine (Cambridge Isotope Labs), and 25 μg/mL of the remaining 19 naturally occurring L-amino acids, using previously published methods (3, 4). Site-directed mutagenesis of KSI was performed as previously described (5), and sequences of all constructs were confirmed by sequencing miniprep plasmid DNA. Protein masses were confirmed by ESI LC/MS and were within 5 Da of the expected values in all cases.

**NMR Spectroscopy of KSI Variants Containing $^{13}$C-Tyrosine or a $^{13}$C-Nitrile.** One-dimensional $^{13}$C NMR spectra were acquired on Varian $^{\text{UNITY}}$INOVA NMR spectrometers (500 or 600 MHz, $^1$H frequency) running VNMR v6.1C and equipped with a 5 mm PFG switchable probe or a 10 mm broadband probe operating at ambient temperature (20 °C). Samples contained 1 mM KSI, 2 mM ligand (bound samples only), 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% $D_2O$ as the lock solvent. Spectra were acquired in a 5 mm or 10 mm Shigemi symmetrical microtube for approximately 5,000 scans and processed using a 10 Hz line broadening. Peak positions are reported as the chemical shift value corresponding to the position of maximum intensity for each peak. $^{13}$C chemical shifts were referenced to an external standard of sodium 3-trimethylsilylpropionate-2,2,3,3-$d_4$ (0 ppm) under identical solvent and temperature conditions.

**KSI X-ray Crystallography.** Crystals of KSI-CN variants in space group $P2_1$ were obtained using hanging drop vapor diffusion by mixing 2 μL of 30 mg/mL KSI (also containing equimolar equilenin in the cases of the F86C-CN/D40N•equilenin and M105C-CN/D40N•equilenin structures) with 2 μL of reservoir solution (1.1 M ammonium sulfate, 40 mM potassium phosphate (pH 7.2), 1 mM EDTA). Blade-shaped crystals appeared after incubation for 1 week at 20 °C. Cryoprotection was achieved by soaking crystals in a 1 : 1 mixture of mother liquor and 2.9 M sodium malonate followed by a direct soak in 2.9 M sodium malonate prior to flash-freezing in liquid nitrogen. Diffraction data for each KSI-CN variant from single crystals maintained at 100 K were collected at beamline 8.2.2 of the Advanced Light Source (Lawrence Berkeley National Laboratory). Data were integrated and scaled using DENZO and SCALEPACK, respectively. (6) Five percent of the observed data were set aside for cross-validation. Data collection and refinement statistics are summarized in Table S1.

An initial protein model for each structure was obtained by molecular replacement with Phaser (7) using the coordinates from the previously published structure of phenol bound to pKSI D40N (PDB ID code 2PZV). After inspection of the $\sigma_A$-weighted $2F_o$-$F_c$ and $F_o$-$F_c$ electron density maps, a complete model was constructed for each structure using Coot. (8) Simulated annealing refinement was carried out using a maximum-likelihood amplitude-based target function as implemented in PHENIX. (9) Further refinement and water picking were carried out with PHENIX, interspersed with manual corrections and model rebuilding in Coot. A final round of refinement in PHENIX treated each monomer as an independent TLS group. Structure factors and coordinates have been deposited in the RCSB Protein Data Bank, and the accession codes are given in Table S1. All structural figures were prepared using MacPyMOL (10).

**FTIR Spectroscopy.** IR spectra were obtained on a Bruker Vertex 70 FTIR spectrometer equipped with an InSb detector and a band-pass (2,000–2,500 cm$^{-1}$) interference filter from Spectrogon. For room temperature measurements, samples were held in a gas-tight demountable liquid cell (Bruker Optics) with sapphire windows and offset spacers (one 75 μm, and one 100 μm spacer on either side). Absorbance spectra were measured relative to a background with buffer. Baselines were calculated using a polynomial fit (fourth to sixth order, depending on curvature) with roots defined at least 15 cm$^{-1}$ distant from the peak maximum. Peak positions were determined using a second-derivative-based method built into the OPUS FTIR software (Bruker Photonics) and are reported as the average value from three independently prepared samples. Pictured spectra are comprised of an unweighted average of three independently prepared samples, after each spectrum is normalized to a peak height of 1. For pH dependence of the IR spectrum, the buffer 2-(*N*-morpholino)ethanesulfonic acid (MES) was used because its buffering range allowed a single agent to be used across the pH range of interest and because the protein stability in this buffer allowed for measurements down pH values as low as 5.6 at the high (mM) protein concentrations required for IR. The desired pH was achieved by mixing appropriate ratios of the buffer and its sodium salt at 1 M concentration and adding it to a concentrated solution of the protein in water, resulting in a final buffer concentration of 100 mM and final protein concentration of 2–3 mM. The pH of the protein/buffer solution was confirmed by in situ measurements with a pH microelectrode that was calibrated daily.

Vibrational Stark experiments and low temperature absorption spectra were acquired as described previously (11,12) using a home-built liquid cell consisting of a pair of 1 mm thick, 13 mm diameter sapphire windows (Meller Optics) with 40 Å of nickel vacuum deposited on the surfaces facing the sample. The nickel electrodes were connected to a high voltage DC power supply (Trek Instruments Inc.), whose output voltage was synchronized to the FTIR scan timing with a home-built control unit. The windows were separated from each other by a pair of 26-μm thick Teflon spacers and held in place with a metal clamp. The sample was rapidly frozen by immersing the cell in a custom liquid nitrogen cryostat (13) to form a glass. 512 scans with an applied field of 1 MV/cm were alternated with 512 scans with no applied field. Field-off scans were averaged and subtracted from the averaged field-on scans. The Stark tuning rate was determined for each KSI-CN construct as previously described (11, 12), and the reported values are the average and standard deviation of three independently prepared samples.

**UV-vis Absorption Measurements.** UV-vis spectra were obtained for the nitrile-modified enzymes as a function of pH using a Perkin Elmer Lambda 4 spectrophotometer with a 1-cm path length quartz cell. Difference spectra were acquired by subtraction of the absorption spectrum at pH 4.7 for wild type and D40N (Fig. 6), and at pH 5.3 for KSI-CN variants (Fig. S5). The higher pH value for the KSI-CN variants was necessary due to reduced solublity at lower pH. The spectra were baseline-corrected to a value of zero at 320 nm, and the lowest pH spectrum was subtracted from subsequent spectra to calculate the difference spec-

tra. For D40N and WT enzymes, sodium acetate was used for pH 4.7, potassium phosphate was used in the range of 4.7 to 9.1 and glycine was used in the range of 9.0 to 10, with a buffer concentration of 10 mM in all cases. For the KSI-CN variants, 20 mM MES buffer was used and pH-adjusted by adding aliquots of 0.1 M HCl while monitoring the pH with a microelectrode in situ. The enzyme concentration was 10–20 μM.

**Poisson–Boltzmann Modeling.** DelPhi, (14, 15) a finite difference algorithm for solving the Poisson–Boltzmann equation, was employed for each different charge-perturbation modeled above. Charge and radii parameters were taken from the PARSE parameter set (14), which treats only heavy atoms and hydrogens attached to heteroatoms explicitly. Charge and radii parameters for the nonnatural Cys-CN amino acid were taken from a previous study (12). A dielectric of value, ϵ, of 80 was used for the protein exterior. The value for the ϵ assigned to the protein interior presents an important choice: $\varepsilon = 2$ is justified based on the intrinsic polarizability of the chemical groups the protein is composed of (16), whereas it has been recognized that higher dielectric values can be employed in continuum electrostatic calculations to compensate for unaccounted for charge and dipole rearrangements (17,18). Through X-ray crystallography and NMR spectroscopy we have attempted to minimize the structural and ionization state changes, respectively, that are not explicitly accounted for, therefore a lower value of ϵ is called for. Calculations were performed for $\varepsilon = 2$ and $\varepsilon = 4$, with $\varepsilon = 2$ providing a lower rmsd from the experimental dataset; all calculated values reported were therefore taken from the $\varepsilon = 2$ calculation. Heavy atoms positions were determined from the X-ray structures, with hydrogen atoms automatically added to heteroatoms using the program PDB2GMX, (19) based on database values for bond distances and angles, H-bond-satisfaction criteria and unperturbed $pK_a$ values. To avoid artifacts from the finite grid treatment, a seven tier focusing routine was used. A calculation of the potential on a $50 \times 50 \times 50$ grid of 0.5 grid lines per Ångstrom, centered around the CN-bond and encompassing the whole protein provided the initial the boundary conditions for a subsequent calculation at two times finer grid spacing. This process was repeated until the grid spacing was 32 lines per Ångstrom. The gradient in the potential along the CN-bond axis was calculated at each tier, and monitored for convergence; all values reported are from the final tier.

**Error Analysis for Correction of Electric Field Measurements for Hydrogen bond Formation.** IR frequencies were corrected in cases of direct hydrogen bonding to a nitrile probe by subtracting a constant term from the vibrational frequency based on the magnitude of its departure from the IR/NMR correlation line (Fig. 3) for ethylthiocyanate in simple solvents, as described in the main text. This correction, which assumes that systematic nonelectrostatic effects do not dominate the observed chemical shift across the series of aprotic solvent observations (see ref. 2), introduces uncertainty in the corrected values for IR frequencies. To estimate this uncertainty, we evaluated the degree of scatter in data points from the $^{13}C$ NMR/IR correlation line of ethylthiocyanate in aprotic solvents (slope = −1.66 (±28%) [$cm^{-1}$/ppm]) that was used to make the hydrogen bond correction. The scatter in these data points presumably arises from environmental perturbations of unspecified origin to either the $^{13}C$ NMR or IR observable (e.g., anisotropic ring current effects on the chemical shift due to solute-toluene complexes). We calculated the standard error in the slope of this correlation using Eq. **S1** below, where $x$ is the measured $^{13}C$ NMR chemical shift and $y$ is the measured IR frequency for each ethylthiocyanate data point and obtained a relative error of 28%. This value was used as the error estimate for each corrected IR frequency generated by subtraction of its offset from the ethylthiocyanate correlation line.

$$\text{Standard error} = \sqrt{\frac{\sum (y-\bar{y})^2 - \frac{[\sum (x-\bar{x})(y-\bar{y})]^2}{\sum (x-\bar{x})^2}}{n-2}} \qquad \textbf{[S1]}$$

The Stark tuning rate has two sources of error, statistical error (10% standard deviation across the three probe-sites) and the systematic error due to the unknown local field correction factor, which for this situation is estimated as 10% (20). The chemical shift error is estimated to be 0.02 ppm, or approximately 2%. The total relative error in electrostatic fields calculated when comparing hydrogen-bonded and non-hydrogen-bonded cases is the sum of all of these relative errors, or 50%. In cases where the hydrogen bonds are not made or broken, the error is calculated using the uncertainty in the Stark tuning rate and the statistical error (i.e., standard deviation) in the measurement of the peak frequency by IR.

## II. Discussion of IR and NMR transition linewidths.

As discussed in the main text, the linewidths of the nitrile IR stretching transitions for the apo-D40N KSI variants follow the order M105C-CN < M116C-CN < F86C-CN, both at room temperature (Fig. 2 *D–F*) and at 80 K (Fig. 4). The overall linewidth of a spectroscopic transition results from the complex interplay between inhomogeneous broadening (due to distinct subpopulations with unique transition frequencies within the ensemble), motional narrowing (due to rapid interconversion between these subpopulations), and system-dependent excited-state lifetime-limited broadening. The observation of identical hierarchies at temperatures low enough to freeze out many motional narrowing mechanisms suggests that the observed difference in linewidths at the three probe-sites is primarily due to inhomogeneous broadening. Additional evidence that linewidth differences are due to differential contributions from inhomogenous broadening is provided by the observation of an identical hierarchy of linewidths in the $^{13}C$ NMR spectra of the nitrile probes (Fig. S3). The time scale defined by the full-width-at-half-maxima (FWHM) for the NMR peaks, approximately 100 Hz, is nine orders of magnitude slower than the time scale defined by the IR linewidths of approximately 10 $cm^{-1}$ or 300 GHz. Inhomogeneity that persists on the long time scale defined by NMR will necessarily be detected on the much faster timescale of IR. We propose that the observed differences in IR and NMR linewidth for the individual nitriles arise from differences in conformational heterogeneity for each nitrile arising from differences in local packing interactions, as discussed in the main text.

## III. Discussion of Monitoring Tyr ionization by NMR and UV-vis.

We attempted to use NMR to ascertain the $pK_a$ of the most downfield tyrosine resonance in Fig. 7. Increasing the pH to 8 resulted in no detectable change in the NMR spectrum. Lowering the pH to 6.2 resulted in peak broadening and a diminished signal-to-noise (Fig. S6). The broadening increased until the signal was obscured but before new resonances from the protonated tyrosine network were observed. This observation is common for widely separated chemical shifts in dynamic exchange whose rate of exchange depends on pH (21). This effect limits the utility of NMR for determining a $pK_a$ for this ionization. Furthermore, the broadened resonances do not appear to disappear due to changes in relaxation time with pH. At pH 6.4, where the intensity of the peak at 165.5 ppm is diminished relative to pH 7.1, increasing the recycle delay 4× (more than doubling the spacing between pulses) had no effect on the peak height. Fitting the downfield peak with

a lorentzian lineshape allows determination of the linewidth at each pH. Loss in peak height from pH 7.1 to pH 6.5 is exactly compensated for by increase in peak width, so that net area is unchanged.

The change in the ultraviolet absorbance spectra upon ionization has also been used to study the tyrosine ionization state in KSI (22) and provides the means to study the pH titration without the signal being obscured by chemical exchange. An increase in intensity and shift of the absorption maximum from 275 to 293 nm accompanies ionization of tyrosine; additionally a strong absorbance at 240 nm is present for ionized tyrosine. UV-vis spectra were obtained for the nitrile-modified enzymes as a function of pH and difference spectra were calculated relative to the absorption spectrum at pH 5.3 (Fig. S4). The choice to reference the spectra at pH 5.3 was based on the observation of increased absorbance due to light scattering at lower pH values, which suggested protein unfolding and precipitation. The pH-dependent changes confirmed that the probe-modified proteins experienced a change in protonation state of a tyrosine in the pH ranges explored by in Fig. 2 *J–L*.

## IV Discussion of $pK_a$ Prediction Algorithms.

To test the ability of continuum electrostatic calculations to predict active site $pK_a$ values for D103 and Y16 in D40N KSI, we submitted the crystal structure of the D40N and D40N/D103N mutants to three web-based, continuum electrostatic calculation engines for the prediction of protein $pK_a$ values: H++ (23), PROPKA (24), and Karlsberg+ (25) (Table S2). These electrostatic calculation engines not only fail to predict a low $pK_a$ tyrosine in D40N but universally predict tyrosine $pK_a$ values perturbed upward, even for the D40N/D103N mutant, which eliminates any charge-charge repulsions predicted due to an anion incorrectly predicted at D103. The failures of continuum electrostatic methods for quantitatively predicting $pK_a$ values have previously been noted by others (26, 27), and we have previously demonstrated that such disagreement between experiment and theory extends to electrostatic fields as well (28). However, disagreement often hinges on ambiguous structural information, especially concerning the positions of protons, and in one case the agreement was shown to improve upon relaxing modeled structures using molecular dynamics (17). Nevertheless, molecular dynamics-based structural adaptation is built in to the Karlsberg+ method (Table S2), yet this still does not lead to agreement with the experiment. In this case, a reasonable protonation state was used as a starting point; however, molecular dynamics by itself will not allow for the intrinsically quantum mechanical movement of charge from one site to another or account for novel situations in which charge is delocalized.

## V. Discussion of Electrostatic Field Calculations in Response to Charge Perturbations.

To obtain an estimate of the spectroscopic change expected to accompany ionizations of active site residues to compare to the experimental values, we employed the software package DelPhi (15) to calculate electric fields and Eq. **1** to calculate the spectral shift. During this study, we also explored molecular dynamics-based methods for calculating electrostatic fields present in the course of a simulation, using methods described previously (12). This approach was abandoned due to the preponderance of water in the immediate vicinity of the nitrile and the well-known difficulties in accurately representing the interaction potential between nitriles and water (29, 30).


1. Sigala PA, Fafarman AT, Bogard PE, Boxer SG, Herschlag D (2007) Do ligand binding and solvent exclusion alter the electrostatic character within the oxyanion hole of an enzymatic active site? *J Am Chem Soc* 129:12104–12105.
2. Fafarman AT, Sigala PA, Herschlag D, Boxer SG (2010) Decomposition of vibrational shifts of nitriles into electrostatic and hydrogen-bonding effects. *J Am Chem Soc* 132:12811–12813.
3. Markley JL, Kainosho M (1993) in *NMR of Macromolecules: A Practical Approach*, ed Roberts GCK (Oxford Univ Press, Oxford), pp 101–152.
4. Kato-Toma Y, Iwashita T, Masuda K, Oyama Y, Ishiguro M (2003) PKa measurements from nuclear magnetic resonance of tyrosine-150 in class C beta-lactamase. *Biochem J* 371:175–181.
5. Kraut D, et al. (2006) Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4:501–519.
6. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode *Methods Enzymol* 276:307–326.
7. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458–64.
8. Emsley P, Cowtan K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
9. Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
10. Schrödinger, LLC (2010) The PyMOL Molecular Graphics System, version 1.3r1.
11. Andrews SS, Boxer SG (2000) Vibrational stark effects of nitriles I. Methods and experimental results. *J Phys Chem A* 104:11853–11863.
12. Fafarman AT, Boxer SG (2010) Nitrile bonds as infrared probes of electrostatics in ribonuclease S. *Phys Chem B* 114:13536–13544.
13. Andrews SS, Boxer SG (2000) A liquid nitrogen immersion cryostat for optical measurements. *Rev Sci Instrum* 71:3567–3569.
14. Sitkoff D, Sharp K, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 98:1978–1988.
15. Gilson MK, Sharp K, Honig B (1988) Calculating the electrostatic potential of molecules in solution—method and error assessment. *J Comput Chem* 9:327–335.
16. Harvey SC, Hoekstra P (1972) Dielectric-relaxation spectra of water adsorbed on lysozyme. *J Phys Chem* 76:2987–2994.
17. Suydam I, Snow C, Pande V, Boxer S (2006) Electric fields at the active site of an enzyme: Direct comparison of experiment with theory. *Science* 313:200–204.
18. Schutz CN, Warshel A (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* 44:400–417.
19. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Modeling* 7:306–317.
20. Andrews SS, Boxer SG (2002) Vibrational Stark effects of nitriles II. Physical origins of stark effects from experiment and perturbation models. *J Phys Chem A* 106:469–477.
21. Sudmeier JL, Evelhoch JL, Jonsson NB-H (1980) Dependence of NMR lineshape analysis upon chemical rates and mechanisms: Implications for enzyme histidine titrations. *J Magn Reson* 40:377–390.
22. Li YK, Kuliopulos A, Mildvan AS, Talalay P (1993) Environments and mechanistic roles of the tyrosine residues of delta-5-3-ketosteroid isomerase. *Biochemistry* 32:1816–1824.
23. Gordon JC, et al. (2005) H++: A server for estimating $pK_a$s and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33:W368.
24. Bas DC, Rogers DM, Jensen J (2008) Very fast prediction and rationalization of $pK_a$ values for protein-ligand complexes. *Proteins* 73:765–783.
25. Kieseritzky G, Knapp E (2008) Optimizing $pK_a$ computation in proteins with pH adapted conformations. *Proteins* 71:1335–1348.
26. Antosiewicz J, McCammon JA, Gilson MK (1996) The determinants of $pK_a$s in proteins. *Biochemistry* 35:7819–7833.
27. Warshel A (1981) Calculations of enzymatic-reactions—calculations of $pK_a$, proton-transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry* 20:3167–3177.
28. Webb LJ, Boxer SG (2008) Electrostatic fields near the active site of human aldose reductase: 1. New inhibitors and vibrational stark effect measurements. *Biochemistry* 47:1588–1598.
29. Choi J-H, Oh K-I, Lee H, Lee C, Cho M (2008) Nitrile and thiocyanate IR probes: Quantum chemistry calculation studies and multivariate least-square fitting analysis. *J Chem Phys* 128:134506.
30. Lindquist BA, Corcelli SA. Nitrile groups as vibrational probes: Calculations of the C N infrared absorption line shape of acetonitrile in water and tetrahydrofuran. *J Phys Chem B* 112:6301–6303.

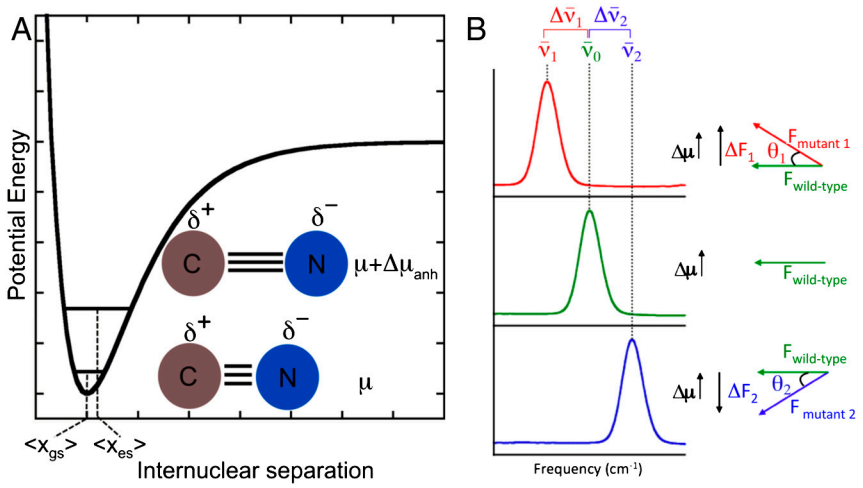


**Fig. S1.** Origins of the linear Stark effect. (*A*) An anharmonic potential energy function describing the CN stretching mode. As a consequence of the anharmonicity, different average interatomic separation distances in the ground and vibrationally excited states, $\langle x_{gs} \rangle$ and $\langle x_{es} \rangle$, respectively, lead to an increase in the dipole moment, $\Delta\mu_{anh}$, in the excited state. The increase in the dipole moment leads to a different relative effect of an electric field on the two states, and thus a field-dependent change in the energy of excitation. The total influence of the field on the CN stretching transition also includes an effect on the force constant of the bond, $\Delta\mu_{bond}$. For details see ref. 20. (*B*) Spectra for three variants of a hypothetical protein with a fixed nitrile probe. The field projection F in the three separate constructs—wild-type, mutant 1, and mutant 2—is different in magnitude and/or angle, $\theta$ as a consequence of the mutation, leading to different projections ΔF, on the fixed position of $\Delta\bar{\mu}$. This leads to changes in the transition frequency, $\Delta\bar{\nu}$, relative to WT.

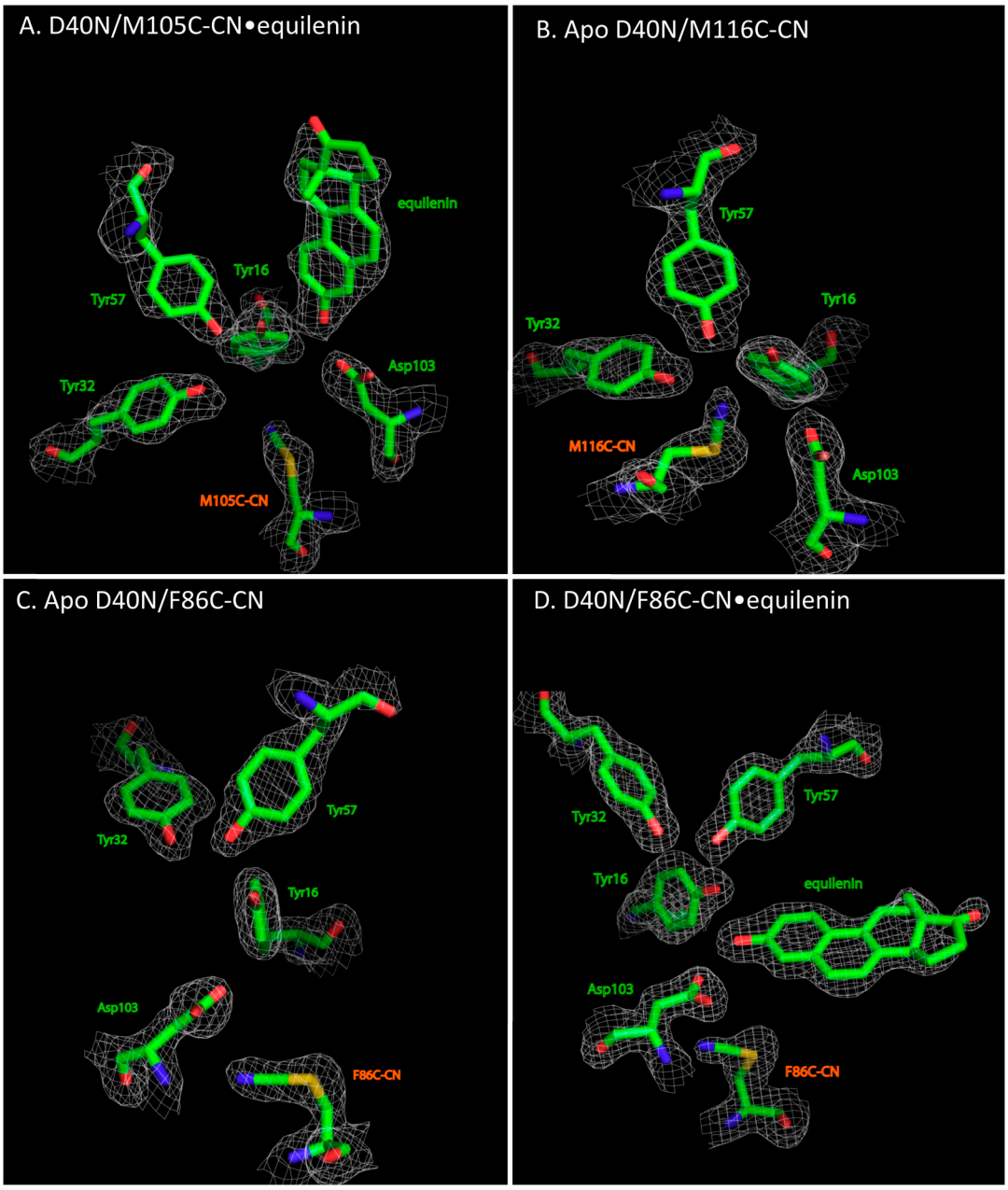


**Fig. S2.** $2F_o$-$F_c$ Electron density maps KSI-CN variants. The electron density, contoured at 1.5 sigma, is displayed as a wire mesh superimposed upon the structural model shown in stick format. Carbon, oxygen, nitrogen, and sulfur atoms are shown in green, blue, red, and gold, respectively.

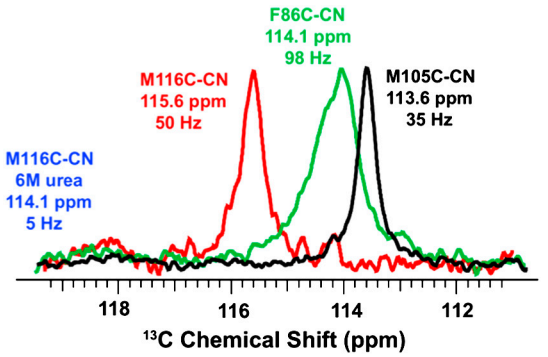


**Fig. S3.** NMR spectra of $^{13}$C-nitrile labeled pKSI D40N variants. Spectral parameters are also shown for M116C-CN/D40N unfolded in 6 M urea.

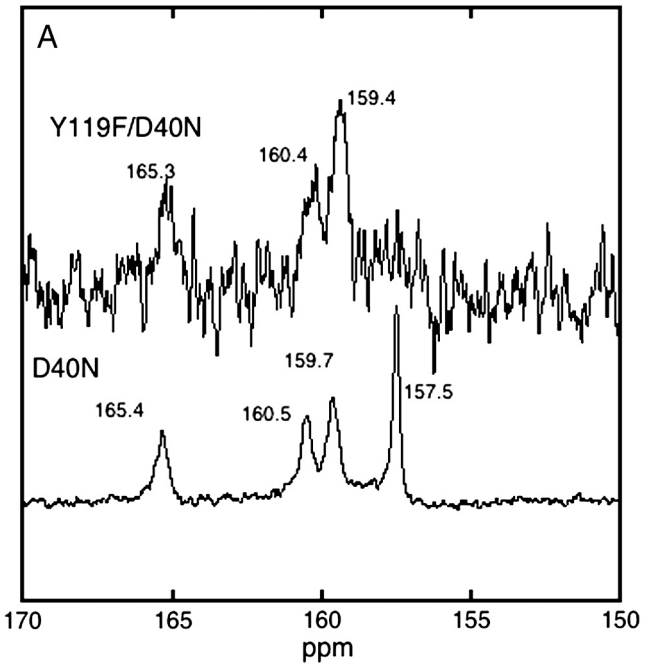


**Fig. S4.** $^{13}C_{\zeta}$-tyrosine labeled KSI NMR spectra of the Y119F/D40N mutant (top) compared to D40N (bottom, reproduced from Fig. 7).

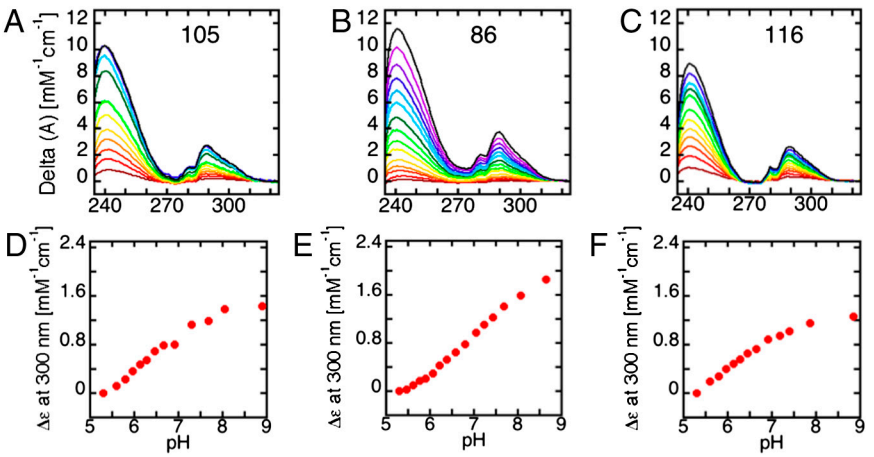


**Fig. S5.** UV-Vis difference-absorption spectra of probe-modified D40N KSI variants. (*A–C*) Difference spectra for pH values 5.3 to 9 (red to black) are calculated relative to the spectrum at pH 5.3, for M105C-CN, F86C-CN and M116C-CN, respectively. (*D–F*) The change in absorbance at the absorption maximum of tyrosinate, 293 nm, as a function of measured pH, for M105C-CN, F86C-CN and M116C-CN, respectively.

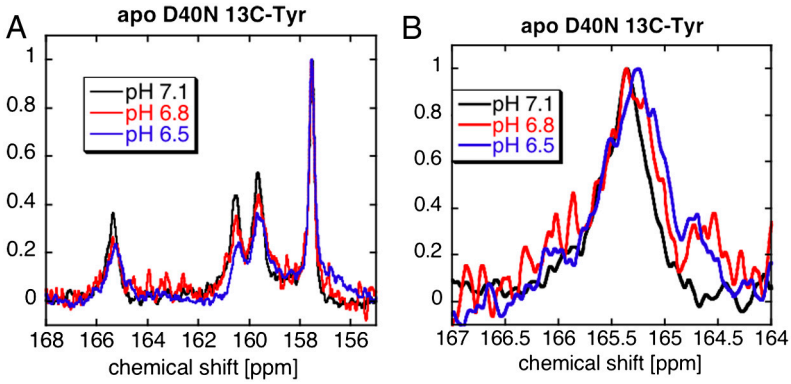


**Fig. S6.** $^{13}$C NMR spectra of KSI D40N with labeled tyrosines. (*A*) The effect of decreasing the pH. All spectra were taken on the same sample (buffer = 40 mM phosphate, no other additives; pH decreased with dropwise addition of 1 M $KH_2PO_4$). (*B*) Expanded view of downfield region of *A*, height-normalized to 1.

**Table S1. X-ray data collection and structure refinement statistics**

| Dataset | D40N/M116C-CN | D40N/F86C-CN | D40N/F86C-CN | D40N/M105C-CN |
|---|---|---|---|---|
| Ligand | – | – | Equilenin | Equilenin |
| PDB ID code | 3OXA | 3OX9 | 3OWU | 3OWY |
| Resolution range (Å) | 41.67–1.89 | 41.13–2.00 | 36.14–1.70 | 36.36–2.30 |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ | $P2_1$ |
| a, Å | 42.9 | 44.1 | 45.2 | 45.3 |
| b, Å | 93.6 | 129.8 | 130.3 | 129.4 |
| c, Å | 75.4 | 75.4 | 45.1 | 84.0 |
| α, ° | 90.0 | 90.0 | 90.0 | 90.0 |
| β, ° | 103.9 | 114.5 | 112.7 | 96.5 |
| γ, ° | 90.0 | 90.0 | 90.0 | 90.0 |
| No. unique reflections | 39,383 | 23,898 | 32,425 | 39,308 |
| Completeness, % | 85.0 | 76.4 | 61.6 | 92.4 |
| Multiplicity | 4.4 | 3.8 | 4.4 | 4.8 |
| $R_{merge}$, % * | 9.4 | 18.9 | 11.7 | 25.5 |
| $I/\sigma_{overall}$ ($I/\sigma_{highres}$) | 14.6 (1.5) | 6.7 (1.1) | 11.0 (1.0) | 6.0 (1.2) |
| Refinement statistics | | | | |
| No. residues | 130 | 129 | 129 | 129 |
| No. waters | 44 | 77 | 218 | 132 |
| $R_{work}$, % † | 24.4 | 26.5 | 21.6 | 28.0 |
| $R_{free}$, % ‡ | 29.0 | 29.0 | 27.0 | 34.6 |
| rmsd bond, Å | 0.005 | 0.009 | 0.005 | 0.004 |
| rmsd angle, ° | 0.62 | 1.08 | 1.06 | 0.85 |

*$R_{merge} = \sum_{hkl} \sum_i |I(hkl)_i - [I(hkl)]| / \sum_{hkl} \sum_i I(hkl)_i$

†$R_{work} = \sum_{hkl} |F(hkl)_o - [F(hkl)_c]| / \sum_{hkl} F(hkl)_o$

‡$R_{free}$ was calculated exactly as $R_{work}$ where $F(hkl)_o$ were taken from 5% of the data not included in refinement.

**Table S2. p$K_a$ predictions from web-based continuum electrostatic calculation engines**

| | M116A/D40N | | | | M116A/D40N/D103N | | |
|---|---|---|---|---|---|---|---|
| | Asp103 | Tyr16 | Tyr57 | Tyr32 | Tyr16 | Tyr57 | Tyr32 |
| H++ | 6.0 | 27 | 24 | 34 | 22 | 11 | 31 |
| PROPKA | 6.4 | 20 | 15 | 11 | 14 | 19 | 11 |
| Karlsberg | 7.8 | >20 | >20 | 18.9 | >20 | >20 | 12.7 |

**Table S3. Distances and angles between midpoint of nitrile bond and side chains that could undergo ionization changes**

| Atom position | | Dist [Å] * | Angle $\theta_{NMX}$ † | cos $\theta_{NMX}$† | Projected field [MV/cm] ‡ |
|---|---|---|---|---|---|
| ASP 103OD2 § | M116C-CN | 4.1 | 110.7 | −0.353 | −15.0 |
| | F86C-CN | 3.5 | 67.5 | 0.383 | 22.4 |
| | M105C-CN | 5.7 | 106.8 | −0.289 | −6.4 |
| ASN 40OD1 § | M116C-CN | 5.7 | 27.7 | 0.885 | 19.5 |
| | F86C-CN | 9.2 | 90.7 | −0.012 | −0.1 |
| | M105C-CN | 11.1 | 117.5 | −0.462 | −2.7 |
| EquileninO1 § | M116C-CN | 4.1 | 80.1 | 0.172 | 7.3 |
| | F86C-CN | 5.2 | 93.1 | −0.054 | −1.4 |
| | M105C-CN | 7.1 | 96.6 | −0.115 | −1.6 |
| Tyr 16OH § | M116C-CN | 4.5 | 89.1 | 0.016 | 0.6 |
| | F86C-CN | 7.3 | 86.8 | 0.056 | 0.8 |
| | M105C-CN | 6.5 | 84.9 | 0.089 | 1.5 |
| Tyr 57OH § | M116C-CN | 4.1 | 80.7 | 0.162 | 6.9 |
| | F86C-CN | 9.5 | 83.1 | 0.120 | 1.0 |
| | M105C-CN | 7.6 | 91.4 | −0.024 | −0.3 |
| Tyr 32OH § | M116C-CN | 4.3 | 110.2 | −0.345 | −13.3 |
| | F86C-CN | 10.2 | 69.2 | 0.355 | 2.4 |
| | M105C-CN | 6.3 | 100.7 | −0.186 | −3.4 |

*The distance is calculated between the location of the closest atom of the side-chain to the and the midpoint of the nitrile bond.

†The angle between the CN-bond axis and the location of the ionizing atom, X, measured from the midpoint of the nitrile bond (N–midpoint–X).

‡Fields according to Coulomb's law: $\frac{q}{4\pi\varepsilon\varepsilon_0 r^2}$, with $\epsilon = 2$, projected on the nitrile bond axis.

§The residue and atom names refer to the names given in the PDB structures.