Supporting Information

Sigala et al. 10.1073/pnas.1302191110

SI Text

A. SI Materials and Methods

1. Materials. Phenol and steroid ligands were the highest purity commercially available and were used without further purification. Steroids were purchased from Steraloids, and phenols were acquired as previously described (1). Phenol and steroid pK_a values were taken from a study by Kraut et al. (1), except for 5-androsten-3-ol-17-one (5-Andro) and 4-androsten-3-ol-17-one (4-Andro), whose aqueous pK_a values were calculated using the ACD/Labs I-Lab 2.0 pK_a Prediction Module.

2. Ketosteroid Isomerase Mutagenesis, Expression, and Purification. Ketosteroid isomerase (KSI) mutants were prepared using QuikChange (Stratagene) site-directed mutagenesis and confirmed by DNA sequencing. KSI was expressed from a pKK223-3 plasmid and purified from *Escherichia coli* BL21 bacteria as previously described (1). Nitrile incorporation was accomplished as previously reported (2–4). Protein purity was confirmed by a Coomassie-stained SDS/PAGE gel (>95% purity in all cases), and protein concentration was determined by absorbance using the calculated molar extinction coefficient of 16,960 $M^{-1}\cdot cm^{-1}$ (5).

3. ^{13}C-Tyr Labeling of Recombinant and Semisynthetic KSI. *Pseudomonas putida* KSI (pKSI) D40N containing $^{13}C_\zeta$-Tyr labels at all four Tyr residues was prepared as previously reported (4). Semisynthetic pKSI D40N (also containing the R15K/D21N/D24C mutations) containing $^{13}C_\zeta$-Tyr labels at Y32, Y57, and Y119 (Y16 unlabeled) was prepared as described in the following subsections. NMR spectra of recombinant unliganded D40N and D40N/R15K/D21N/D24C containing $^{13}C_\zeta$-Tyr labels at all four Tyr residues showed identical Tyr peak positions (Fig. S9).

3a. Construction of the His-tagged D24C-131 KSI plasmid. The sequence encoding residues 24–131 was PCR-amplified out of the pKK223-3 plasmid containing KSI using a forward primer containing an AscI site, followed by the KSI sequence starting at Ile25 and a reverse primer containing the terminal KSI sequence, a stop codon, and a PacI site. Following digestion with the appropriate restriction enzymes, this PCR product was cloned between the AscI and PacI sites of a vector containing a His_6 tag and the small ubiquitin-like modifier (SUMO) (a gift from Aaron Straight, Stanford University, Stanford, CA). QuikChange site-directed mutagenesis was used to mutate the residue at position 24 to a Cys, generating a His-tagged SUMO-D24C-131 construct. The product was confirmed by sequencing miniprep DNA from DH5α cells.

3b. Peptide synthesis and purification. A peptide comprising the N-terminal 23 amino acids of KSI with a C-terminal thioester for ligation was synthesized manually on β-mercaptopropionyl-Leu-phenylacetamidomethyl resin using tert-butoxycarbonyl (BOC) in situ neutralization protocols (6, 7). The peptide was deprotected using trifluoromethanesulfonic acid and thioanisole (6, 7). The peptide was purified by reverse-phase HPLC using a gradient elution between A (water, 0.1% TFA) and B (9:1 acetonitrile/water, 0.09% TFA). Fractions containing the peptide product were pooled and lyophilized. The product mass was confirmed by electrospray MS.

3c. Expression and purification of the recombinant fragment containing an N-terminal Cys and $^{13}C_\zeta$-Tyr labeled at Y32, Y57, and Y119. The ^{13}C-Tyr–labeled fusion protein with Y32, Y57, and Y119 $^{13}C_\zeta$-labeled was expressed in BL21(DE3) cells grown in M9 minimal media supplemented with L-Tyr [50 mg/L phenol-4-^{13}C (95–99%); Cambridge Isotope Laboratories, Inc.] and the remaining 19 unlabeled amino acids. Cells were grown at 37 °C to an OD of ∼0.6, followed by the addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside and a further 10 h of growth at 25 °C. Cells were harvested and resuspended in 20 mM sodium phosphate (pH 7.2) and 150 mM NaCl (lysis buffer), and were then lysed by passage through a French pressure cell. Inclusion bodies containing the fusion protein were isolated by solubilization of membranes by addition of 1% Triton X-100, 20 mM sodium phosphate (pH 7.2), and 150 mM NaCl, followed by centrifugation at 8,000 × *g*. The inclusion bodies were then washed several times by resuspension in 20 mM sodium phosphate (pH 7.2) and 150 mM NaCl to remove detergent.

Inclusion bodies were resolubilized in 7 M urea, 20 mM sodium phosphate (pH 7.2), and 150 mM NaCl. The samples were centrifuged to remove aggregated protein. The supernatant was loaded on a nickel–nitriloacetic acid column preequilibrated with 7 M urea, 20 mM sodium phosphate (pH 7.2), and 150 mM NaCl. The column was washed with 7 M urea, 20 mM sodium phosphate (pH 7.2), and 150 mM NaCl until the A_{280} dropped to ∼0 (∼10 column volumes). The product was eluted in one step using 250 mM imidazole, 7 M urea, 20 mM sodium phosphate (pH 7.2), and 150 mM NaCl.

The eluted material was diluted at 4 °C by drop-wise addition in 20 mM sodium phosphate (pH 7.2) and 150 mM NaCl to a final urea concentration of 2 M to allow refolding of the SUMO protein. The material was concentrated using an Amicon centrifugal filter unit and was then buffer-exchanged by passing through a HiPrep 26/10 (GE Healthcare Life Sciences) desalting column preequilibrated with 2 M urea, 50 mM Tris·HCl (pH 8.0), and 150 mM NaCl. The purity of the fusion protein was >95%, as determined by SDS/PAGE.

SUMO protease (1 mg of protease per 100 mg of fusion protein) was added to cleave the fusion protein. The cleavage reaction was carried out in 2 M urea, 50 mM Tris·HCl (pH 8.0), 150 mM NaCl, and 2 mM DTT at 30 °C for 2 h. To minimize aggregation of the cleaved products, the concentration of the fusion protein in the cleavage reaction was below 100 μM. Cleavage efficiency was typically >95%, as determined by SDS/PAGE. Following the reaction, solid urea was added directly to the mixture to a final concentration of 8 M in the reaction mixture. The mixture was centrifuged to remove aggregated material. Cleavage products were purified by loading the mixture on a Superose-12 gel filtration column preequilibrated with 7 M urea and 20 mM sodium phosphate (pH 7.2). Fractions containing the KSI fragment were identified by SDS/PAGE, pooled, concentrated to a final concentration of ∼2 mM using a 3-kDa cutoff centrifugal filter unit, and stored at 4 °C. The KSI fragment and the cleaved SUMO tag, which were removed in a subsequent step as described below, comprised >98% of the protein in the final concentrate as measured by Coomassie-stained SDS/PAGE.

3d. Native chemical ligation. The peptide containing a C-terminal thioester was ligated to the $^{13}C_\zeta$-Tyr–labeled recombinant fragment containing an N-terminal Cys using native chemical ligation (6). The lyophilized peptide was dissolved in 7 M urea and 20 mM sodium phosphate (pH 7.2) to a concentration of ∼4 mM. The peptide and recombinant fragments were combined to give final concentrations of ∼4 mM and ∼2 mM, respectively, in 7 M urea and 20 mM sodium phosphate (pH 7.2). Sodium 4-mercaptophenylacetic acid was added to a final concentration of 1 M (8). The ligation was allowed to proceed for 2 h at 25 °C. Protein in the ligation mixture was then refolded by a 20-fold dilution into 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, and 2 mM DTT, followed by stirring for 1 h at 4 °C. The refolded protein was

purified by deoxycholate affinity chromatography as typically performed for recombinant KSI (1), followed by buffer exchange into 40 mM potassium phosphate (pH 7.2),1 mM EDTA, and 2 mM DTT in a 10-kDa cutoff concentrator. Final purity was >99% on a Coomassie-stained SDS/PAGE gel. Protein concentration was determined using a calculated molar extinction coefficient at 280 nm. The yield relative to the limiting recombinant fragment was ~40%, and 4.2 mg of pure KSI was recovered.

4. NMR Spectroscopy. ^{19}F and ^{13}C NMR spectra were acquired at 25 °C on 500 and 600 MHz (proton frequency) Varian ^{UNITY}INOVA spectrometers using sample conditions and acquisition parameters as previously reported (4, 9, 10). Samples for ^{13}C detection contained 1 mM KSI, 2 mM ligand (when present), 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% D_2O as the lock solvent. Samples for ^{19}F detection contained 1.5 mM KSI, 1 mM ligand, 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% D_2O as the lock solvent. ^{19}F NMR chemical shifts were referenced to an external sample of TFA in buffer (−76.1 ppm relative to $CFCl_3$). ^{13}C NMR spectra were referenced to an external sample of sodium-3-trimethylsilylpropionate-2,2,3,3-d_4 in buffer (0 ppm). Spectra were processed using a 10-Hz line broadening and baseline correction applied over the peaks of interest. Peak positions are reported as the chemical shift value corresponding to the position of maximum intensity for each peak. Uncertainty in chemical shift measurements is estimated as ±0.1 ppm based on replicate measurements of identical samples.

For all phenol complexes studied, the limiting component (protein in ^{13}C spectra or ligand in ^{19}F spectra) was ≥95% bound, based on the sample concentrations above and the low micromolar K_d values observed for binding of substituted phenols to D40N (1). Under these conditions, the reported NMR spectra therefore reflect the properties of fully or nearly fully (≥95%) bound KSI (^{13}C spectra) or phenol (^{19}F NMR), and any chemical exchange with the minor population (≤5%) of the dissociated species does not contribute significantly to the observed peak positions. This conclusion is supported by observation of identical ^{13}C-Tyr peak positions in spectra of KSI–phenol complexes acquired with differing excess phenol concentrations, ruling out a significant contribution to peak position from ligand exchange. As discussed in *SI Text*, section D, our observation of a single ^{13}C NMR peak for each labeled Tyr residue indicates that proton transfers within the KSI–phenol complexes are much faster than the NMR time scale (defined below) and that the observed peak positions are the population-weighted average of all ionization states present for a given residue. Additional discussion of chemical exchange contributions to NMR peak positions can be found elsewhere (11, 12). NMR spectra of KSI–phenol complexes were observed to be pH-independent (pH 7–8), as expected for the effectively intramolecular nature of the proton transfer between the bound phenol and KSI oxyanion hole residues.

5. IR Spectroscopy. Room temperature and 80-K IR spectra for phenols bound to nitrile-modified KSI (KSI-CN) variants were acquired in the CN stretching frequency region using sample conditions and experimental parameters as previously described (2–4). The 80-K spectrum of 4-F-phenol in the stretching region from C to F required selective deuteration of the phenol (2,6-d_2-4-F-phenol) to disrupt complicating Fermi resonances, and the spectra were acquired in a CaF_2 sample cell with 12- to 23-μM spacers as previously described (11).

6. X-Ray Crystallography. Cocrystals of pKSI D40N–3-F-4-NO_2-phenolate and D40N/M116C-CN–equilenin were obtained at 20 °C using hanging drop vapor diffusion according to previously published crystallization conditions in ammonium sulfate and potassium phosphate buffer (1, 4, 9, 13). X-ray diffraction data for D40N–3-F-4-NO_2-phenolate and D40N/M116C-CN–equilenin were collected at the Stanford Synchrotron Radiation Laboratory and the Advanced Light Source (Lawrence Berkeley National Laboratory), respectively, and structure refinement was carried out as previously described (1, 4, 9). Structural coordinates and structure factors were deposited with the Protein Data Bank (PDB) under the ID codes 3VGN (D40N–3-F-4-NO_2-phenolate) and 3OWS (D40N/M116C-CN–equilenin). Structure figures were prepared using MacPyMOL (14).

7. Quantitative Modeling of Active Site Protonation States as a Function of Phenol pK_a. Fractional ionizations of the bound phenol, Y16, and Y57 were calculated as shown in Fig. 4*A*, based on the following assumptions: (*i*) These three groups are the only species that significantly ionize in the KSI active site; (*ii*) the sum of the fractional ionizations of these three groups at each phenol pK_a value is equal to 1; (*iii*) the chemical shifts observed for Y16 and Y57 in the KSI complex with 4-nitrophenol (pK_a = 7.1), which is bound as the fully ionized phenolate (1, 11, 15) (Fig. 2*A*), represent the chemical shifts for fully neutral Y16 and Y57 within the phenol-bound KSI active site; and (*iv*) the chemical shift of fully ionized Y16 and Y57 is 166 ppm, the chemical shift of a fully ionized Tyr within a peptide in basic aqueous solution (16).

The fractional ionizations of Y16 and Y57 at discrete phenol pK_a values ($X_i^{Y^-}$) were calculated as $X_i^{Y^-} = \frac{\delta_i - \delta_{7.1}}{166 - \delta_{7.1}}$, where δ_i is the observed chemical shift at each pK_a value and $\delta_{7.1}$ is the chemical shift with bound 4-nitrophenol. The fractional phenol ionization at each pK_a value ($X_i^{Ph^-}$) was calculated as $X_i^{Ph^-} = 1 - (X_i^{Y16^-} + X_i^{Y57^-})$. For p$K_a$ values ≥10.4, where $X_i^{Y16^-} + X_i^{Y57^-}$ calculated as above gave a sum that was greater than 1, $X_i^{Ph^-}$ was assumed to be 0 and the values of $X_i^{Y16^-}$ and $X_i^{Y57^-}$ were linearly scaled to make $X_i^{Y16^-} + X_i^{Y57^-}$ equal to 1. The $X_i^{Ph^-}$ values determined indirectly in this fashion were very similar to those previously reported based on direct ligand measurements for phenols (11), naphthols (17), and equilenin (17) bound to pKSI D40N (Fig. S3) and are qualitatively consistent with the ^{19}F NMR changes observed for bound 4-F–substituted phenols in Fig. 2*A* and the observation of a single C-F peak in a low-temperature (80 K) spectrum of 2,6-d_2-4-F-phenol bound to D40N/M116C-CN (Fig. S7*B*). We estimate the uncertainty in our fractional ionization estimates for each point in Fig. 4*A* to be ±0.15, based on the 2.5- to 3-ppm difference in the ^{13}C chemical shift of neutral Y16 (158.5 ppm) and Y57 (158 ppm) within the KSI–4-nitrophenol complex vs. a neutral Tyr residue within a peptide in aqueous solution (155.5 ppm) divided by the 11-ppm ^{13}C chemical shift dispersion between a neutral and ionized Tyr in aqueous solution (16). We note that the central conclusions of the paper, which are based on the systematic changes in fractional ionizations across the series of phenols rather than the absolute values of fractional ionizations for any given complex, are unaffected by this uncertainty.

The observed changes in fractional ionization for the phenol, Y16, and Y57 as a function of phenol pK_a were globally fit by nonlinear regression using GraphPad Prism to the following models derived from the proton transfer equilibria between these groups (A = phenol, B = Y16, and C = Y57, where K_a is the apparent acid dissociation constant for each group and X_i is the fractional ionization of each group). As noted in the main text, the apparent pK_a values used in these models are a proxy for the relative proton affinities of the phenol, Y16, and Y57 within KSI–phenol complexes and do not represent the actual pK_a values for these groups within the phenol-bound KSI active site:

$$A^- \bullet\bullet\bullet H\text{-}B \bullet\bullet\bullet H\text{-}C \overset{K_1}{=\!=} A\text{-}H \bullet\bullet\bullet B^- \bullet\bullet\bullet H\text{-}C \overset{K_2}{=\!=} A\text{-}H \bullet\bullet\bullet B\text{-}H \bullet\bullet\bullet C^-$$

$$K_1 = \frac{[B^-]}{[A^-]} = \frac{K_a^B}{K_a^A} = 10^{\left(pK_a^A - pK_a^B\right)} \quad \textbf{[S1]}$$

$$K_2 = \frac{[C^-]}{[B^-]} = \frac{K_a^C}{K_a^B} = 10^{\left(pK_a^B - pK_a^C\right)} \quad \textbf{[S2]}$$

$$K_1 \times K_2 = K_3 = \frac{[C^-]}{[A^-]} = \frac{K_a^C}{K_a^A} = 10^{\left(pK_a^A - pK_a^C\right)} \quad \textbf{[S3]}$$

$$X_{A^-} + X_{B^-} + X_{C^-} = 1 \quad \textbf{[S4]}$$

$$X_{A^-} = \frac{[A^-]}{[A^-]+[B^-]+[C^-]} = \frac{1}{1 + \frac{[B^-]}{[A^-]} + \frac{[C^-]}{[A^-]}} = \frac{1}{1 + 10^{\left(pK_a^A - pK_a^B\right)} + 10^{\left(pK_a^A - pK_a^C\right)}} \quad \textbf{[S5]}$$

$$X_{B^-} = \frac{[B^-]}{[A^-]+[B^-]+[C^-]} = \frac{1}{1 + \frac{[A^-]}{[B^-]} + \frac{[C^-]}{[B^-]}} = \frac{1}{1 + 10^{\left(pK_a^B - pK_a^A\right)} + 10^{\left(pK_a^B - pK_a^C\right)}} \quad \textbf{[S6]}$$

$$X_{C^-} = \frac{[C^-]}{[A^-]+[B^-]+[C^-]} = \frac{1}{1 + \frac{[A^-]}{[C^-]} + \frac{[B^-]}{[C^-]}} = \frac{1}{1 + 10^{\left(pK_a^C - pK_a^A\right)} + 10^{\left(pK_a^C - pK_a^B\right)}} \quad \textbf{[S7]}$$

We first fit the data in Fig. 4*A* using expressions **S5**–**S7**, which treat pK_a^A as an independent variable and pK_a^B and pK_a^C as adjustable parameters. This model assumes constant proton affinities for B (Y16) and C (Y57) regardless of the pK_a value of A (the bound phenol). Global fits of this model to the data (Fig. S4*A*) failed to account for the decrease in X_{Y16} and increase in X_{Y57} observed above a pK_a of 10. Rather, this model predicts that constant Y16 and Y57 proton affinities will result in constant X_{Y16} and X_{Y57} values in this region, which is not what was observed (Fig. S4*B*).

Based on prior linear free energy studies and known physical properties of hydrogen bonds (1, 18–24), we posited that energetic changes in the phenol–Y16 hydrogen bond with increasing phenol pK_a would alter the stability of the Y16 anion, and thus modulate its ability to ionize relative to Y57. To account for this behavior, we next fit the data with modified expressions in which we allowed the apparent pK_a of Y16 to vary linearly with changes in phenol pK_a according to the equation $pK_a^B = a + b \times pK_a^A$, in which a and b were adjustable parameters. These modified expressions, shown below, were globally fit to the data as shown in Fig. 4*A* and accurately accounted for the observed changes:

$$X_{A^-} = \frac{1}{1 + 10^{\left(pK_a^A - \left(a + b \times pK_a^A\right)\right)} + 10^{\left(pK_a^A - pK_a^C\right)}} \quad \textbf{[S8]}$$

$$X_{B^-} = \frac{1}{1 + 10^{\left(\left(a + b \times pK_a^A\right) - pK_a^A\right)} + 10^{\left(\left(a + b \times pK_a^A\right) - pK_a^C\right)}} \quad \textbf{[S9]}$$

$$X_{C^-} = \frac{1}{1 + 10^{\left(pK_a^C - pK_a^A\right)} + 10^{\left(pK_a^C - \left(a + b \times pK_a^A\right)\right)}} \quad \textbf{[S10]}$$

8. Electrostatic Modeling. Theoretical calculations of electrostatic fields within the KSI active site were carried out using DelPhi (25, 26), a finite difference algorithm for solving the Poisson–Boltzmann equation. Calculations were performed using the protein heavy atoms positions determined in the X-ray structures of the equilenin-bound D40N mutants of M116C-CN, M105C-CN, and F86C-CN (respective PDB ID codes 3OWS, 3OWY, and 3OWU). Heavy atom positions for bound ligands [4-NO_2-phenolate, 3,4-$(NO_2)_2$-phenolate and unsubstituted phenolate] were modeled using the equilenin *A* ring in the above structures as a scaffold to which the relevant phenolate functional groups were added by comparison with the X-ray structure of D40N with bound 3-F-4-NO_2-phenolate (PDB ID code 3VGN).

Charge and radii parameters for protein atoms were taken from the PARSE parameter set (25), which treats only heavy atoms and hydrogen atoms attached to heteroatoms explicitly. Charge and radii parameters for the nonnatural Cys-CN amino acid were taken from a previous study (27). A dielectric value (ε) of 80 was used for the protein exterior, and a value of 2 was assigned to the protein interior to account for the intrinsic polarizability of the constituent chemical groups of the protein (28) and based on previous studies of this enzyme in which this value provided the best fit to the data (4). Ligand partial charges were added based on a quantum chemical calculation of substituted phenolate anions in the gas phase using Gaussian '03, with Mulliken population analysis (29). The calculated Mulliken partial charges on the phenolate oxygen were −0.51, −0.43, and −0.37 for phenolate, 4-nitrophenolate, and 3,4-dinitrophenolate, respectively. This decrease in calculated phenolate oxygen partial charge with decreasing solution pK_a value is consistent with prior published calculations of phenolate charges (30, 31).

Hydrogen atoms were automatically added to the protein heteroatoms using the program PDB2GMX (32), which uses database values for bond distances and angles, hydrogen bond satisfaction criteria, and unperturbed pK_a values for protein groups. Calculations were performed using three different protein protonation states, with anionic charge residing on the hydroxylic oxygen of the bound phenol, Y16, or Y57. To avoid artifacts from the finite grid treatment, a seven-tier focusing routine was used. A calculation of the potential on a 50 × 50 × 50 grid of 0.5 grid lines per angstrom, centered around the CN bond and encompassing the whole protein, provided the initial boundary conditions for a subsequent calculation at twofold finer grid spacing. This process was repeated until the grid spacing was 32 lines per angstrom. 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.

To convert calculated electric fields into nitrile IR frequencies (in inverse cm, cm^{-1}), the 4-nitrophenol complex (pK_a = 7.1) was used as a reference point. We calculated the change in electric field (in megavolts per centimeter, MV/cm) along each nitrile relative to the calculated electric field for the 4-nitrophenol complex. We then divided this number by the previously determined vibrational Stark tuning rate of 0.65 cm^{-1}/(MV/cm) (4) to convert the electric field changes into vibrational frequency changes (in cm^{-1}). To calculate the predicted position of the IR peak of a given tautomer, the calculated vibrational frequency change was added to the experimental IR frequency for each probe with bound 4-nitrophenol. To calculate the predicted peak frequency in cases in which multiple tautomers coexist, Gaussian curves centered at the calculated frequency for each tautomer (with FWHM line widths equal to the experimental line widths with bound 4-nitrophenol) were assigned a relative weight according to the fractions in Fig. 4*A* and summed together as in Fig. 8*A*. The maximum value of the resulting composite peak was reported as the predicted peak position. An overlay of the experimental and calculated IR frequencies for each probe is shown in Fig. 8*B*.

9. Modeling of Nitrile IR Peak Positions for the Series of KSI-CN–Phenol Complexes. The observed increase in M116C-CN IR frequency with increasing phenol pK_a in the region with low pK_a values (5–8), where the phenol is bound exclusively as the ionized phenolate (Fig. 4*A*), led us to hypothesize that increased negative charge localization on the phenolate oxygen with increased solution pK_a was responsible for the electric field increase at the nitrile. To test this hypothesis, we performed electrostatic calculations for enzyme complexes with 3,4-dinitrophenol (pK_a = 5.4), 4-nitrophenol (pK_a = 7.1), and unsubstituted phenol (pK_a = 10.0) using the quantum mechanics (QM)-derived partial charges for ligand atoms described above. These calculations predicted IR shifts for M116C-CN of +3.4, +4.3, and +4.2 cm^{-1} with each of the above phenolates, respectively. Although caution is warranted because only three points are calculated and these points do not comprise a monotonically increasing series, a line of best fit to these points has a slope of 0.2 cm^{-1} per pK_a ($R^2 = 0.55$), which is similar to the experimental slope of 0.3 cm^{-1} per pK_a in the low pK_a regime, where the ligand is fully ionized. This same procedure was carried out for F86C-CN and M105C-CN. We calculated a slope of −0.3 cm^{-1} per pK_a ($R^2 = 0.66$) for F86C-CN, which is in poor agreement with the experimental data (further discussion on this probe is provided below), and a slope of −0.009 cm^{-1} per pK_a ($R^2 = 0.99$) for M105C-CN, which is consistent with the data that showed no detectable change over the series. The decreasing fraction of ionized phenolate at higher pK_a values limits the impact of this parametrization mainly to the calculated values in the pK_a 5–8 regime.

The line shape analysis and low-temperature IR data presented in the main text indicate that exchange between tautomeric states due to proton transfer was slow on the IR time scale, such that at least two tautomers exist as distinct states with overlapping IR spectra (Fig. 7 and analysis in the main text). To calculate IR shifts for each KSI-CN–phenol complex, the IR peak for each possible tautomer was modeled as a single Gaussian-shaped peak of line width equal to that observed for the KSI-CN–4-nitrophenolate complex. A numerical sum of the three peaks at each pK_a was calculated for each probe, and the maximum of this aggregate peak was reported.

10. Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed as previously described (27) using the MD program GROMACS 3.3.1 (32, 33), with the AMBER-99 force field ported to GROMACS (34). The nitrile-derivatized Cys residue and equilenin were parameterized using Antechamber and Leap from the AMBER 9 software suite (35), using the GAFF atom force field extension to AMBER (36). Hydrogens were added using the utility PDB2GMX as described above. For each different position of the nitrile, one monomer of KSI was simulated in explicit water using the SPC/E model. Simulations were equilibrated through 20 ps of energy minimization, followed by 20 ps of heavy atom position-restrained refinement that was monitored for convergence. Two nanoseconds of MD were simulated under simultaneously fixed temperature, using a Nose–Hoover thermostat (298 K), and fixed pressure (1 atm), using a Parinello–Raman barostat. The particle mesh Ewald model for the calculation of long-range electrostatics was used with a 1-nm cutoff. The distribution of values taken on by the dihedral angle describing the orientation of the nitrile relative to the protein backbone ($\angle C_\alpha$-C_β-S-CN) was calculated over the whole trajectory and plotted in Fig. S8. Additionally, the autocorrelation function, the probability that dihedral angles measured at times t_1 and t_2, separated by an interval Δt, will have the same value, was calculated using the GROMACS utility "g_angle" and plotted as a function of Δt in Fig. 8*C*. The apoprotein was simulated for M116C-CN, M105C-CN, and F86C-CN. For F86C-CN, significant dihedral motion was observed; thus, we additionally simulated the effect of the bound ligand equilenin on the mobility of this probe. Additional discussion of the MD simulations is provided in *SI Text*, section E.

B. Discussion of ^{13}C-Tyr Chemical Shift Changes Due to Hydrogen Bond Shortening and Ionization

Prior ^{1}H NMR and quantum mechanical/molecular mechanics (QM/MM) studies of phenols bound to pKSI D40N strongly suggest that the oxyanion hole hydrogen bonds formed by D103 and Y16 to the phenol oxygen shorten progressively with increasing phenol pK_a (1, 37). For a hydrogen bond formed between a Tyr residue and an ionized phenolate, polarization of the Tyr O-H bond accompanying hydrogen bond shortening can result in deshielding of the C_ζ-Tyr nucleus. Such deshielding can be ∼1 ppm in magnitude, based on ^{13}C NMR studies of hydrogen-bonded complexes between phenols and trimethylamine-*N*-oxide (38), but the large peak shifts we observed for Y16 and Y57 to values ≥160 ppm cannot be accounted for by O-H bond polarization alone. This observation suggests ionization of Y16 at an intermediate phenol pK_a and ionization of Y57 at a high phenol pK_a, consistent with the other measures of phenol ionization state described in the main text and presented in Fig. 2*A* and Fig. S3. The ∼1-ppm down-field shifts for the ^{13}C peak of Y32 with bound 4-Andro and 5-Andro (Fig. 3*D*) are consistent with polarization of the Y32 O-H bond on ionization of Y57, and our analysis assumes that Y32 does not ionize (see above).

C. Contributions to Nitrile Peak Shifts for Phenols Bound to KSI-CN Variants

Binding of the fully ionized 3,4-dinitrophenol (pK_a = 5.4) results in a dispersion of 13 cm^{-1} in the stretch frequency observed across the three nitrile probe sites (Fig. 6*A*). As elucidated previously (3, 4), this dispersion is the result of two superimposed contributions to relative peak position: (*i*) a large and constant offset in stretch frequency imparted by the variable hydrogen bonding state of the three probes (the nitriles of F86C-CN and M116C-CN accept hydrogen bonds from a backbone amide and water, respectively, whereas the M105C-CN nitrile is free of hydrogen bonding) and (*ii*) a smaller and unique shift in the stretch frequency of each nitrile due to the differing projection of the local electrostatic field on the unique position and orientation of each probe. Because the hydrogen bonding state of each probe and its associated contribution to the IR stretch frequency remain nearly constant across a series of bound ligands (based on tandem IR and ^{13}C NMR measurements of each nitrile) (3) (Fig. S10), changes in the IR frequency for each nitrile across the series of bound phenols report on changes in the local electrostatic field experienced by each probe due to charge rearrangements within the hydrogen bond network with increasing phenol pK_a.

As an independent test of the sensitivity of the M116C-CN nitrile to charge and proton transfers within the active site, we acquired ^{13}C NMR spectra of this variant bearing a ^{13}C-labeled nitrile (−^{13}CN) bound to the same series of phenols. The ^{13}C chemical shift of the nitrile changes little from a pK_a of 5–8 but increases steadily above a pK_a of 8 (Fig. S11). This inflection is similar to the observed inflection in M116C-CN nitrile stretch frequency (Fig. 6*B*), suggesting that both IR frequency and NMR chemical shift respond to the same active site charge rearrangements that accompany increases in phenol pK_a in this region. Nevertheless, there are differences in the observed dependence of the nitrile IR and ^{13}C NMR frequencies on phenol pK_a. Understanding the physical origin of these differences will require a deeper understanding of the contributions to ^{13}C NMR chemical shift within the KSI active site.

D. Quantitative Limit on Proton Transfer Rate for pK_a-Matched Hydrogen Bonds Within KSI

Our ability to resolve distinct nitrile stretch peaks (Fig. 7) by IR spectroscopy at both low temperature (independently resolved peaks) and room temperature (resolvable peak shoulders) for the individual tautomers of ionized Y16 and ionized Y57 with phenols of pK_a near 10 bound to M116C-CN (Fig. 4) allows us to place limits on the time scale of the proton transfer that results in

interconversion of these tautomers. This limit can be derived from the time-energy uncertainty relation $\Delta E \Delta t > h$, where ΔE is the energy difference between two states, Δt is the approximate time for interconversion between the two states, and h is Planck's constant. In frequency terms, this expression has the form $\Delta \nu \Delta t > 1$. The two distinct states of KSI under study here, assigned to the different electrostatic environments of the M116C-CN probe in which negative charge resides on either Y16 or Y57, have associated M116C-CN nitrile stretching transitions with a difference of 4 cm^{-1} or 1×10^{11} Hz in energy or frequency, respectively. From this frequency difference and a model in which the two ionized tautomers are interconverting via proton transfer, a minimum proton residence time of at least 10 ps can be estimated. A proton transfer rate faster than 10^{11} Hz would be expected to result in a single average IR peak rather than the discrete peaks that we observe.

Our observation of only single population-averaged ^{13}C-Tyr NMR peaks for Y16 and Y57 with bound phenols of pK_a of ~10 (Fig. 3), rather than observation of discrete peaks for the interconverting neutral and ionized states of both Y16 and Y57 populated at this pK_a (Fig. 4*B*), indicates an exchange rate that is fast relative to the NMR-measured frequency difference between these two ionization states. A lower limit on the rate of proton transfer between Y16 and Y57 can be estimated from the ~8-ppm chemical shift difference between a neutral (157 ppm) and ionized (165 ppm) Tyr residue measured on a 125-MHz (^{13}C frequency) NMR spectrometer, which corresponds to a frequency difference of 10^3 Hz and a proton residence time of 1 ms. Based on these quantitative limits, we conclude that proton transfer between Y16 and Y57 within the D40N active site with a bound phenol of pK_a near 10 occurs faster than 10^3 s^{-1} but slower than 10^{11} s^{-1}.

E. Additional Discussion of MD Simulations

The dihedral angle $\angle C_\alpha$-C_β-S-C, calculated for MD trajectories of M105C-CN and M116C-CN, showed a narrow range of sampled values (Fig. S8), with time-dependent autocorrelation values near unity (Fig. 8*C*) over the 2 ns of the simulation. In other words, the S-CN group at these positions did not rotate significantly relative to the peptide backbone. F86C-CN, however, behaved differently. The dihedral autocorrelation decayed rapidly, indicating low barriers to rotation, and the angles sampled by the probe were widely distributed. This qualitatively different behavior of the F86C-CN nitrile relative to the other two probe positions in trajectories of the apoproteins led us to question whether the dynamic behavior of F86C-CN would be more restricted in the ligand-bound state. To test whether the position of the probe was more restricted in the presence of a bound ligand, we simulated F86C-CN with the phenolic steroid equilenin. The mobility of the S-CN group was significantly restricted with bound equilenin relative to the apoprotein but remained substantially more mobile than observed for the nitriles of M105C-CN and M116C-CN in the unliganded state (Fig. 8*C* and Fig. S8).

In Fig. 8*C*, the C_α-C_β-S-CN dihedral angle autocorrelation of F86C-CN with bound equilenin decayed to 0.6 in 2 ns, suggesting that the barriers to rotation of this group are spontaneously crossed; however, such events are sufficiently rare that major reorientations are infrequently observed on the hundreds of picosecond time scale sampled. These MD observations suggest a physical basis for understanding the differential behavior of the F86C-CN vs. M105C-CN and M116C-CN nitriles observed by IR spectroscopy (Fig. 6) and by modeling (Fig. 8), where the F86C-CN nitrile is the only probe that did not show a systematic change in probe frequency with changing ligand pK_a. The observation of spontaneous structural rearrangement of F86C-CN in MD simulations suggests that this probe has enhanced conformational mobility relative to the other two probes and can structurally rearrange to adopt a new average equilibrium conformation in response to an energetic perturbation.

Based on its heightened conformational mobility and our observation of limited sensitivity of the F86C-CN nitrile to electrostatic perturbations due to varying phenol pK_a, we propose that this probe conformationally reorients on binding of distinct phenols to minimize the interaction energy between the electrostatic field associated with each bound phenol and the ground state dipole moment of the F86C-CN nitrile. This reorientation may reduce the magnitude of the field change experienced by the F86C-CN nitrile to below the limit of detection by IR spectroscopy, accounting for the relative invariance in the measured IR peak position of this nitrile with increasing phenol pK_a. This effect would not have been captured by our point charge or DelPhi electrostatic models, which assumed a rigid probe position based on the F86C-CN crystal structure, and can thus account for the different behaviors of the F86C-CN observed by experiment vs. computation.

These results are consistent with a previous time-resolved IR study that showed no rearrangement of F86C-CN under the influence of a near-instantaneous energetic perturbation on the tens of picoseconds time scale (39). Our MD simulations, however, suggest that spontaneous rearrangement is possible on the hundreds of picoseconds time scale.

1. Kraut DA, et al. (2006) Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4(4):e99.
2. 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(40):12104–12105.
3. 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(37):12811–12813.
4. Fafarman AT, et al. (2012) Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase. *Proc Natl Acad Sci USA* 109(6):E299–E308.
5. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182(2):319–326.
6. Kent SB (1988) Chemical synthesis of peptides and proteins. *Annu Rev Biochem* 57: 957–989.
7. Hackeng TM, Griffin JH, Dawson PE (1999) Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc Natl Acad Sci USA* 96(18):10068–10073.
8. Johnson EC, Kent SB (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 128(20):6640–6646.
9. Sigala PA, et al. (2008) Testing geometrical discrimination within an enzyme active site: Constrained hydrogen bonding in the ketosteroid isomerase oxyanion hole. *J Am Chem Soc* 130(41):13696–13708.
10. Bagchi S, Fried SD, Boxer SG (2012) A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields. *J Am Chem Soc* 134(25):10373–10376.
11. Fried SD, Boxer SG (2012) Evaluation of the energetics of the concerted acid-base mechanism in enzymatic catalysis: The case of ketosteroid isomerase. *J Phys Chem B* 116(1):690–697.
12. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1996) *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego).
13. Kraut DA, Sigala PA, Fenn TD, Herschlag D (2010) Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole. *Proc Natl Acad Sci USA* 107(5):1960–1965.
14. DeLano WL (2007) *MacPyMOL: A PyMOL-Based Molecular Graphics Application for MacOS X* (DeLano Scientific LLC, Palo Alto, CA).
15. Petrounia IP, Pollack RM (1998) Substituent effects on the binding of phenols to the D38N mutant of 3-oxo-delta5-steroid isomerase. A probe for the nature of hydrogen bonding to the intermediate. *Biochemistry* 37(2):700–705.
16. Richarz R, Wuthrich K (1978) Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* 17(9):2133–2141.
17. Childs W, Boxer SG (2010) Proton affinity of the oxyanion hole in the active site of ketosteroid isomerase. *Biochemistry* 49(12):2725–2731.
18. Hibbert F, Emsley J (1990) Hydrogen bonding and chemical reactivity. *Adv Phys Org Chem* 26:255–379.
19. Jeffrey GA (1997) *An Introduction to Hydrogen Bonding* (Oxford Univ Press, New York).
20. Shan SO, Herschlag D (1996) The change in hydrogen bond strength accompanying charge rearrangement: Implications for enzymatic catalysis. *Proc Natl Acad Sci USA* 93(25):14474–14479.
21. Steiner T, Saenger W (1994) Lengthening of the covalent O-H bond in O-H-O hydrogen bonds re-examined from low-temperature neutron diffraction data of organic compounds. *Acta Crystallogr B* 50:348–357.
22. Shan SO, Herschlag D (1999) Hydrogen bonding in enzymatic catalysis: Analysis of energetic contributions. *Methods Enzymol* 308:246–276.
23. Stahl N, Jencks WP (1986) Hydrogen bonding between solutes in aqueous solution. *J Am Chem Soc* 108(14):4196–4205.

24. Abraham MH, Grellier PL, Prior DV, Morris JJ, Taylor PJ (1990) Hydrogen bonding. Part 10. A scale of solute hydrogen-bond basicity using log K values for complexation in tetrachloromethane. *J Chem Soc Perkin Trans 2:* 521–529.
25. Sitkoff D, Sharp K, Honig B (1994) Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 98:1978–1988.
26. 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.
27. Fafarman AT, Boxer SG (2010) Nitrile bonds as infrared probes of electrostatics in ribonuclease S. *J Phys Chem B* 114(42):13536–13544.
28. Harvey SC, Hoekstra P (1972) Dielectric relaxation spectra of water adsorbed on lysozyme. *J Phys Chem* 76(21):2987–2994.
29. Frisch MJ, et al. (2004) *Gaussian 03* (Gaussian, Inc., Wallingford, CT), Revision C.02.
30. Gross KC, Seybold PG (2001) Substituent effects on the physical properties and pKa of phenol. *Int J Quantum Chem* 85:569–579.
31. Gross KC, Waybold PG, Hadad CM (2002) Comparison of different atomic charge schemes for predicting pK_a variations in substituted anilines and phenols. *Int J Quantum Chem* 90:445–458.
32. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model* 7:306–317.
33. Berendsen HJC, van der Spoel D, van Drunen R (1995) A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91:43–56.
34. Sorin EJ, Pande VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 88(4):2472–2493.
35. Pearlman DA, et al. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* 91:1–41.
36. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general AMBER force field. *J Comput Chem* 25(9):1157–1174.
37. Hanoian P, Sigala PA, Herschlag D, Hammes-Schiffer S (2010) Hydrogen bonding in the active site of ketosteroid isomerase: Electronic inductive effects and hydrogen bond coupling. *Biochemistry* 49(48):10339–10348.
38. Brycki B, Brzezinski B, Zundel G, Keil T (1992) ^{1}H and ^{13}C NMR studies of the proton transfer in complexes of substituted phenols with trimethylamine N-oxide. *Magn Reson Chem* 30(6):507–510.
39. Jha SK, Ji M, Gaffney KJ, Boxer SG (2011) Direct measurement of the protein response to an electrostatic perturbation that mimics the catalytic cycle in ketosteroid isomerase. *Proc Natl Acad Sci USA* 108(40):16612–16617.

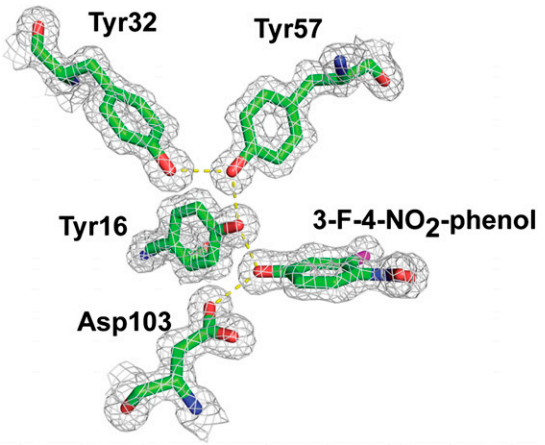

Fig. S1. $2F_o$-F_c electron density map (contoured at 1.5 σ) for the 1.30-Å resolution structure of 3-F-4-NO_2-phenol bound to pKSI D40N.

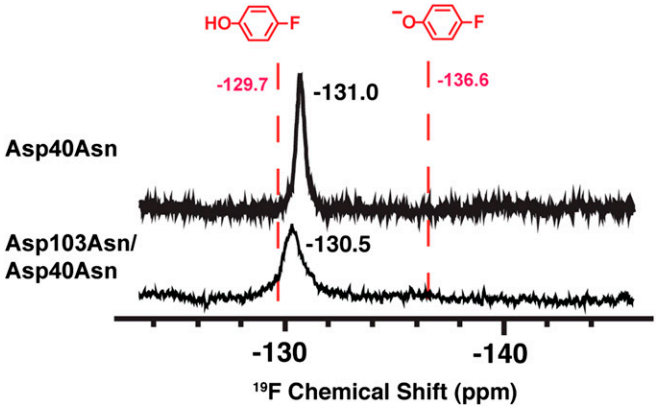

Fig. S2. ^{19}F NMR spectrum of 1 mM 4-F-3-Me-phenol bound to 1.5 mM pKSI D40N or D103N/D40N in 40 mM potassium phosphate (pH 7.2) and 1 mM EDTA.

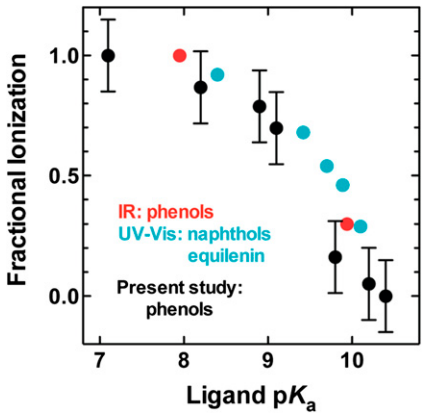

Fig. S3. Comparison of fractional ionization measurements for pKSI D40N-bound ligands. Phenols from the present study are shown in black, and their fractional ionizations and associated uncertainties were estimated indirectly from the ^{13}C-Tyr NMR data as explained in the main text and *SI Materials and Methods*. Phenols previously studied by FTIR are shown in red (11). Fractional ionizations previously reported for naphthols and equilenin based on UV-visible (Vis) absorbance are shown in cyan (17).

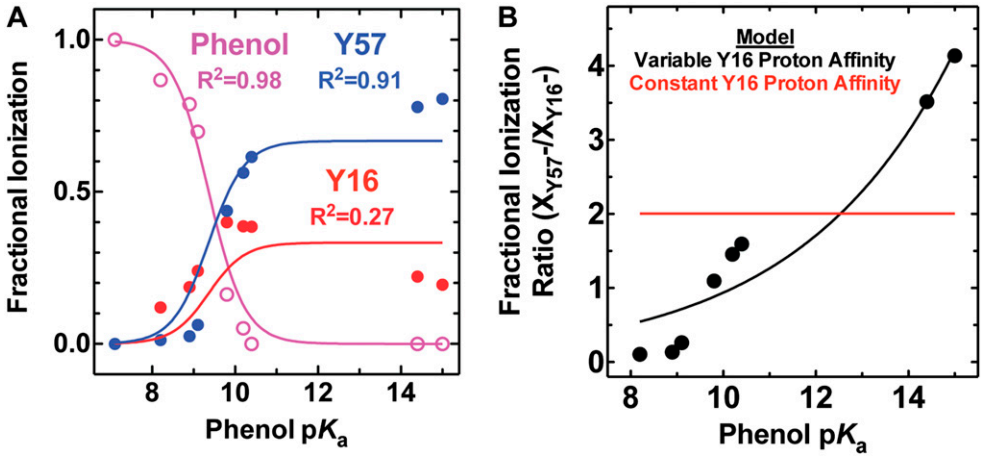

Fig. S4. Global fit of ionization data to equilibrium titration model assuming constant proton affinity values for Y16 and Y57. (*A*) Fitting to this model fails to account for the apparent decrease in Y16 fractional ionization at high phenol pK_a and poorly fits the observed fractional ionization changes for Y16 and Y57 in the low phenol pK_a region. The expressions used for fitting are given in *SI Materials and Methods*. (*B*) The model in which Y16 and Y57 have constant proton affinities (red) regardless of phenol pK_a predicts a constant ratio of ionized Y57 to ionized Y16 ($X_{Y57}-/X_{Y16}-$), contradicting the experimental observation by ^{13}C NMR that this ratio increases with increasing phenol pK_a. A model with variable Y16 proton affinity (black) can account for the increase in $X_{Y57}-/X_{Y16}-$ ($R^2 = 0.92$).

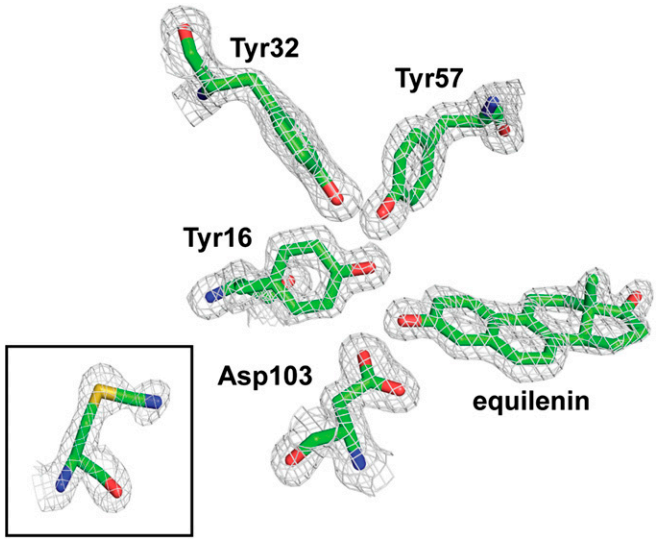

Fig. S5. $2F_o$-F_c electron density map (contoured at 1.5 σ) for the 1.7-Å resolution structure of equilenin bound to D40N/M116C-CN. (*Inset*) For clarity, the M116C-CN group was removed from the view shown for the other groups and is shown in the box.

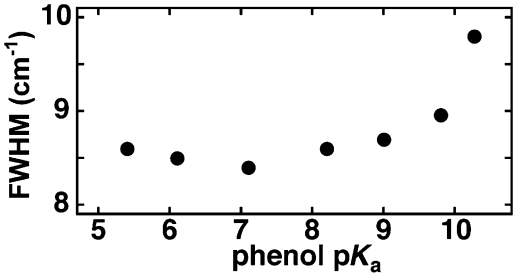

Fig. S6. Observed nitrile IR peak width (FWHM) for D40N/M116C-CN bound to phenols of increasing pK_a (data are from Table S2).

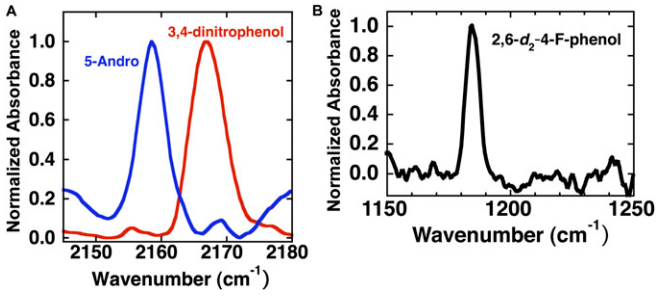

Fig. S7. (*A*) Low-temperature (80 K) IR spectra of the nitrile stretch for M116C-CN/D40N bound to 5-Andro (blue, pK_a = 15) or 3,4-dinitrophenol (red, pK_a = 5.4). (*B*) IR spectrum at 80 K of the C-F stretch for 2,6-d_2-4-F-phenol bound to M116C-CN/D40N. The single C-F peak for this complex at 80 K is similar to the peak observed at 1190 cm^{-1} for neutral 2,6-d_2-4-F-phenol in room temperature spectra taken in aqueous solution (pH 2) (11).

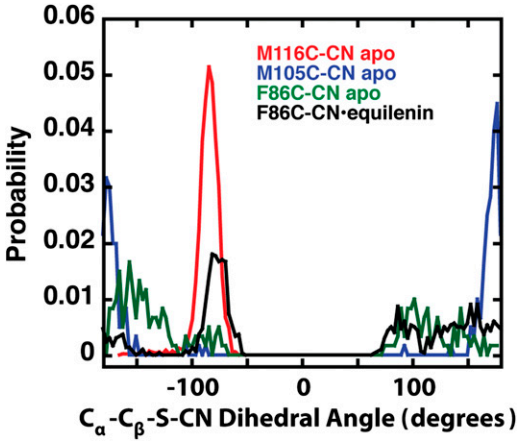

Fig. S8. Distribution of the C_α-C_β-S-CN dihedral angle for each probe during MD simulations of KSI-CN.

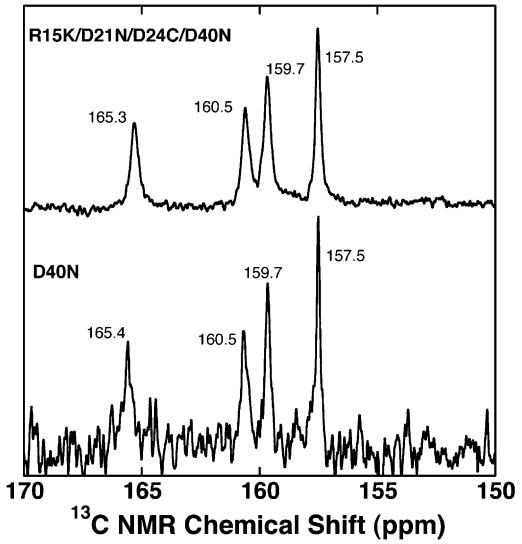

Fig. S9. ^{13}C NMR spectra of unliganded pKSI D40N and D40N/R15K/D21N/D24C containing $^{13}C_\zeta$-Tyr labels at all four Tyr residues.

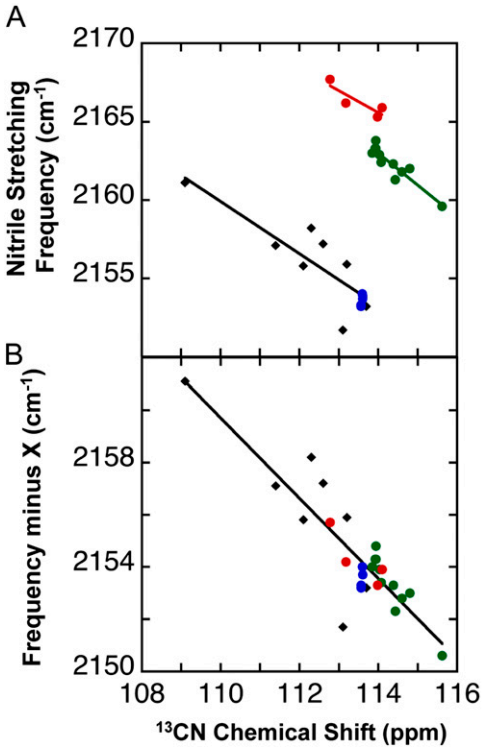

Fig. S10. Nitrile stretching frequency vs. ^{13}C NMR chemical shift in model compounds and in KSI-CN variants. (*A*) Plot of uncorrected IR stretch frequency vs. NMR chemical shift for ethylthiocyanate in aprotic solvents [black, compound identities provided by Fafarman et al. (3); green, M116C-CN/D40N–phenolate complexes; red, F86C-CN/D40N–phenolate complexes; and blue, M105C-CN/D40N–phenolate complexes]. (*B*) Corrected plot of IR stretch frequency vs. NMR chemical shift after subtracting a constant offset of 13 cm^{-1} (F86C-CN) or 10 cm^{-1} (M116C-CN) from the measured nitrile stretch frequency for phenolate complexes of M116C-CN or F86C-CN to correct for hydrogen bond formation to these probes. Symbol definitions are the same as for panel *A*.

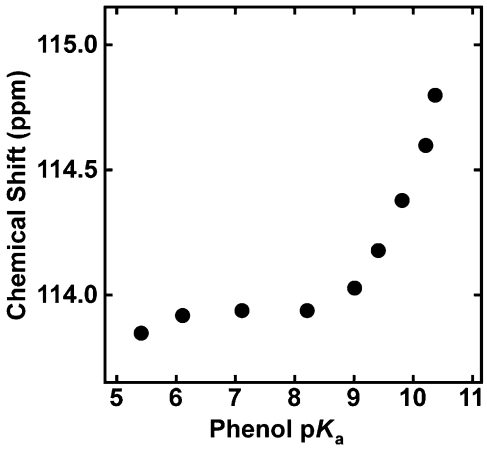

Fig. S11. ^{13}C NMR chemical shift of the nitrile of D60N/M116C-CN bound to phenols of differing pK_a.

Table S1. X-ray data collection and refinement statistics

KSI variant	D40N	D40N/M116C-CN
Ligand	3-F-4-NO_2-phenol	Equilenin
PDB ID code	3VGN	3OWS
Dataset		
Resolution range, Å	35.0–1.30	36.41–1.71
Space group	$P2_12_12_1$	$P2_1$
a, Å	35.48	45.6
b, Å	72.32	129.7
c, Å	95.28	45.5
α, °	90.0	90.0
β, °	90.0	112.8
γ, °	90.0	90.0
No. of unique reflections	55,738	41,740
Completeness, %	91.3	78.8
Multiplicity	4.1	4.8
R_{merge}, %*	7.6	12.6
$I/\sigma_{overall}$ ($I/\sigma_{high\ res}$)	10.4 (1.9)	9.8 (1.3)
Refinement statistics		
No. of residues	254	129
No. of waters	171	290
R_{work}, %†	16.9	21.2
R_{free}, %‡	21.8	25.7
rmsd bond, Å	0.012	0.006
rmsd angle, °	1.53	1.08

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

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

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

Table S2. Nitrile IR stretch frequencies and ^{13}C NMR chemical shifts for bound phenols

Bound phenol	Phenol pK_a	F86C-CN IR frequency, cm^{-1}	M105C-CN IR frequency, cm^{-1}	M116C-CN IR frequency, cm^{-1}	M116C-CN IR FWHM, cm^{-1}	M116C-CN ^{13}C NMR chemical shift, ppm
3,4-$(NO_2)_2$	5.4	2,166.4	2,153.4	2,163.0	8.6	113.85
3-F-4-NO_2	6.1	2,166.4	2,153.3	2,163.3	8.5	113.92
3-CF_3-4-NO_2	6.3			2,163.6		
4-NO_2	7.1	2,166.6	2,153.3	2,163.8	8.4	113.94
3,4,5-Cl_3	7.6			2,164.1		
3-NO_2-4-Cl	7.8			2,163.6		
4-CN	8.0			2,164.0		
3,4,5-F_3	8.2	2,166.2	2,153.3	2,163.3	8.6	113.94
3-NO_2	8.4			2,163.5		
4-CF_3	8.9					
3-Cl	9.0	2,166.6	2,153.5	2,162.9	8.7	114.03
3,4-F_2	9.1					
4-Cl	9.4					114.18
4-F-3-Me	9.8	2,166.2	2,153.7	2,162.3	9.0	114.38
4-MeO	10.2		2,153.8	2,161.8	9.8	114.60
3,4-Me_2	10.4			2,162.0		114.80

IR frequencies and line widths are averages from triplicate measurements (SD ± 0.1 cm^{-1}). IR and NMR data for M116C-CN with bound 4-F-3-Me-phenol were previously published in ref. 3. IR and NMR data for all three KSI-CN variants with bound 3-F-4-NO_2-phenol and 4-NO_2-phenol were previously published in ref. 4.

Table S3. Measured distances between nitrile probes and ionizing active site residues

	Y57, Å	Y16, Å	Phenol, Å
M116C-CN	4.74	5.18	4.46
M105C-CN	7.72	6.05	7.60
F86C-CN	9.32	7.15	5.07

Distances were measured from the nitrogen atom of each nitrile probe to the hydroxylic oxygen atom of Y57, Y16, or the bound phenol and are based on the overlay in Fig. 5.