

# Quantitative dissection of hydrogen bond-mediated proton transfer in the ketosteroid isomerase active site

Paul A. Sigala<sup>a,1,2</sup>, Aaron T. Fafarman<sup>b,1,3</sup>, Jason P. Schwans<sup>a,1,4</sup>, Stephen D. Fried<sup>b</sup>, Timothy D. Fenn<sup>c</sup>, Jose M. M. Caaveiro<sup>d,5</sup>, Brandon Pybus<sup>d,6</sup>, Dagmar Ringe<sup>d</sup>, Gregory A. Petsko<sup>d</sup>, Steven G. Boxer<sup>b,7</sup>, and Daniel Herschlag<sup>a,b,7</sup>

Departments of <sup>a</sup>Biochemistry, <sup>b</sup>Chemistry, and <sup>c</sup>Molecular and Cellular Physiology, Stanford University, Stanford, CA 94305; and <sup>d</sup>Departments of Biochemistry and Chemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454

Edited by David Baker, University of Washington, Seattle, WA, and approved May 29, 2013 (received for review February 1, 2013)

**Hydrogen bond networks are key elements of protein structure and function but have been challenging to study within the complex protein environment. We have carried out in-depth interrogations of the proton transfer equilibrium within a hydrogen bond network formed to bound phenols in the active site of ketosteroid isomerase. We systematically varied the proton affinity of the phenol using differing electron-withdrawing substituents and incorporated site-specific NMR and IR probes to quantitatively map the proton and charge rearrangements within the network that accompany incremental increases in phenol proton affinity. The observed ionization changes were accurately described by a simple equilibrium proton transfer model that strongly suggests the intrinsic proton affinity of one of the Tyr residues in the network, Tyr16, does not remain constant but rather systematically increases due to weakening of the phenol–Tyr16 anion hydrogen bond with increasing phenol proton affinity. Using vibrational Stark spectroscopy, we quantified the electrostatic field changes within the surrounding active site that accompany these rearrangements within the network. We were able to model these changes accurately using continuum electrostatic calculations, suggesting a high degree of conformational restriction within the protein matrix. Our study affords direct insight into the physical and energetic properties of a hydrogen bond network within a protein interior and provides an example of a highly controlled system with minimal conformational rearrangements in which the observed physical changes can be accurately modeled by theoretical calculations.**

computational modeling | enzyme catalysis | protein electrostatics | protein semisynthesis | active site environment

Hydrogen bond networks are ubiquitous structural features within proteins, and they play key roles linking secondary and tertiary structural elements and spanning protein–protein interfaces. Such networks are especially common within enzyme active sites, where they position protein and substrate groups for catalysis, stabilize charge rearrangements during chemical transformations, and mediate proton transfers (1). Despite the prevalence and critical structural and functional roles of hydrogen bond networks, incisive dissection of their physical properties within the idiosyncratic interior of folded proteins remains difficult.

Hydrogen-bonded protons are not observed in the vast majority of protein X-ray structures due to the low X-ray scattering power of hydrogen atoms (2). Thus, the presence of hydrogen bond networks is typically inferred from the proximity and orientation of hydrogen bond donor and acceptor groups within refined protein structural models. The inherent inability of most X-ray diffraction studies to monitor proton positions imposes additional challenges for dissecting the physical features that influence the equilibrium protonation states of specific residues along a hydrogen-bonded proton transfer network. Furthermore, it remains extremely challenging to study the electrostatic consequences of charge rearrangements that accompany hydrogen bond-mediated proton transfers. Few experimental methods exist to vary the ionization properties of discrete protein groups incrementally, and structural

rearrangements within the protein matrix that typically accompany charge rearrangements complicate computational modeling and the straightforward interpretation of the electrostatic properties of protein active sites and interiors (3–5).

Bacterial ketosteroid isomerase (KSI) from *Pseudomonas putida* KSI (pKSI) and *Comamonas testosteroni* KSI has been a powerful system with which to study the physical properties of hydrogen bonds within an enzyme active site (6–13). KSI uses a general base, D40 (pKSI numbering), to deprotonate steroid substrates and form a dienolate reaction intermediate that is stabilized by hydrogen bonds donated by Y16 and protonated D103. Y16 is further linked via hydrogen bonds to Y57 and Y32, forming an extended active site hydrogen bond network in pKSI (Fig. 1A). Phenolic ligands, such as single-ring phenols, two-ring naphthols, and four-ring steroids like equilenin or estradiol, can bind in the KSI active site as negatively charged oxyanions and accept

## Significance

Hydrogen bond networks play critical structural and functional roles in proteins but have been challenging to study within this complex environment. We incorporated spectroscopic probes into the active site of the bacterial enzyme ketosteroid isomerase to systematically dissect the proton transfer equilibrium within a key hydrogen bond network formed to bound transition state analogs. Our study provides direct insight into the physical and energetic properties of a hydrogen bond network within an enzyme and presents a simple computational model of electrostatic effects within this protein that succeeds due to detailed knowledge of ionization states and a tightly controlled experimental system.

Author contributions: P.A.S., A.T.F., S.G.B., and D.H. designed research; P.A.S., A.T.F., J.P.S., S.D.F., T.D.F., J.M.M.C., and B.P. performed research; P.A.S., A.T.F., J.P.S., S.D.F., T.D.F., J.M.M.C., B.P., D.R., G.A.P., S.G.B., and D.H. analyzed data; and P.A.S., A.T.F., J.P.S., S.D.F., S.G.B., and D.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID codes 3VGN and 3OWS).

<sup>1</sup>P.A.S., A.T.F., and J.P.S. contributed equally to this work.

<sup>2</sup>Present address: Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110.

<sup>3</sup>Present address: Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104.

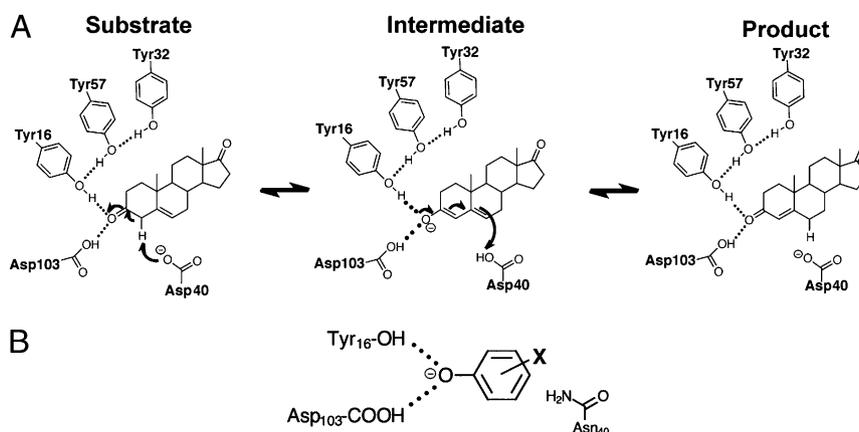
<sup>4</sup>Present address: Department of Chemistry and Biochemistry, California State University, Long Beach, CA 90840.

<sup>5</sup>Present address: Laboratory of Medical Proteomics, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan.

<sup>6</sup>Present address: Experimental Therapeutics Branch, Walter Reed Army Institute of Research, Silver Spring, MD 20910.

<sup>7</sup>To whom correspondence may be addressed. E-mail: [sboxer@stanford.edu](mailto:sboxer@stanford.edu) or [herschla@stanford.edu](mailto:herschla@stanford.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302191110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302191110/-DCSupplemental).



**Fig. 1.** KSI reaction and reaction intermediate analog. (A) KSI reaction mechanism for isomerization of 5-androstene-3,17-dione. (B) Schematic depiction of an ionized substituted phenol bound at the KSI D40N active site.

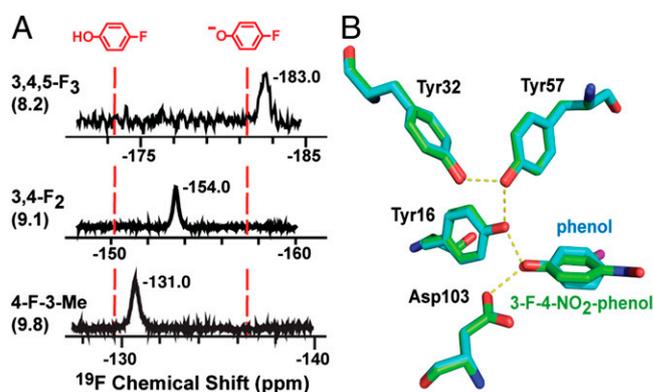
hydrogen bonds from Y16 and D103, mimicking the oxyanion charge localization of the dienolate reaction intermediate and dienolate-like transition states (6–8, 14, 15) (Fig. 1B). A homologous series of bound phenols or naphthols bearing different electron-withdrawing substituents provides a deft experimental tool with which to incrementally vary the proton affinity and negative charge density of the phenolic oxygen (16). These changes tune the structure and strength of hydrogen bonds formed to phenolic ligands (17–19), and thus provide a systematic probe of the physical and energetic properties of the hydrogen bond network within the KSI oxyanion hole (6–9, 11, 13) and the response of the surrounding protein matrix to such changes (20, 21).

Recent studies with the D40N pKSI mutant, which mimics the protonated D40 present in the KSI–dienolate intermediate complex (Fig. 1A), have provided evidence that ligands of increasing  $pK_a$  are bound as an increasing population of neutral, protonated phenol (11, 13, 22). These results suggest that an unspecified active site residue can ionize with increasing phenol  $pK_a$ , resulting in a net proton transfer to the bound ligand. We have used site-specific NMR and IR probes and KSI semisynthesis to determine that either of two different Tyr residues within the extended hydrogen bond network can ionize, and we have systematically mapped the changes in their equilibrium ionization states as a function of the proton affinity and hydrogen bonding capability of the phenolic ligand. We further measured the electric field changes at discrete active site positions due to charge rearrangements within the hydrogen bond network. We demonstrate that a static continuum electrostatic model with a low dielectric can accurately describe these changes, suggesting a high degree of structural organization and the absence of substantial conformational rearrangement in response to charge rearrangement within the active site.

## Results and Discussion

**Ionization States of Substituted Phenols Bound to pKSI D40N.** Recent spectroscopic studies suggested that single- and multiple-ring phenolic ligands are bound to pKSI D40N with a neutral, protonated fraction that increases with ligand  $pK_a$  and reaches a ratio of 50:50 ionized/neutral for a ligand with a solution  $pK_a$  of 9.7 (13, 22). To dissect and understand the nature and properties of proton transfer within the KSI oxyanion hole, we turned to systematic studies with single-ring phenols, because these compounds are available over a wider  $pK_a$  range than naphthols or steroids. To probe the ionization state of phenols bound to KSI, we acquired  $^{19}\text{F}$  NMR spectra of 4-fluoro-substituted phenols bound to the D40N mutant, because the chemical shift of the 4-fluoro nucleus is a sensitive reporter of changes in phenol ionization state and shifts 7–8 ppm down-field on ionization (8) (Fig. 2A).

Prior UV-visible (Vis) and IR absorbance studies have suggested that phenols with  $pK_a$  values  $\leq 8$  are bound to D40N in their ionized, phenolate form (8, 22). In agreement with these prior results, the 4-fluoro nucleus of 3,4,5- $\text{F}_3$ -phenol ( $pK_a = 8.2$ ) has a  $^{19}\text{F}$  chemical shift when bound to pKSI D40N that is 1.5 ppm up-field of that observed for the phenolate anion in solution (Fig. 2A). This modest up-field shift relative to solution is consistent with the ability of aromatic ring currents and other shielding differences between a protein interior and water to result in chemical shift differences of 1–2 ppm for a  $^{19}\text{F}$  nucleus (23). In contrast, 3,4- $\text{F}_2$ -phenol ( $pK_a = 9.1$ ) bound to pKSI D40N displays a chemical shift that is intermediate between the observed values for its neutral (phenol) and ionized (phenolate) forms in solution, suggesting that this phenol binds to pKSI D40N as a mixture of neutral phenol and ionized phenolate and that these forms are rapidly interconverting relative to their  $^{19}\text{F}$  NMR frequency difference (additional discussion is provided in *SI Text*). The 4-F-3-Me-phenol, with a still higher  $pK_a$  of 9.8, exhibits a chemical shift when bound that is 1.3 ppm up-field of the neutral phenol in solution (Fig. 2A). This series of spectra strongly suggests that single-ring phenols bind to the pKSI D40N mutant with a neutral phenol



**Fig. 2.** Spectroscopic and structural analysis of phenols bound to pKSI D40N. (A)  $^{19}\text{F}$  NMR spectra and chemical shifts of the 4-F group for F-substituted phenols bound to KSI at pH 7.2 (black peaks) or free in solution (red dashes) at pH 2 (neutral) or pH 12 (ionized). Phenol  $pK_a$  values are shown in parentheses. (B) Superposition of the 1.30-Å resolution D40N–3-F-4- $\text{NO}_2$ -phenol (carbon atoms are colored green; PDB ID code 3VGN) and 1.25-Å resolution D40N–phenol (carbon atoms are colored cyan; PDB ID code 2PZV) X-ray crystal structures. Oxygen, nitrogen, and fluorine atoms are colored red, blue, and magenta, respectively.

fraction that becomes significant at a phenol  $pK_a$  greater than 8 and increases with the phenol  $pK_a$  until the protonated phenol is the predominantly bound form at or above a  $pK_a$  of 10. This conclusion is consistent with prior UV and IR studies (13, 22), and it is further supported by  $^{13}\text{C}$  NMR and IR results described below.

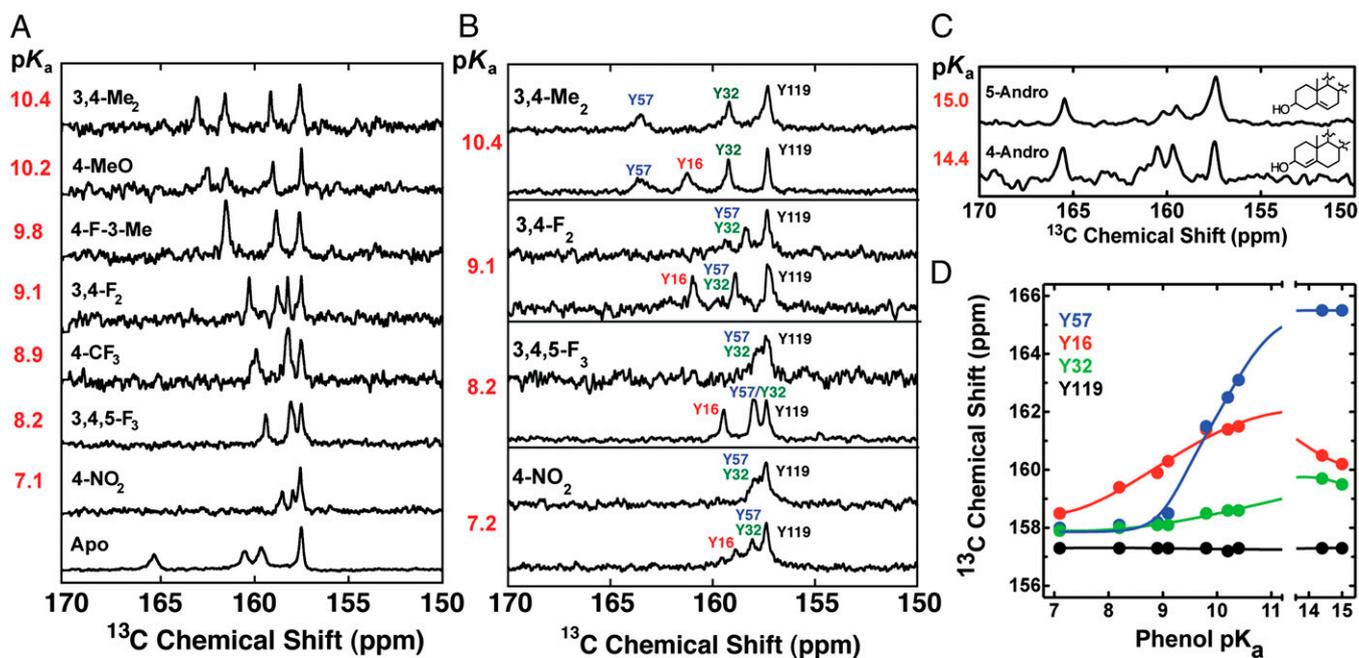
**Structural Similarity of pKSI D40N Bound to Low vs. High  $pK_a$  Phenols.** To assess whether the change in phenol ionization state for low vs. high  $pK_a$  phenols is accompanied by conformational rearrangements within or surrounding the hydrogen bond network of the pKSI oxyanion hole, we determined the 1.30-Å resolution X-ray crystal structure of 3-F-4-NO<sub>2</sub>-phenol ( $pK_a = 6.1$ ) bound to pKSI D40N [data collection and refinement statistics are shown in Table S1, and electron density map is shown in Fig. S1; Protein Data Bank (PDB) ID code 3VGN] and compared this structure with the previously published 1.25-Å resolution structure of the same pKSI mutant bound to unsubstituted phenol ( $pK_a = 10.0$ ) (8). Superposition of the two structures reveals that the overall structures are indistinguishable, that the bound ligands are similarly positioned within the KSI active site, and that the oxyanion hole residues that form the hydrogen bond network to the hydroxylic oxygen of each phenol are nearly identically positioned in both structures (rmsd = 0.130 Å; Fig. 2B). These observations rule out the possibility that gross structural rearrangements within the hydrogen bond network accompany changes in the ionization state of bound phenols. Such rearrangements can complicate the analysis and modeling of electrostatic effects, and their absence in this case is a simplifying feature for the experimental and computational studies that follow.

**NMR Identification of Tyr Ionizations in KSI-Phenol Complexes Using Site-Specific  $^{13}\text{C}$ -Labeled Tyrs.** We considered either D103 or Y16, both of which directly donate hydrogen bonds to the bound ligand (Figs. 1A and 2B), as the most likely residue to be ionized in the presence of higher  $pK_a$  phenols. A priori, the carboxylic acid

moiety of an aspartic acid (solution  $pK_a$  of  $\sim 4$ ) would be expected to be much more acidic than that of a Tyr hydroxyl group (solution  $pK_a$  of  $\sim 10$ ) (24). However, D103 is surrounded by hydrophobic residues that elevate its  $pK_a$  well above its typical solution value (15, 25, 26), and recent quantum mechanics/molecular mechanics (QM/MM) studies of phenols bound to pKSI D40N have suggested that D103 may be less acidic than Y16 (11). Furthermore, we observed that the D103N/D40N mutant, which preserves hydrogen bonding to residue 103 but ablates its ability to transfer a proton, still binds 4-F-3-Me-phenol ( $pK_a = 9.8$ ) predominantly in its neutral form (Fig. S2), strongly suggesting that D103 is not the residue that ionizes as bound phenols become protonated. Based on this result and on our prior observation of an ionized Tyr in unliganded pKSI D40N (20), we considered Y16 the most likely residue to be ionized when the bound phenol is neutral. The results described below provide evidence for ionization of both Y16 and Y57.

pKSI D40N contains four Tyrs. Y119 is a surface residue located far from the active site, whereas Y32 and Y57 form an active site hydrogen bond network to Y16, which directly donates a hydrogen bond to bound phenols (Figs. 1A and 2B). The chemical shift of the C<sub>ε</sub> carbon of Tyr, the carbon atom adjacent to the hydroxyl group, is highly sensitive to the ionization state of Tyr, shifting down-field from 155.5 to 166.3 ppm upon ionization in Tyr-containing peptides in aqueous solution (24). We therefore used  $^{13}\text{C}$  NMR of phenol complexes of pKSI D40N containing  $^{13}\text{C}$ -labeled Tyr residues to determine whether Y16 or any of the active site Tyr residues ionize upon phenol binding.

As previously reported by Fafarman et al. (20), the  $^{13}\text{C}$  spectrum of unliganded pKSI D40N displays four well-resolved C<sub>ε</sub>-Tyr peaks (Fig. 3A, lowest spectrum). The far down-field peak at 165.1 ppm, indicative of an ionized Tyr, was tentatively assigned to Y57 based on comparisons of data and computational models of electrostatic field effects (20). On binding of 4-NO<sub>2</sub>-phenol ( $pK_a = 7.1$ ),



**Fig. 3.**  $^{13}\text{C}$  NMR spectra of pKSI D40N containing  $^{13}\text{C}$ -Tyr labels. (A) Spectra of recombinant D40N uniformly labeled with  $^{13}\text{C}$ -Tyr and bound to a series of phenols with increasing solution  $pK_a$  (values are colored red, and phenol substituent groups are indicated above each spectrum). Spectra of D40N apoenzyme and 4-nitrophenol-bound D40N were previously published (20). (B) Spectra of recombinant (Lower) or semisynthetic (Upper) D40N/R15K/D21N/D24C (explanations of additional mutations are provided in the main text and *SI Materials and Methods*) bearing  $^{13}\text{C}$ -Tyr labels at all four Tyrs (recombinant) or only at Y32/Y57/Y119 (semisynthetic) and bound to the indicated phenols. Peaks have been assigned as described in the text. (C) Spectra of recombinant D40N uniformly labeled with  $^{13}\text{C}$ -Tyr and bound to 5-Andro or 4-Andro. For simplicity, only the A and B steroid rings are shown. (D) Chemical shift for each assigned Tyr peak is plotted as a function of phenol  $pK_a$ . Trend lines are empirical fits to guide the eye.

which is fully ionized in complex with pKSI D40N (8), all the observed  $^{13}\text{C}$ -Tyr peaks in D40N are shifted up-field to less than 159 ppm (20), as expected for the protonated forms of all four Tyr residues (Fig. 3A). The 1.2-ppm chemical shift dispersion for the discrete  $\text{C}_\alpha$ -Tyr peaks is within the range of 1–2 ppm expected for differential shielding contributions arising from the unique local structural environment of each Tyr (27). Binding of phenols with  $\text{pK}_a$  values increasing from 7.1 to 10.4 resulted in no changes in the position of the most up-field peak at 157.3, previously assigned via mutagenesis to the surface Y119 (20), but steadily shifted the position of the remaining active site Tyr peaks further down-field (Fig. 3A and B). At the highest phenol  $\text{pK}_a$  of 10.4, two peaks have shifted  $\geq 3$  ppm down-field (relative to their positions in the D40N–4-NO<sub>2</sub>-phenolate spectrum) to 163.1 ppm and 161.5 ppm.

Because simple phenols with  $\text{pK}_a$  values  $>10.4$  are not readily available, we acquired spectra of pKSI D40N bound to the cyclohexanolic steroids 5-androsten-3-ol-17-one (5-Andro) and 4-androsten-3-ol-17-one (4-Andro) to mimic binding of phenols with high  $\text{pK}_a$ . These compounds have estimated  $\text{pK}_a$  values of 15.0 (5-Andro) and 14.4 (4-Andro), and 5-Andro was previously shown to bind pKSI D40N in a structurally similar manner to phenols, with the hydroxyl oxygen of its cyclohexanolic A-ring positioned within hydrogen bonding distance of Y16 and D103 (28). Furthermore, prior studies have shown that the multiple distal rings of a steroid do not alter the electrostatic environment within the oxyanion hole, relative to a bound single-ring phenol (29). The  $^{13}\text{C}$  NMR spectra of the D40N–5-Andro and D40N–4-Andro complexes showed three up-field peaks at  $<161$  ppm and a single far down-field peak at 165.5 ppm (Fig. 3C), strongly suggesting the presence of a single predominantly ionized Tyr at a ligand  $\text{pK}_a$  of  $\sim 15$ . In the analyses that follow, we interpret the Andro spectra as reflecting the properties of bound single-ring phenols with equivalent  $\text{pK}_a$  values.

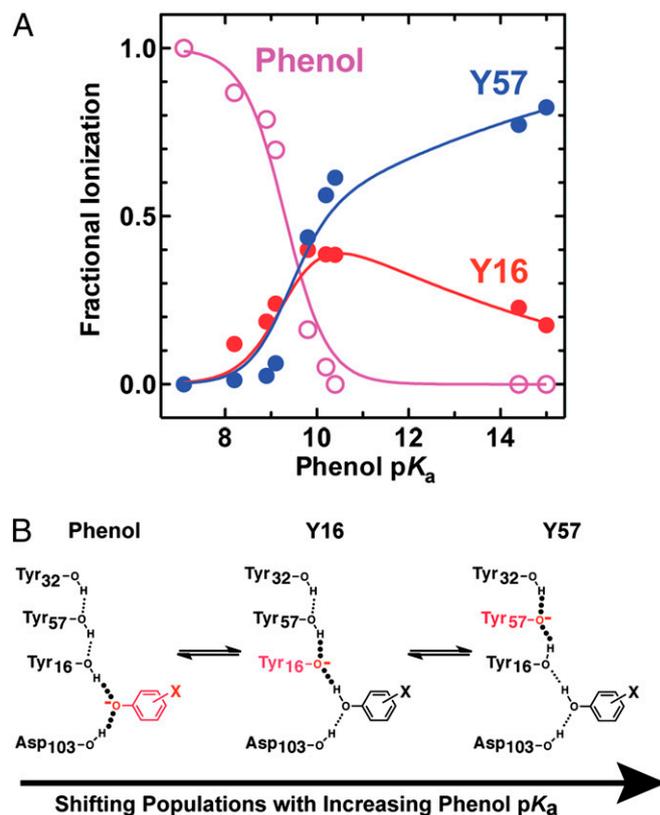
To assign the observed  $^{13}\text{C}$ -Tyr peaks in spectra of D40N–phenol complexes (Fig. 3A), we prepared semisynthetic pKSI D40N (a complete description is provided in *SI Materials and Methods*) by ligating a synthetic peptide containing unlabeled Y16 to a recombinant peptide fragment containing  $^{13}\text{C}_\alpha$ -labeled Y32, Y57, and Y119, and refolding the full-length enzyme out of urea. We acquired  $^{13}\text{C}$  spectra for semisynthetic (Y16 unlabeled) and recombinant (uniformly  $^{13}\text{C}$ -Tyr-labeled) D40N bound to a series of phenols with  $\text{pK}_a$  values of 7.1–10.4 (Fig. 3B). In the presence of bound phenols with  $\text{pK}_a$  values of 7.1–9.1, the most down-field resonance observed for uniformly  $^{13}\text{C}$ -Tyr-labeled D40N is absent in spectra of semisynthetic D40N (Y16 unlabeled), identifying this peak as arising from Y16 in these complexes. In contrast, in a complex between semisynthetic D40N and 3,4-Me<sub>2</sub>-phenol ( $\text{pK}_a = 10.4$ ), the second-most down-field peak at 161.5 pm is missing, identifying this peak as Y16 and indicating a crossover in the identity of the most down-field peak between  $\text{pK}_a$  values of 9.1 and 10.4.

By elimination, the most down-field peak at 163.1 ppm in the 3,4-Me<sub>2</sub>-phenol spectrum corresponds to either Y57 or Y32. Y32 is the terminal residue in the hydrogen bond network (Figs. 1A and 2B) and, other than Y57, is surrounded by hydrophobic residues that are expected to destabilize ionized Y32. In contrast, Y57 is positioned within hydrogen bonding distance of both Y16 and Y32, and prior results suggest that it has a highly perturbed  $\text{pK}_a$  of 6.3 in the pKSI D40N apoenzyme (20). We therefore assigned the 163.1 ppm peak in the D40N–3,4-Me<sub>2</sub>-phenol spectrum to Y57. This information and the spectral comparisons with semisynthetic KSI described above allowed us to assign the  $^{13}\text{C}$ -Tyr peaks in each spectrum as indicated in Fig. 3B, and the observed chemical shift for each assigned peak is plotted in Fig. 3D as a function of phenol  $\text{pK}_a$ .

**Quantitative Modeling of Ionization States Within the KSI Hydrogen Bond Network.** As a basis for modeling and understanding the changes in protonation state of the bound phenol, Y16, and Y57

as a function of phenol  $\text{pK}_a$  and how these charge rearrangements are sensed within the active site, we used the  $^{13}\text{C}$  NMR chemical shift changes in Fig. 3 to estimate the fraction of each of these three groups present in its ionized form at equilibrium with differing bound phenols (additional discussion is provided in *SI Text*). To assign these fractional ionizations (Fig. 4A), we assumed that only a single group is ionized in the KSI active site at any given time, consistent with the observed pH dependence for phenol binding to D40N (8). In this model, proton transfers between the phenol, Y16, and Y57 shift the equilibrium population of each ionized group (Fig. 4B), with the sum of the fractions of each group present in its ionized form at each phenol  $\text{pK}_a$  always equal to 1.

The fractional populations of Y16 and Y57 present as ionized tyrosinates ( $X_{Y16}$  and  $X_{Y57}$ ) were estimated at each phenol  $\text{pK}_a$  value by first calculating the chemical shift difference between the observed peak position for each Tyr and that observed at a  $\text{pK}_a$  of 7.1 (Fig. 3), where the bound phenol is fully ionized (8, 22) (Fig. 2A) and Y16 and Y57 are thus fully neutral. This value was then divided by the total expected  $^{13}\text{C}$  chemical shift dispersion between an ionized and neutral Tyr in the KSI active site (166 ppm minus the observed chemical shift at a phenol  $\text{pK}_a$  of 7.1; additional discussion is provided in *SI Materials and Methods*) to estimate fractional ionization ( $X_i$ ) values for each Tyr residue at



**Fig. 4.** Quantitative fractional ionization model of hydrogen bonding groups within the KSI D40N active site with bound phenols of increasing solution  $\text{pK}_a$ . (A) Fractional ionization values (estimated uncertainty  $\pm 0.15$ ) were derived from the  $^{13}\text{C}$  NMR data as explained in the main text and *SI Materials and Methods*. The data were globally fit ( $R^2$  values of individual fits = 0.92–0.97) with the equilibrium proton transfer model given in *Methods* to give best-fit values of 8.6 for parameter  $a$ , 0.1 for parameter  $b$  (the slope of the linear dependence of Y16 acidity on phenol  $\text{pK}_a$ ), and 9.6 for the apparent  $\text{pK}_a$  of Y57. (B) Schematic depiction of the ionization states present within the active site hydrogen bond network whose fractional populations shift as a function of increasing phenol  $\text{pK}_a$ .

each phenol  $pK_a$ . Fractional phenol ionizations ( $X_{phenol}$ ) at each  $pK_a$  value were then estimated indirectly by subtracting the sum of  $X_{16}$  and  $X_{57}$  from 1. The values of  $X_{phenol}$  calculated in this fashion are similar to those previously reported for bound phenols, naphthols, and equilenin based on FTIR and UV-Vis absorbance experiments (13, 22) (Fig. S3) and are qualitatively consistent with the  $^{19}\text{F}$  NMR changes described above (Fig. 2A). As discussed in *SI Materials and Methods*, we estimate the uncertainty in the assigned  $X_i$  values as  $\pm 0.15$ .

As quantitatively modeled in Fig. 4A and schematically depicted in Fig. 4B, the proton transfer equilibrium changes systematically as a function of the  $pK_a$  of the bound phenol. This incremental redistribution indicates that the ionization states of individual residues within the KSI-phenol hydrogen bond network are systematically altered by varying the proton affinity of a single distal group on the bound phenol. The change in the identity of the predominantly ionized Tyr from Y16 to Y57 at low vs. high phenol  $pK_a$  (e.g.,  $pK_a$  of 7.1–10.4) and the continuing shifts in the ionized populations of these groups above a  $pK_a$  of 10 (Fig. 4A) directly suggest that the relative proton affinities of these two Tyr groups do not remain constant but change with increasing phenol  $pK_a$  to favor ionization of Y57 over Y16, a behavior that we are able to accurately model below.

To understand the origin of these ionization changes better and to test whether increases in the phenol proton affinity alone can accurately explain the observed changes, we derived an equilibrium proton transfer model for the hydrogen bond network based on the solution  $pK_a$  values of the phenols and the apparent  $pK_a$  values of Y16 and Y57 within the phenol-bound complexes (full derivation and additional discussion are provided in *SI Materials and Methods*). The apparent  $pK_a$  values used in this model are a proxy for the relative proton affinities of the hydrogen-bonded groups and do not represent true  $pK_a$  values in the active site environment. Because our goal is to understand the physical and energetic properties that underpin the observed ionization changes rather than to assign microscopic acid dissociation constants, our analysis does not depend on this simplification. We first attempted to fit the data globally with the simpler model, which assumes constant proton affinities for Y16 and Y57 regardless of the phenol  $pK_a$ . As expected, this model failed to account for the predominant ionization of Y16 rather than Y57 below a  $pK_a$  of 10 and could not explain the observed decrease in  $X_{Y16}$  above a  $pK_a$  of 10 (Fig. S4A and B).

Based on prior linear free energy studies and known physical properties of hydrogen bonds (8, 17, 30–35), 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 effect, we modified our equilibrium proton transfer model to allow the proton affinity of the Y16 anion to vary linearly with that of the phenol (modified expressions are shown in *Materials and Methods*, with additional discussion provided in *SI Materials and Methods*). A global fit of this model to the fractional ionization data (Fig. 4A and Fig. S4B) accurately accounted for the observed decrease in  $X_{Y16}$  and increase in  $X_{Y57}$  at high phenol  $pK_a$  values. On the basis of this fit, we conclude that these changes in the proton transfer equilibrium occur as the Y16 proton affinity surpasses that of Y57 due to physical and energetic changes in the phenol–Y16 hydrogen bond with increasing phenol  $pK_a$ . We discuss the physical origins of these hydrogen bond changes in *Conclusions*.

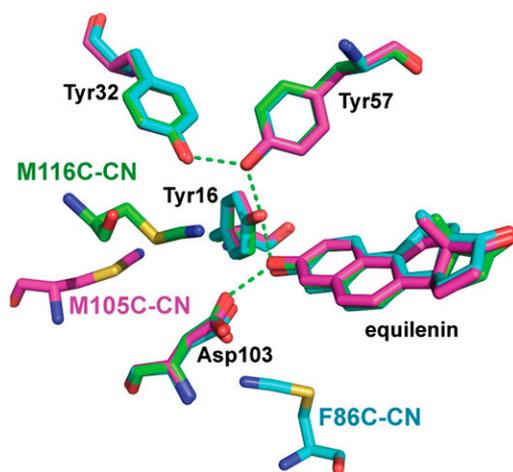
**Incorporation of Nitrile Electric Field Probes and Structural Analysis of KSI Variants.** To evaluate the electrostatic field changes within the protein that accompany proton transfers within the active site hydrogen bond network, we incorporated nitrile (–CN) electric field probes into the pKSI active site. The IR stretching frequency of a nitrile group (in inverse centimeters,  $\text{cm}^{-1}$ ) is linearly sensitive to electrostatic fields (36), and experimental IR frequency shifts

can be used to determine the change in local electric field projected along the nitrile bond axis. This conversion is accomplished by using the average linear Stark tuning rate  $|\Delta\bar{\nu}|$  of  $0.65 \text{ cm}^{-1}/(\text{MV}/\text{cm})$  previously determined for KSI-CN probes (20) and the vibrational Stark effect (VSE) equation  $\Delta\bar{\nu} = -\Delta\bar{\mu} \cdot \Delta\vec{F}$ , where  $\Delta\bar{\nu}$  (in  $\text{cm}^{-1}$ ) is the observed IR peak shift between different bound phenols,  $\Delta\bar{\mu}$  is the difference-dipole moment and is parallel to the –CN bond axis, and  $\Delta\vec{F}$  is the field change associated with the changing identity of the bound phenol (further discussion of the VSE can be found in ref. 20 and references therein).

In prior work, we site-specifically incorporated nitrile groups into pKSI by cyanylating a unique Cys introduced by mutation at position M116, M105, or F86 within a Cys-free C69S/C81S/C97S background (20, 29, 37). These positions were selected on the basis of their proximity (3–11 Å) to the key catalytic groups (Fig. 5). Our previous studies of these KSI-CN variants indicated that nitrile incorporation resulted in minimal perturbation to KSI structure, catalytic activity, and ligand binding (20, 29).

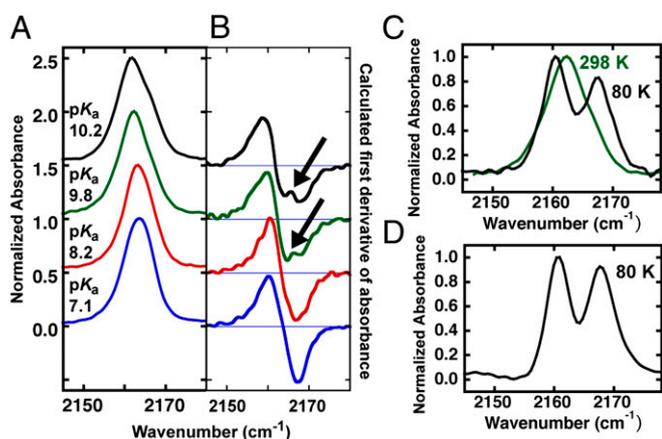
To test for differences in the active site structures of our KSI-CN variants bound to a common ligand, we determined the 1.7-Å resolution X-ray structure of equilenin bound to D40N/M116C-CN (data collection and refinement statistics are shown in Table S1, and an electron density map is shown in Fig. S5; PDB ID code 3OWS) and compared it with the previously published structures of equilenin bound to D40N/F86C-CN (1.7-Å resolution; PDB ID code 3OWU) and D40N/M105C-CN (2.3-Å resolution; PDB ID code 3OWY) (20). As shown in Fig. 5, equilenin binding to the three KSI-CN variants resulted in nearly identical positioning of the bound ligand and of groups within the hydrogen bond network (rmsd = 0.18 Å), and no overall structural changes were observed. Furthermore, the nitrile probe in each structure refined to a single, well-ordered conformation, positioning each nitrile with a unique vantage from which to monitor the electrostatic field effects of ionization changes within the hydrogen bond network.

**Measuring the Electrostatic Field Changes from Proton Transfers Within the Hydrogen Bond Network.** The movement of negative charge from the phenolate oxygen to Y16 and Y57 with increasing phenol  $pK_a$  is expected to alter local electrostatic fields within the KSI active site (20). To directly measure these field changes in



**Fig. 5.** Structural comparison of KSI-CN variants bound to equilenin. Superposition of the 1.7-Å resolution D40N/M116C-CN–equilenin (carbon atoms are colored green; PDB ID code 3OWS), the 1.7-Å resolution D40N/F86C-CN–equilenin (carbon atoms are colored cyan; PDB ID code 3OWU), and the 2.3-Å resolution D40N/M105C-CN–equilenin (carbon atoms are colored magenta; PDB ID code 3OWY) X-ray crystal structures. Oxygen, nitrogen, and sulfur atoms are colored red, blue, and gold, respectively. Each KSI-CN variant contained a single nitrile group.





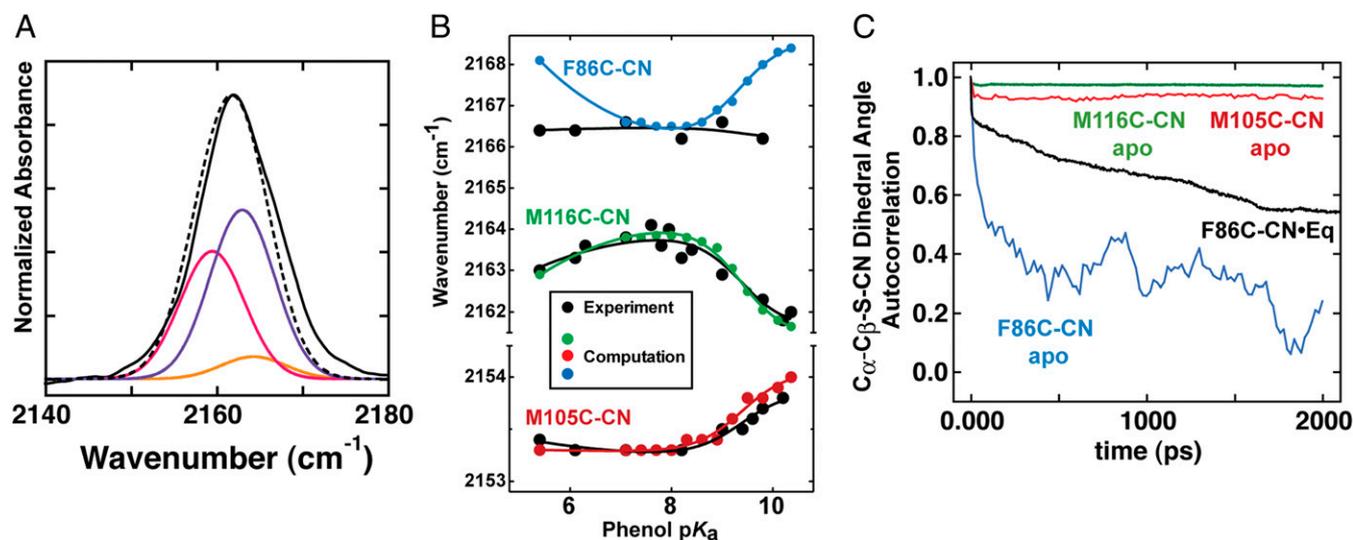
**Fig. 7.** Analysis of spectral width and asymmetry of the nitrile IR absorption spectra of D40N/M116C-CN bound to phenols of differing  $pK_a$ . (A) Room temperature IR absorption spectra in the nitrile stretching region of D40N/M116C-CN bound to 4-NO<sub>2</sub>-phenol (blue), 3-Cl-phenol (red), 4-F-3-Me-phenol (green), and 4-MeO-phenol (black). (B) First derivative of the absorption spectrum calculated from the data in A, highlighting the appearance of a shoulder (inflections, which indicate a shoulder, are marked with arrows) and increasing peak width and asymmetry (i.e., increasing ratio of peak height to trough depth) for higher  $pK_a$  phenols. (C) Superposition of the nitrile stretch peak observed for 4-F-3-Me-phenol ( $pK_a = 9.8$ ) bound to D40N/M116C-CN at 298 K in buffer (green) or at 80 K in 50% glycerol/buffer (black). (D) Nitrile stretch peak observed for D40N/M116C-CN with bound 2,6-*d*<sub>2</sub>-4-F-phenol ( $pK_a = 10.0$ ) at 80 K in 50% glycerol/buffer. A single C-F stretch peak corresponding to the neutral phenol was also observed for this complex at 80 K (Fig. S7B), consistent with our assignment in Fig. 4A that a bound phenol with a  $pK_a$  of ~10 is predominantly neutral and supporting assignment of the two peaks in C and D to the KSI tautomers with ionized Y16 or Y57.

the electrostatic properties of proteins, key limitations in each of these approaches have prevented clean, quantitative measurements of electrostatic fields within proteins (ref. 20 and references therein). Our nitrile-based measurements of electric field

changes provide a metric by which to test and improve computational models of the electrostatic effects of charge rearrangement within a protein interior. Based on our analysis of the shifting ionization equilibria within the hydrogen bond network (Fig. 4), we carried out continuum electrostatic calculations using the software package DelPhi (described in detail in *SI Materials and Methods*) to predict the effect of these ionization changes on the stretching frequency of each nitrile probe within the KSI active site. These computations used an internal protein dielectric of 2 to account for the intrinsic polarizability of amino acids but otherwise assumed the absence of structural rearrangements (i.e., the same static structural model was used in all calculations).

We first calculated the expected electric field change and IR peak shift for each probe due to negative charge residing fully on the oxygen of the bound phenol, Y16, or Y57, and then used the fractional charge distribution from Fig. 4A to calculate a population-weighted IR peak shift as a function of ligand  $pK_a$ . Absolute nitrile stretching frequencies were obtained by adding the calculated peak shift to a chosen reference state, which we selected to be the IR frequency for each KSI-CN variant when bound to 4-nitrophenol ( $pK_a = 7.1$ ) (additional explanation is provided in *SI Materials and Methods*). This referencing is the only adjustable parameter in calculating peak positions and does not affect the trend or scale of the calculated frequency changes.

To account for differences in charge density on the phenolate oxygen for phenol with different solution  $pK_a$  values (16, 38), we used QM-calculated atomic charges for three phenolates with differing electron-withdrawing groups as inputs for the DelPhi calculations. We then used the calculated electrostatic field values for the three explicitly modeled phenolates to construct a linear correlation (for each probe position) between phenol  $pK_a$  and the local electrostatic field due to negative charge on the phenolate oxygen (*SI Materials and Methods*). For calculating IR peak positions with bound phenols above a  $pK_a$  of 8, the coexistence of distinct ionization populations (Fig. 4A and B) that are detectable as discrete species in IR spectra (Fig. 7) required that electric



**Fig. 8.** Comparison of computational and experimental results for KSI-CN variants. (A) Experimental spectrum of D40N/M116C-CN bound to 4-MeO-phenol (solid black line) modeled with a three-spectra basis set (dashed black line) composed of calculated spectra for negative charge on the hydroxylic oxygen of 4-MeO-phenolate (orange), Y16 (red), or Y57 (violet) and weighted by the fractional populations of these species at a  $pK_a$  of 10.2 as shown in Fig. 4. (B) Superposition of experimental (black) and calculated (colored) nitrile IR peak shifts for phenols of increasing  $pK_a$  bound to F86C-CN (cyan), M116C-CN (green), and M105C-CN (red). Computed IR peak shifts were modeled as described in the main text and *SI Materials and Methods*. Trend lines are empirical fits to guide the eye. Note that the  $y$ -axis scale has been expanded for M105C-CN. (C) Plot of the dihedral angle autocorrelation for each probe as a function of time during molecular dynamics simulations of nitrile mobility in the KSI-CN variants. Eq, equilenin.

fields be calculated for each separate tautomer, with negative charge on the ligand, Y16, or Y57.

As shown for M116C-CN with bound 4-MeO-phenol ( $pK_a = 10.2$ ) in Fig. 8A and described in more detail in *SI Materials and Methods*, a population-weighted spectral envelope (dashed black line) was calculated from the sum of three basis spectra (colored lines) whose relative intensities and peak frequencies correspond to the fractional populations and calculated IR frequencies, respectively, for the three tautomers present at this  $pK_a$ . This modeled composite spectrum closely resembles the observed IR peak for this complex (Fig. 8A, solid black line), supporting assignment of the high-energy features resolved in spectra of M116C-CN bound to high  $pK_a$  phenols (peak shoulder in Fig. 7A and higher energy peak in Fig. 7C and D) to the population of ionized Y57. The spectral envelope for the stretching transition of each nitrile probe was calculated in this manner for a series of phenols with discrete  $pK_a$  values that span the experimental range.

A plot of the absorbance maximum of each modeled spectrum (colored) vs. the observed IR peak maximum (black) for each probe as a function of phenol  $pK_a$  is shown in Fig. 8B. The close correspondence between the predicted and observed IR shifts for M116C-CN and M105C-CN with varying phenol  $pK_a$  suggests that the electrostatic field changes due to charge rearrangement in the active site hydrogen bond network accurately account for the observed trends in IR peak maxima for these two probes.

In contrast to M116C-CN and M105C-CN, our modeling is unable to explain the observed IR frequency trend for F86C-CN. Our calculations predict a significant IR frequency dependence on ligand  $pK_a$ , but no systematic change is observed experimentally (Fig. 8B). This disagreement can be accounted for by a model in which F86C-CN has greater conformational freedom relative to the other two probe sites due to the steric vacancy generated by mutation of the bulky phenyl ring of the parent F86 residue to a thiocyanate. The additional space may permit this nitrile probe to reorient in response to changes in charge localization between discrete ionized tautomers present across the phenol series, an effect not captured by our calculations, which relied on a static structural model. Indeed, molecular dynamics simulations (described in detail in *SI Materials and Methods*) suggest that the nitrile of F86C-CN has substantially greater conformational mobility than nitriles at the other two sites and samples a wide distribution of rotamers on the nanosecond time scale (Fig. 8C and Fig. S8; additional discussion is provided in *SI Text*). These results are consistent with the model above that posits a wide equilibrium distribution of conformers for the F86C-CN nitrile. Nevertheless, recent time-resolved IR studies indicate that this probe does not substantially rearrange on the time scale of tens of picoseconds (21).

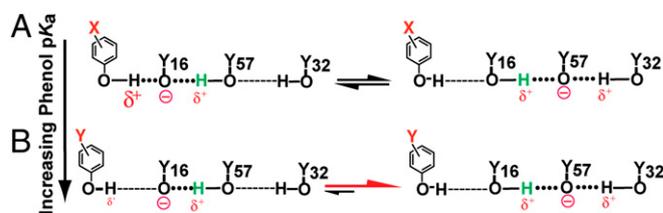
## Conclusions and Implications

Pauling and Corey recognized over 60 y ago that hydrogen bonds and the extended networks they frequently form are ubiquitous and central components of biological structure and function (46, 47). Hydrogen bonds are typically probed in a coarse fashion by ablating them via site-directed mutagenesis and evaluating the functional consequence of their removal. Although this approach can highlight their general functional importance, it does not reveal the physical properties of the intact hydrogen bonds that underpin their functional roles (8–10, 48), and the energetic effects of mutations can have as much to do with surrounding structural rearrangements in response to hydrogen bond ablation as they do with properties of the hydrogen bonds themselves (49–52). In contrast to these common mutagenic approaches, we have leveraged favorable features of KSI to interrogate the physical and energetic properties of the intact hydrogen bond network formed in the KSI active site and to study the effects of internal charge rearrangement on electrostatic fields within the active site.

**Electrostatic Effects of Charge Rearrangement Within the Active Site Hydrogen Bond Network.** A hallmark of protein catalysts and a distinguishing feature from reactions in bulk solution is that enzymes provide a highly structured and chemically heterogeneous solvation environment that has a limited ability for electrostatic rearrangement (1, 49, 53–55). Nevertheless, it has been a formidable challenge to carry out direct experimental tests of the nature and properties of this environment. Our nitrile probes, combined with our ability to reposition charge incrementally within the active site by varying the identity of bound phenols, provided a highly controlled system to test how specific charge rearrangements are sensed in discrete regions of an enzyme active site and how the surrounding protein matrix responds to this charge rearrangement.

Using continuum electrostatic calculations based on a static structural model and an internal dielectric of 2, we found surprisingly close agreement between the calculated and observed field changes for M116C-CN and M105C-CN, the two most ordered sites. This quantitative agreement suggests the absence of substantial electrostatic reorientation within the surrounding protein matrix in response to charge rearrangement within the hydrogen bond network. This finding is consistent with prior time-resolved studies of KSI that suggested a rigid electrostatic environment on the picosecond time scale (21, 56) and extends those observations to the equilibrium time scale. Structural comparison of KSI bound to low vs. high  $pK_a$  phenols (Fig. 2B) also suggested negligible structural rearrangement at equilibrium, but small rearrangements on the 0.1-Å scale would be difficult to detect reliably by X-ray crystallography even at the 1.3-Å resolution of our structures. Rearrangement of a single charged group by 0.1 Å, however, could be readily detected by IR spectroscopy as a  $2\text{-cm}^{-1}$  shift if movement of this group were parallel to a nitrile probe located  $\sim 4$  Å away, given an internal dielectric of 2.

The quantitative agreement between experiment and computation for M116C-CN and M105C-CN contrasts with prior studies of electrostatic changes in proteins due to pH changes, side-chain mutation, and ligand binding, in which qualitative agreement, at best, has been observed between experiment and theory (20, 44, 45, 57, 58). We hypothesize that our simple computational approach succeeded in the present study due to the subtle perturbations of the homologous series of bound phenols (differing only in their *meta* and *para* substituent groups) and to the absence of structural and solvent rearrangements that are likely to accompany more gross perturbations, such as pH changes, mutation, or protein–ligand association. This success also suggests that undetermined ionization states and rearrangements that are uncharacterized and difficult to model may be general problems that severely limit computational accuracy. Indeed, many studies have attempted to account for these unknown features by using a higher internal dielectric (e.g.,  $\epsilon = 20$ ) in continuum electrostatic calcu-



**Fig. 9.** Schematic model for preferential ionization of Y57 vs. Y16 with increasing phenol  $pK_a$ . Increasing the solution  $pK_a$  of the bound phenol going from A to B weakens its hydrogen bond to Y16, which destabilizes Y16 ionization and shifts the proton transfer equilibrium toward ionization of Y57, where charge is stabilized by hydrogen bonds from Y16 and Y32. For clarity, the transferred proton is shown in green and the D103–phenol hydrogen bond has been omitted.

lations (3–5). Our study provides an example in which a more controlled system with minimal structural rearrangements and incisive knowledge of ionization states may more cleanly isolate electrostatic effects and substantially improve computational accuracy. Nevertheless, we stress that continuum electrostatic models based on a uniform protein dielectric are an incomplete description of the heterogeneous and anisotropic protein environment and are inadequate to describe all the properties and behaviors of these environments. Indeed, the behavior of the F86C-CN probe appears to provide an example of this more complex behavior.

**Quantitative Dissection of Hydrogen Bond-Mediated Proton Transfer in the KSI Active Site.** Site-specific NMR probes allowed us to resolve ionization changes in the active site hydrogen bond network that were otherwise invisible in KSI X-ray structures with resolutions as high as 1.1–1.25 Å (8, 15). The ability to fit the ionization data with a simple model based only on the proton affinities of the interacting groups (Fig. 4A) suggests that these incremental ionization changes predominantly arise due to physical and energetic changes in the hydrogen-bonded groups themselves rather than from the propagated effects of conformational rearrangements of distal groups. Internal proton transfer is an expected property of hydrogen bonds formed between groups with similar proton affinities. Our observation of incremental proton transfer from Y16 to the bound phenolate as the proton affinity of the phenol approaches that of Y16 is consistent with extensive studies of hydrogen bond-mediated proton transfer based on small-molecule complexes in solution (19, 34, 42, 59) and is not a unique property of hydrogen bonds formed within the heterogeneous protein environment.

The systematic variation in relative ionized fractions of Y16 and Y57 is readily accounted for by a model in which the energetic stabilization of the Y16 anion provided by the phenol–Y16 hydrogen bond decreases with increasing phenol  $pK_a$ . Extensive prior hydrogen bond studies in small molecules provide strong evidence that increasing the  $pK_a$  of a phenol or other hydrogen bond donor (via substituent effects) weakens its ability to donate a hydrogen bond to a common acceptor of lower  $pK_a$  (17, 18, 32, 34, 60, 61) due, in part, to the decrease in positive charge character of the hydroxylic proton and its weaker interaction with an anionic acceptor (17, 60). Increasing phenol  $pK_a$  is also expected to lengthen the hydrogen bond to the Y16 anion as the proton affinities of the two groups become increasingly mismatched (8), reducing charge-transfer across the hydrogen bond and increasing negative charge localization on the Y16 anion (30, 31). These changes, which progressively weaken the phenol–Y16 hydrogen bond, destabilize the Y16 anion and thereby favor formation of the Y57 anion instead, which is stabilized by hydrogen bonds from Y16 and Y32 (Fig. 9).

This systematic and quantitative dissection of equilibrium proton transfer within the oxyanion hole hydrogen bond network of KSI provides one of the cleanest isolations and interrogations of specific hydrogen bond properties within a complex, heteroge-

neous, and highly idiosyncratic protein interior. Refinement of these measurements in KSI and determination of related measures in other proteins, along with additional tests in model systems, have the potential to provide basic insights into the energetic properties of hydrogen bonds underlying enzymatic catalysis and the fundamental properties of protein interiors.

## Materials and Methods

A full description of all experimental and computational methods and KSI semisynthesis is given in *SI Materials and Methods*, with additional detail and discussion in *SI Text* and Figs. S9–S11. KSI mutants were expressed and purified from *Escherichia coli* using published methods (8). Nitrile labeling and uniform  $^{13}\text{C}$ -Tyr incorporation were performed as previously described (20, 29, 37).  $^{19}\text{F}$  and  $^{13}\text{C}$  NMR spectra were acquired at 20 °C on 500- and 600-MHz (proton frequency) Varian UNITYINOVA NMR spectrometers using previously published methods (8, 9, 37). IR spectra were acquired at room temperature and at 80 K as previously reported (20, 29, 37). Cocrystals of pKSI D40N–3-F-4-NO<sub>2</sub>-phenolate and D40N/M116C-CN–equilenin were obtained at 20 °C using hanging drop vapor diffusion in accordance with previously published methods (8, 20). X-ray diffraction data were collected at the Stanford Synchrotron Radiation Laboratory and the Advanced Light Source (Lawrence Berkeley National Laboratory), and structure refinement was carried out as previously described (9, 20). Electrostatic calculations were performed with DelPhi (62), as previously published (20).

The  $X_i$  of each group in Fig. 4A was globally fit via nonlinear regression using GraphPad Prism to the following equilibrium titration expressions, in which  $pK_a^{\text{phenol}}$  was the independent variable, and apparent  $pK_a$  values, as defined above, were used to describe the relative proton affinities of the Tyr groups. In the expressions below,  $pK_a^{Y57}$  was fit as an adjustable parameter and  $pK_a^{Y16}$  was fit as a linear function of  $pK_a^{\text{phenol}}$  according to  $pK_a^{Y16} = a + b \times pK_a^{\text{phenol}}$ , with  $a$  and  $b$  as adjustable parameters. A full derivation and description of the fitting is given in *SI Materials and Methods*.

$$X_{\text{phenol}} = \frac{1}{1 + 10^{(pK_a^{\text{phenol}} - (a + b \times pK_a^{\text{phenol}}))} + 10^{(pK_a^{\text{phenol}} - pK_a^{Y57})}} \quad [1]$$

$$X_{Y16} = \frac{1}{1 + 10^{((a + b \times pK_a^{\text{phenol}}) - pK_a^{\text{phenol}})} + 10^{((a + b \times pK_a^{\text{phenol}}) - pK_a^{Y57})}} \quad [2]$$

$$X_{Y57} = \frac{1}{1 + 10^{(pK_a^{Y57} - pK_a^{\text{phenol}})} + 10^{(pK_a^{Y57} - (a + b \times pK_a^{\text{phenol}}))}} \quad [3]$$

**ACKNOWLEDGMENTS.** We thank Corey Liu and Steve Lynch for assistance with NMR experiments, Aaron Straight for reagents, Jessica DeMott for contributions to nitrile-labeling experiments, Eliza Reuben for assistance with QM calculations, and John Brauman and Peter Tolstoy for helpful discussions. Portions of this research were conducted at the Advanced Light Source, a national user facility operated by the Lawrence Berkeley National Laboratory; at the Stanford Synchrotron Radiation Laboratory, which is supported by the Department of Energy and the National Institutes of Health (NIH); and at the Stanford Magnetic Resonance Laboratory, which is supported, in part, by the Stanford University Medical School. P.A.S. was supported, in part, by Howard Hughes Medical Institute and G. Lieberman predoctoral fellowships. S.D.F. was supported, in part, by National Science Foundation (NSF) and Stanford Bio-X predoctoral fellowships. Funding was provided by grants to S.G.B. (NIH Grant GM27738) and D.H. (NSF Grant MCB-1121778).

- Fersht AR (1999) *Structure and Mechanism in Protein Science* (Freeman, New York).
- Ladd MFC, Palmer RA (2003) *Structure Determination by X-Ray Crystallography* (Kluwer Academic/Plenum, New York).
- Fitch CA, et al. (2002) Experimental  $pK_a$  values of buried residues: Analysis with continuum methods and role of water penetration. *Biophys J* 82(6):3289–3304.
- Antosiewicz J, McCammon JA, Gilson MK (1996) The determinants of  $pK_a$ s in proteins. *Biochemistry* 35(24):7819–7833.
- Karp DA, et al. (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys J* 92(6):2041–2053.
- Petrounia IP, Pollack RM (1998) Substituent effects on the binding of phenols to the D38N mutant of 3-oxo- $\Delta^5$ -steroid isomerase. A probe for the nature of hydrogen bonding to the intermediate. *Biochemistry* 37(2):700–705.
- Petrounia IP, Blotny G, Pollack RM (2000) Binding of 2-naphthols to D38E mutants of 3-oxo- $\Delta^5$ -steroid isomerase: Variation of ligand ionization state with the nature of the electrophilic component. *Biochemistry* 39(1):110–116.
- Kraut DA, et al. (2006) Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4(4):e99.
- 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.
- Sigala PA, Caaveiro JM, Ringe D, Petsko GA, Herschlag D (2009) Hydrogen bond coupling in the ketosteroid isomerase active site. *Biochemistry* 48(29):6932–6939.
- 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.
- Zhao Q, Abeygunawardana C, Gittis AG, Mildvan AS (1997) Hydrogen bonding at the active site of delta 5-3-ketosteroid isomerase. *Biochemistry* 36(48):14616–14626.
- Childs W, Boxer SG (2010) Proton affinity of the oxyanion hole in the active site of ketosteroid isomerase. *Biochemistry* 49(12):2725–2731.
- Kuliopulos A, Mildvan AS, Shortle D, Talalay P (1989) Kinetic and ultraviolet spectroscopic studies of active-site mutants of delta 5-3-ketosteroid isomerase. *Biochemistry* 28(1):149–159.
- Kim SW, et al. (1997) High-resolution crystal structures of delta 5-3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry* 36(46):14030–14036.

16. Gross KC, Seybold PG (2001) Substituent effects on the physical properties and pKa of phenol. *Int J Quantum Chem* 85:569–579.
17. Shan SO, Herschlag D (1999) Hydrogen bonding in enzymatic catalysis: Analysis of energetic contributions. *Methods Enzymol* 308:246–276.
18. Shan SO, Loh S, Herschlag D (1996) The energetics of hydrogen bonds in model systems: Implications for enzymatic catalysis. *Science* 272(5258):97–101.
19. Brycki B, Brzezinski B, Zundel G, Keil T (1992) <sup>1</sup>H and <sup>13</sup>C NMR studies of the proton transfer in complexes of substituted phenols with trimethylamine N-oxide. *Magn Reson Chem* 30(6):507–510.
20. 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.
21. 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.
22. 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.
23. Gerig JT (1989) Fluorine nuclear magnetic resonance of fluorinated ligands. *Methods Enzymol* 177:3–23.
24. 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.
25. Wu ZR, et al. (1997) Solution structure of 3-oxo-delta5-steroid isomerase. *Science* 276(5311):415–418.
26. Thornburg LD, et al. (1998) Electrophilic assistance by Asp-99 of 3-oxo-Delta 5-steroid isomerase. *Biochemistry* 37(29):10499–10506.
27. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1996) *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego).
28. Ha NC, Kim MS, Lee W, Choi KY, Oh BH (2000) Detection of large pKa perturbations of an inhibitor and a catalytic group at an enzyme active site, a mechanistic basis for catalytic power of many enzymes. *J Biol Chem* 275(52):41100–41106.
29. 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.
30. Hibbert F, Emsley J (1990) Hydrogen bonding and chemical reactivity. *Advances in Physical Organic Chemistry* 26:255–379.
31. Jeffrey GA (1997) *An Introduction to Hydrogen Bonding* (Oxford Univ Press, New York).
32. 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.
33. 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.
34. Stahl N, Jencks WP (1986) Hydrogen bonding between solutes in aqueous solution. *J Am Chem Soc* 108(14):4196–4205.
35. 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.
36. Suydam IT, Boxer SG (2003) Vibrational Stark effects calibrate the sensitivity of vibrational probes for electric fields in proteins. *Biochemistry* 42(41):12050–12055.
37. 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.
38. Gross KC, Waybold PG, Hadad CM (2002) Comparison of different atomic charge schemes for predicting pKa variations in substituted anilines and phenols. *Int J Quantum Chem* 90:445–458.
39. Cleland WW, Frey PA, Gerit JA (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 273(40):25529–25532.
40. Cleland WW (2010) The low-barrier hydrogen bond in enzymatic catalysis. *Advances in Physical Organic Chemistry* 44:1–17.
41. de la Vega JR (1982) Role of symmetry in the tunneling of the proton in double minimum potentials. *Acc Chem Res* 15(41):185–191.
42. Koeppel B, Tolstoy PM, Limbach HH (2011) Reaction pathways of proton transfer in hydrogen-bonded phenol-carboxylate complexes explored by combined UV-vis and NMR spectroscopy. *J Am Chem Soc* 133(20):7897–7908.
43. Perrin CL (2010) Are short, low-barrier hydrogen bonds unusually strong? *Acc Chem Res* 43(12):1550–1557.
44. Suydam IT, Snow CD, Pande VS, Boxer SG (2006) Electric fields at the active site of an enzyme: Direct comparison of experiment with theory. *Science* 313(5784):200–204.
45. Fafarman AT, Boxer SG (2010) Nitrile bonds as infrared probes of electrostatics in ribonuclease S. *J Phys Chem B* 114(42):13536–13544.
46. Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* 37(4):205–211.
47. Pauling L, Corey RB (1950) Two hydrogen-bonded spiral configurations of the polypeptide chain. *J Am Chem Soc* 72(11):5349.
48. Sigala PA, Tsuchida MA, Herschlag D (2009) Hydrogen bond dynamics in the active site of photoactive yellow protein. *Proc Natl Acad Sci USA* 106(23):9232–9237.
49. Kraut DA, Carroll KS, Herschlag D (2003) Challenges in enzyme mechanism and energetics. *Annu Rev Biochem* 72:517–571.
50. 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.
51. Plapp BV (1995) Site-directed mutagenesis: A tool for studying enzyme catalysis. *Methods Enzymol* 249:91–119.
52. Fersht AR (1988) Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. *Biochemistry* 27(5):1577–1580.
53. Jencks WP (1987) *Catalysis in Chemistry and Enzymology* (Dover, New York), Ed 2.
54. Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News* 24(10):1375–1377.
55. Warshel A, Aqvist J, Creighton S (1989) Enzymes work by solvation substitution rather than by desolvation. *Proc Natl Acad Sci USA* 86(15):5820–5824.
56. Childs W, Boxer SG (2010) Solvation response along the reaction coordinate in the active site of ketosteroid isomerase. *J Am Chem Soc* 132(18):6474–6480.
57. 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(6):1588–1598.
58. Xu L, Cohen AE, Boxer SG (2011) Electrostatic fields near the active site of human aldose reductase: 2. New inhibitors and complications caused by hydrogen bonds. *Biochemistry* 50(39):8311–8322.
59. Zundel G (2000) Hydrogen bonds with large proton polarizability and proton transfer processes in electrochemistry and biology. *Adv Chem Phys* 111:1–217.
60. Hine J (1972) Hydrogen-bonded intermediates and stepwise mechanisms for proton-exchange reactions between oxygen atoms in hydroxylic solvents. *J Am Chem Soc* 94(16):5766–5771.
61. Gilli P, Pretto L, Bertolasi V, Gilli G (2009) Predicting hydrogen-bond strengths from acid-base molecular properties. The pKa(a) slide rule: Toward the solution of a long-lasting problem. *Acc Chem Res* 42(1):33–44.
62. Gilson MK, Sharp K, Honig B (1988) Calculating the electrostatic potential of molecules in solution—Method and error assessment. *J Comput Chem* 9(4):327–335.

# 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  $16,960 \text{ M}^{-1}\cdot\text{cm}^{-1}$  (5).

**3.  $^{13}\text{C}$ -Tyr Labeling of Recombinant and Semisynthetic KSI.** *Pseudomonas putida* KSI (pKSI) D40N containing  $^{13}\text{C}_\alpha$ -Tyr labels at all four Tyr residues was prepared as previously reported (4). Semisynthetic pKSI D40N (also containing the R15K/D21N/D24C mutations) containing  $^{13}\text{C}_\alpha$ -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  $^{13}\text{C}_\alpha$ -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<sub>6</sub> 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 DH5 $\alpha$  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  $\beta$ -mercapto-propionyl-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  $^{13}\text{C}_\alpha$ -Tyr labeled at Y32, Y57, and Y119.** The  $^{13}\text{C}$ -Tyr-labeled fusion protein with Y32, Y57, and Y119  $^{13}\text{C}_\alpha$ -labeled was expressed in BL21(DE3) cells grown in M9 minimal media supplemented with L-Tyr [50 mg/L phenol-4- $^{13}\text{C}$  (95–99%); Cambridge Isotope Laboratories, Inc.] and the remaining 19 unlabeled amino acids. Cells were grown at 37 °C to an OD of  $\sim 0.6$ , followed

by the addition of 0.4 mM isopropyl- $\beta$ -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  $8,000 \times g$ . 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  $\sim 0$  ( $\sim 10$  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  $\mu\text{M}$ . 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  $\sim 2$  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  $^{13}\text{C}_\alpha$ -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  $\sim 4$  mM. The peptide and recombinant fragments were combined to give final concentrations of  $\sim 4$  mM and  $\sim 2$  mM, respectively, in 7 M urea and 20 mM sodium phosphate (pH 7.2). Sodium 4-mercapto-phenylacetic 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.**  $^{19}\text{F}$  and  $^{13}\text{C}$  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  $^{13}\text{C}$  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%  $\text{D}_2\text{O}$  as the lock solvent. Samples for  $^{19}\text{F}$  detection contained 1.5 mM KSI, 1 mM ligand, 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5%  $\text{D}_2\text{O}$  as the lock solvent.  $^{19}\text{F}$  NMR chemical shifts were referenced to an external sample of TFA in buffer (−76.1 ppm relative to  $\text{CFCl}_3$ ).  $^{13}\text{C}$  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  $\pm 0.1$  ppm based on replicate measurements of identical samples.

For all phenol complexes studied, the limiting component (protein in  $^{13}\text{C}$  spectra or ligand in  $^{19}\text{F}$  spectra) was  $\geq 95\%$  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 ( $\geq 95\%$ ) bound KSI ( $^{13}\text{C}$  spectra) or phenol ( $^{19}\text{F}$  NMR), and any chemical exchange with the minor population ( $\leq 5\%$ ) of the dissociated species does not contribute significantly to the observed peak positions. This conclusion is supported by observation of identical  $^{13}\text{C}$ -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  $^{13}\text{C}$  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  $\text{CaF}_2$  sample cell with 12- to 23- $\mu\text{m}$  spacers as previously described (11).

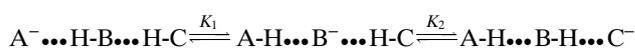
**6. X-Ray Crystallography.** Cocrystals of pKSI D40N–3-F-4- $\text{NO}_2$ -phenolate and D40N/M116C-CN–equilenin were obtained at 20 °C using hanging drop vapor diffusion according to previously

published crystallization conditions in ammonium sulfate and potassium phosphate buffer (1, 4, 9, 13). X-ray diffraction data for D40N–3-F-4- $\text{NO}_2$ -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- $\text{NO}_2$ -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  $\text{pK}_a$ .** Fractional ionizations of the bound phenol, Y16, and Y57 were calculated as shown in Fig. 4A, 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  $\text{pK}_a$  value is equal to 1; (iii) the chemical shifts observed for Y16 and Y57 in the KSI complex with 4-nitrophenol ( $\text{pK}_a = 7.1$ ), which is bound as the fully ionized phenolate (1, 11, 15) (Fig. 2A), 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  $\text{pK}_a$  values ( $X_i^{Y^-}$ ) were calculated as  $X_i^{Y^-} = \frac{\delta_i - \delta_{7.1}}{166 - \delta_{7.1}}$ , where  $\delta_i$  is the observed chemical shift at each  $\text{pK}_a$  value and  $\delta_{7.1}$  is the chemical shift with bound 4-nitrophenol. The fractional phenol ionization at each  $\text{pK}_a$  value ( $X_i^{\text{Ph}^-}$ ) was calculated as  $X_i^{\text{Ph}^-} = 1 - (X_i^{Y16^-} + X_i^{Y57^-})$ . For  $\text{pK}_a$  values  $\geq 10.4$ , where  $X_i^{Y16^-} + X_i^{Y57^-}$  calculated as above gave a sum that was greater than 1,  $X_i^{\text{Ph}^-}$  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  $X_i^{\text{Ph}^-}$  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  $^{19}\text{F}$  NMR changes observed for bound 4-F-substituted phenols in Fig. 2A 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. S7B). We estimate the uncertainty in our fractional ionization estimates for each point in Fig. 4A to be  $\pm 0.15$ , based on the 2.5- to 3-ppm difference in the  $^{13}\text{C}$  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  $^{13}\text{C}$  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  $\text{pK}_a$  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  $\text{pK}_a$  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  $\text{pK}_a$  values for these groups within the phenol-bound KSI active site:



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

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

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

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

$$\begin{aligned} X_{A^-} &= \frac{[A^-]}{[A^-] + [B^-] + [C^-]} = \frac{1}{1 + \frac{[B^-]}{[A^-]} + \frac{[C^-]}{[A^-]}} \\ &= \frac{1}{1 + 10^{(pK_a^A - pK_a^B)} + 10^{(pK_a^A - pK_a^C)}} \end{aligned} \quad [S5]$$

$$\begin{aligned} X_{B^-} &= \frac{[B^-]}{[A^-] + [B^-] + [C^-]} = \frac{1}{1 + \frac{[A^-]}{[B^-]} + \frac{[C^-]}{[B^-]}} \\ &= \frac{1}{1 + 10^{(pK_a^B - pK_a^A)} + 10^{(pK_a^B - pK_a^C)}} \end{aligned} \quad [S6]$$

$$\begin{aligned} X_{C^-} &= \frac{[C^-]}{[A^-] + [B^-] + [C^-]} = \frac{1}{1 + \frac{[A^-]}{[C^-]} + \frac{[B^-]}{[C^-]}} \\ &= \frac{1}{1 + 10^{(pK_a^C - pK_a^A)} + 10^{(pK_a^C - pK_a^B)}} \end{aligned} \quad [S7]$$

We first fit the data in Fig. 4A 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. S4A) 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. S4B).

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. 4A and accurately accounted for the observed changes:

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

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

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

**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<sub>2</sub>-phenolate, 3,4-(NO<sub>2</sub>)<sub>2</sub>-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<sub>2</sub>-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 ( $\epsilon$ ) 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  $50 \times 50 \times 50$  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,  $\text{cm}^{-1}$ ), 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  $0.65 \text{ cm}^{-1}/(\text{MV}/\text{cm})$  (4) to convert the electric field changes into vibrational frequency changes (in  $\text{cm}^{-1}$ ). 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. 4A and summed together as in Fig. 8A. 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. 8B.

**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. 4A), 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  $\text{cm}^{-1}$  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  $\text{cm}^{-1}$  per  $pK_a$  ( $R^2 = 0.55$ ), which is similar to the experimental slope of 0.3  $\text{cm}^{-1}$  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  $-0.3 \text{ cm}^{-1}$  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  $-0.009 \text{ cm}^{-1}$  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 parameterization 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 ( $\angle C_\alpha-C_\beta-S-CN$ ) 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  $\Delta t$ , will have the same value, was calculated using the GROMACS utility “g\_angle” and plotted as a function of  $\Delta t$  in Fig. 8C. 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 $^{13}\text{C}$ -Tyr Chemical Shift Changes Due to Hydrogen Bond Shortening and Ionization

Prior  $^1\text{H}$  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  $C_\zeta$ -Tyr nucleus. Such deshielding can be  $\sim 1$  ppm in magnitude, based on  $^{13}\text{C}$  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  $\geq 160$  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. 2A and Fig. S3. The  $\sim 1$ -ppm down-field shifts for the  $^{13}\text{C}$  peak of Y32 with bound 4-Andro and 5-Andro (Fig. 3D) 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  $\text{cm}^{-1}$  in the stretch frequency observed across the three nitrile probe sites (Fig. 6A). 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  $^{13}\text{C}$  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  $^{13}\text{C}$  NMR spectra of this variant bearing a  $^{13}\text{C}$ -labeled nitrile ( $-^{13}\text{CN}$ ) bound to the same series of phenols. The  $^{13}\text{C}$  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. 6B), 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  $^{13}\text{C}$  NMR frequencies on phenol  $pK_a$ . Understanding the physical origin of these differences will require a deeper understanding of the contributions to  $^{13}\text{C}$  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  $\Delta E$  is the energy difference between two states,  $\Delta t$  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  $4 \text{ cm}^{-1}$  or  $1 \times 10^{11} \text{ Hz}$  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  $10^{11} \text{ Hz}$  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  $^{13}\text{C}$ -Tyr NMR peaks for Y16 and Y57 with bound phenols of  $\text{pK}_a$  of  $\sim 10$  (Fig. 3), rather than observation of discrete peaks for the interconverting neutral and ionized states of both Y16 and Y57 populated at this  $\text{pK}_a$  (Fig. 4B), 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  $\sim 8$ -ppm chemical shift difference between a neutral (157 ppm) and ionized (165 ppm) Tyr residue measured on a 125-MHz ( $^{13}\text{C}$  frequency) NMR spectrometer, which corresponds to a frequency difference of  $10^3 \text{ Hz}$  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  $\text{pK}_a$  near 10 occurs faster than  $10^3 \text{ s}^{-1}$  but slower than  $10^{11} \text{ s}^{-1}$ .

### E. Additional Discussion of MD Simulations

The dihedral angle  $\angle \text{C}_\alpha\text{-C}_\beta\text{-S-C}$ , 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. 8C) 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. 8C and Fig. S8).

In Fig. 8C, the  $\text{C}_\alpha\text{-C}_\beta\text{-S-CN}$  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  $\text{pK}_a$ . 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  $\text{pK}_a$ , 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  $\text{pK}_a$ . 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.

- Kraut DA, et al. (2006) Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol* 4(4):e99.
- 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.
- 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.
- 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.
- Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 182(2):319–326.
- Kent SB (1988) Chemical synthesis of peptides and proteins. *Annu Rev Biochem* 57:957–989.
- 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.
- 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.
- 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.
- 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.
- 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.
- Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1996) *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego).
- 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.
- DeLano WL (2007) *MacPyMOL: A PyMOL-Based Molecular Graphics Application for Mac OS X* (DeLano Scientific LLC, Palo Alto, CA).
- Petrousis IP, Pollack RM (1998) Substituent effects on the binding of phenols to the D38N mutant of 3-oxo- $\Delta^5$ -steroid isomerase. A probe for the nature of hydrogen bonding to the intermediate. *Biochemistry* 37(2):700–705.
- 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.
- Childs W, Boxer SG (2010) Proton affinity of the oxyanion hole in the active site of ketosteroid isomerase. *Biochemistry* 49(12):2725–2731.
- Hibbert F, Emsley J (1990) Hydrogen bonding and chemical reactivity. *Adv Phys Org Chem* 26:255–379.
- Jeffrey GA (1997) *An Introduction to Hydrogen Bonding* (Oxford Univ Press, New York).
- 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.
- Steiner T, Saenger W (1994) Lengthening of the covalent O-H bond in O-H $\cdots$ O hydrogen bonds re-examined from low-temperature neutron diffraction data of organic compounds. *Acta Crystallogr B* 50:348–357.
- Shan SO, Herschlag D (1999) Hydrogen bonding in enzymatic catalysis: Analysis of energetic contributions. *Methods Enzymol* 308:246–276.
- 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 pKa 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)  $^1\text{H}$  and  $^{13}\text{C}$  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.

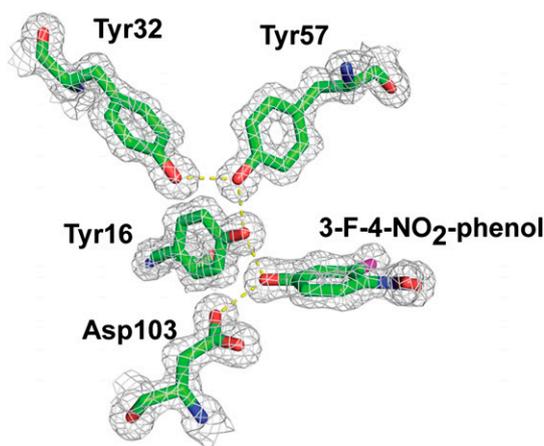


Fig. S1.  $2F_o - F_c$  electron density map (contoured at  $1.5 \sigma$ ) for the 1.30-Å resolution structure of 3-F-4- $\text{NO}_2$ -phenol bound to pKSI D40N.

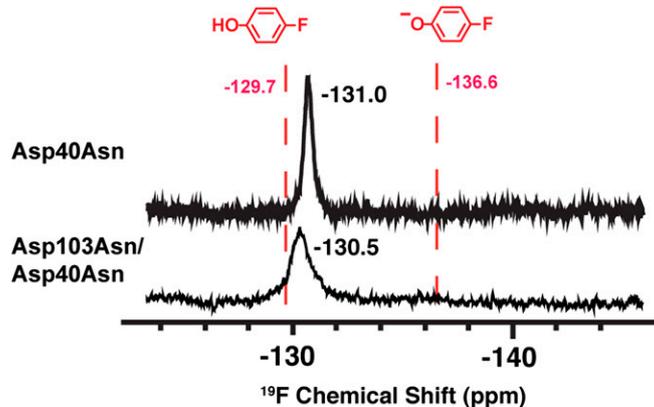
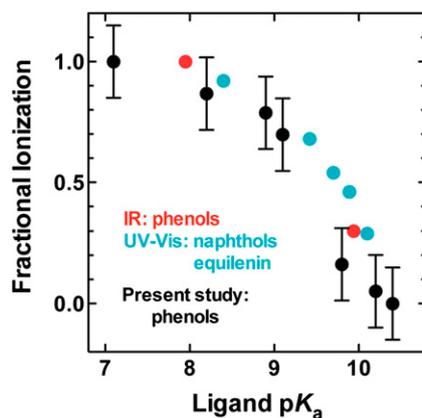
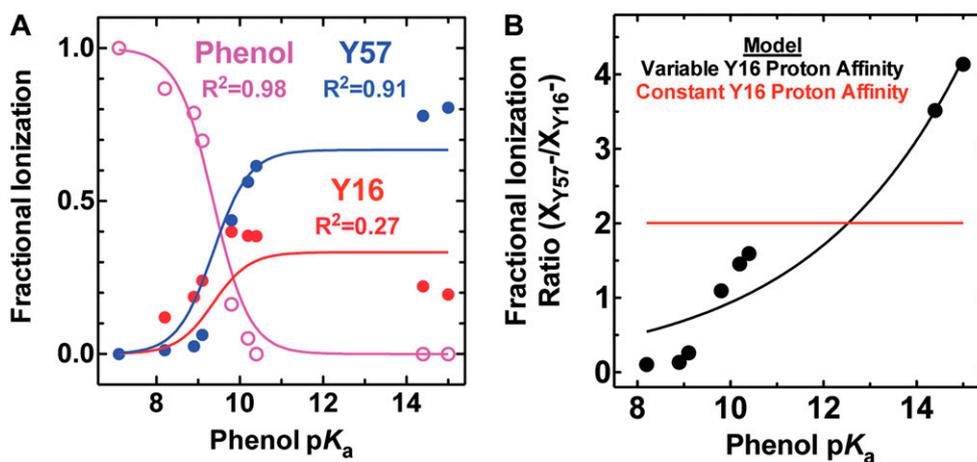


Fig. S2.  $^{19}\text{F}$  NMR spectrum of 1 mM 4-F-3-Me-phenol bound to 1.5 mM pKSI D40N or D103N/D40N in 40 mM potassium phosphate (pH 7.2) and 1 mM EDTA.



**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  $^{13}\text{C}$ -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  $^{13}\text{C}$  NMR that this ratio increases with increasing phenol  $pK_a$ . A model with variable Y16 proton affinity (black) can account for the increase in  $X_{Y57-}/X_{Y16-}$  ( $R^2 = 0.92$ ).

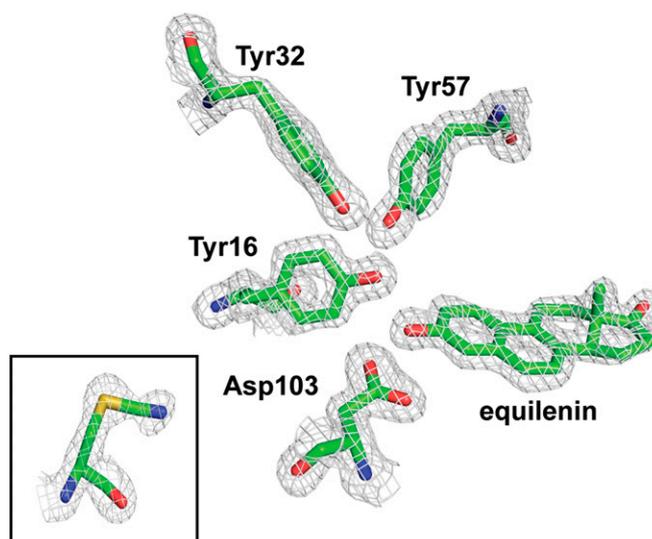


Fig. S5.  $2F_o-F_c$  electron density map (contoured at  $1.5\sigma$ ) for the 1.7-Å resolution structure of equilenin bound to D40N/M116C-CN. (Inset) For clarity, the M116C-CN group was removed from the view shown for the other groups and is shown in the box.

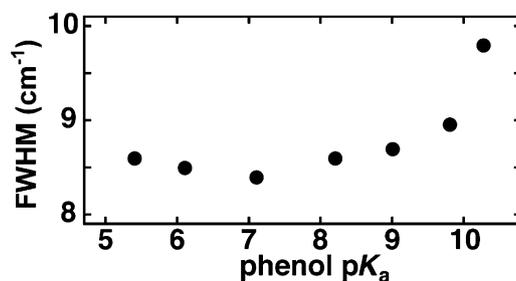


Fig. S6. Observed nitrile IR peak width (FWHM) for D40N/M116C-CN bound to phenols of increasing pK<sub>a</sub> (data are from Table S2).

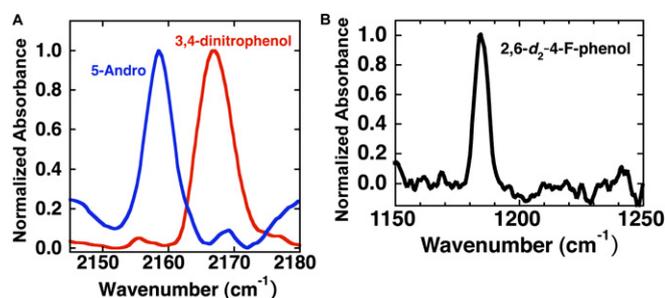


Fig. S7. (A) Low-temperature (80 K) IR spectra of the nitrile stretch for M116C-CN/D40N bound to 5-Andro (blue, pK<sub>a</sub> = 15) or 3,4-dinitrophenol (red, pK<sub>a</sub> = 5.4). (B) IR spectrum at 80 K of the C-F stretch for 2,6-*d*<sub>2</sub>-4-F-phenol bound to M116C-CN/D40N. The single C-F peak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

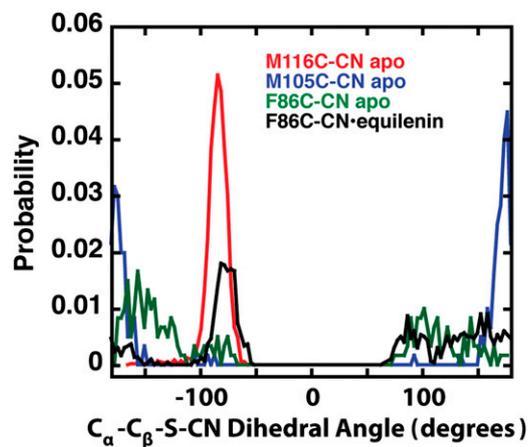


Fig. 58. Distribution of the  $C_{\alpha}$ - $C_{\beta}$ -S-CN dihedral angle for each probe during MD simulations of KSI-CN.

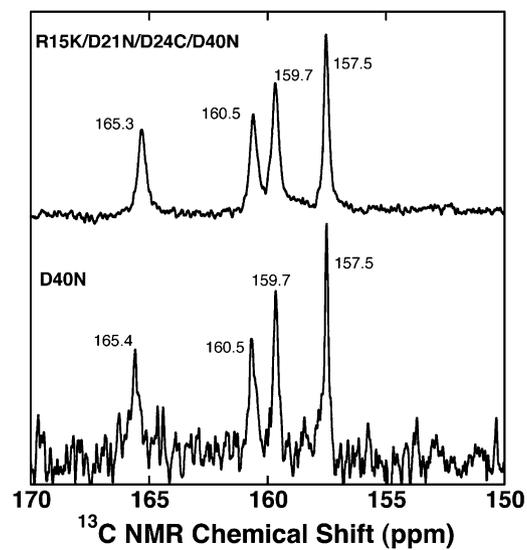
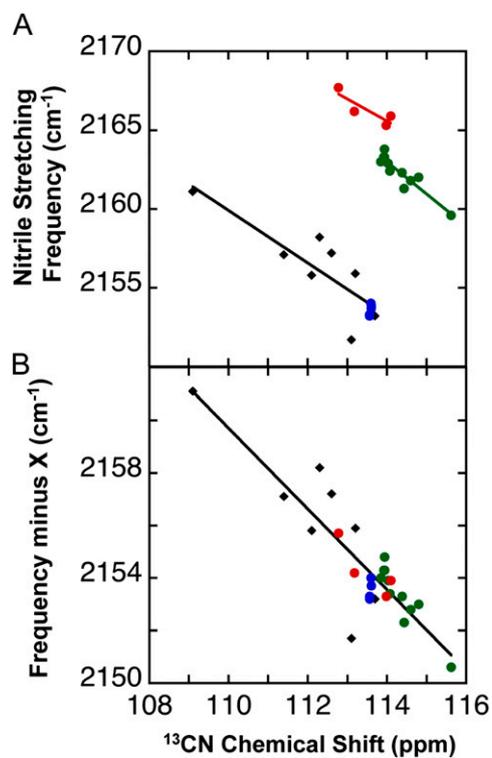
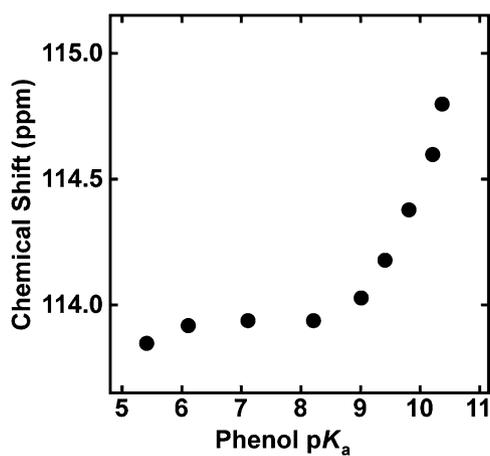


Fig. 59.  $^{13}\text{C}$  NMR spectra of unliganded pKSI D40N and D40N/R15K/D21N/D24C containing  $^{13}\text{C}_{\gamma}$ -Tyr labels at all four Tyr residues.



**Fig. S10.** Nitrile stretching frequency vs.  $^{13}\text{C}$  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  $13\text{ cm}^{-1}$  (F86C-CN) or  $10\text{ cm}^{-1}$  (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.**  $^{13}\text{C}$  NMR chemical shift of the nitrile of D60N/M116C-CN bound to phenols of differing  $\text{pK}_a$ .

