Understanding the electrostatic forces and features within highly heterogeneous, anisotropic, and chemically complex enzyme active sites and their connection to biological catalysis remains a long-standing challenge, in part due to the paucity of incisive experimental probes of electrostatic properties within proteins. To quantitatively assess the landscape of electrostatic fields at discrete locations and orientations within an enzyme active site, we have incorporated site-specific thiocyanate vibrational probes into multiple positions within bacterial ketosteroid isomerase. A battery of X-ray crystallographic, vibrational Stark spectroscopy, and NMR studies revealed electrostatic field heterogeneity of 8 MV/cm between active site probe locations and widely differing sensitivities of discrete probes to common electrostatic perturbations from mutation, ligand binding, and pH changes. Electrostatic calculations based on active site ionization states assigned by literature precedent and computational pK_a predictions were unable to quantitatively account for the observed vibrational band shifts. However, electrostatic models of the D40N mutant gave qualitative agreement with the observed vibrational effects when an unusual ionization of an active site tyrosine with a pK_a near 7 was included. UV-absorbance and 13C NMR experiments confirmed the presence of a tyrosinate in the active site, in agreement with electrostatic models. This work provides the most direct measure of the heterogeneous and anisotropic nature of the electrostatic environment within an enzyme active site, and these measurements provide incisive benchmarks for further developing accurate computational models and a foundation for future tests of electrostatics in enzymatic catalysis.

Numerous biophysical studies have provided important qualitative insights into the differences between solvation environments sampled within proteins versus bulk water or organic solvents. These studies have inferred electrostatic potentials based on free energy measurements [e.g., pK_a shifts of ionizable residues (9–11), reactivity differences among cysteines (12, 13), equilibrium binding constants of charged ligands (14), and redox potential shifts (15)] or spectroscopic variations in probes incorporated into or bound to proteins [e.g., fluorescent dyes (16), 13C and 15N chemical shifts (17–21), and EPR-measured g-factor perturbations of nitroxides (22)]. These studies have spurred the ongoing development of electrostatic calculation methodology (6, 7) and contributed greatly to our general recognition that the solvation environment within proteins is very different from water. However, each of these experimental observables is influenced by additional environmental factors beyond electric field effects, without independent experimental means to distinguish these contributions (8, 12, 23). For example, local electrostatic fields can perturb pK_a values, but the observed pK_a is also sensitive to the making or breaking of hydrogen bonds between the titrating group and nearby groups and to the access of the site to water. These pervasive limitations have precluded accurate quantitative measurement of electrostatic fields in proteins, how such fields vary as a function of position and orientation within enzyme active sites, and how local electrostatic perturbations influence local fields.

Vibrational probes incorporated into proteins at defined positions can provide sensitive and directional reporters of the local electrostatic environment. Nitrile (-CN) groups provide a particularly incisive and powerful probe. Their small size allows incorporation with minimal structural perturbation; they can be easily introduced at discrete and well-defined positions within proteins; their strong absorbance bands lie in an uncluttered region of the IR spectrum; and they are sensitive to both the magnitude and direction of the local electric field (24). Furthermore, they have incorporated site-specific thiocyanate vibrational probes into multiple positions within bacterial ketosteroid isomerase. A battery of X-ray crystallographic, vibrational Stark spectroscopy, and NMR studies revealed electrostatic field heterogeneity of 8 MV/cm between active site probe locations and widely differing sensitivities of discrete probes to common electrostatic perturbations from mutation, ligand binding, and pH changes. Electrostatic calculations based on active site ionization states assigned by literature precedent and computational pK_a predictions were unable to quantitatively account for the observed vibrational band shifts. However, electrostatic models of the D40N mutant gave qualitative agreement with the observed vibrational effects when an unusual ionization of an active site tyrosine with a pK_a near 7 was included. UV-absorbance and 13C NMR experiments confirmed the presence of a tyrosinate in the active site, in agreement with electrostatic models. This work provides the most direct measure of the heterogeneous and anisotropic nature of the electrostatic environment within an enzyme active site, and these measurements provide incisive benchmarks for further developing accurate computational models and a foundation for future tests of electrostatics in enzymatic catalysis.
In contrast to many of the preceding examples, the field sensitivity of the nitrile stretching frequency can be calibrated in situ, obviating the need for a computational model or calibration in terms of an equivalent solution dielectric value that may not be transferable from bulk solvent to the heterogeneous structure of proteins.

The probe’s electric field sensitivity is largely due to the anharmonicity of the nitrile bond that leads to an increase in dipole moment, \( \Delta \mu_{\text{probe}} \), between the ground and first excited vibrational state. This difference dipole moment results in a dependence of the transition energy on the size and direction of the local electrostatic field (Fig. S1 and SI Text). To quantify the electric field sensitivity of a specific vibrational probe, a known external electric field is applied and the effect of the field on the IR absorption spectrum is measured (25–27), a method known as vibrational Stark effect (VSE) spectroscopy. The electric field sensitivity of a vibrational probe directly calibrated in this fashion is called the linear Stark tuning rate, \(~\Delta \mu_{\text{probe}} / \Delta F_{\text{protein}}~\) [in units of \( \text{cm}^{-1}/(\text{MV/cm}) \)].

Eq. 1 can then be employed to translate observed IR frequency shifts, \(~\Delta \nu_{\text{obs}}~\) (in \( \text{cm}^{-1} \)), into changes in the local protein electrostatic field, \(~\Delta F_{\text{protein}}~\) (in \( \text{MV/cm} \)), experienced by a probe at two different sites in a protein or as a result of a side-chain ionization, mutation or ligand binding:

\[
\hbar c \Delta \nu_{\text{obs}} = -\Delta \mu_{\text{probe}} \cdot \Delta F_{\text{protein}} = -|\Delta \mu_{\text{probe}}| |\Delta F_{\text{protein}}| \cos \theta \tag{1}
\]

where \( h \) is Planck’s constant, \( c \) the speed of light, and \( \theta \) is the angle between the vectors defined by \( \Delta \mu_{\text{probe}} \) and \( \Delta F_{\text{protein}} \). The direction of the vibrational probe \( \Delta \mu_{\text{probe}} \) is typically along the oscillator bond axis for simple, linear nitriles (27). Thus, X-ray crystallography can provide the orientation of \( \Delta \mu_{\text{CN}} \) within a nitrile-modified protein and the VSE in an external field calibrates the magnitude of \( \Delta \mu_{\text{CN}} \). Finally, although specific chemical interactions, such as hydrogen bonds, between nitrile probes and the surrounding environment can shift the nitrile IR frequency via a nonelectrostatic mechanism not captured by Eq. 1, we previously published an experimental method to detect and quantitatively correct for such effects (28). Thus, nitrile probes can provide quantitative measures of local electrostatic fields in the presence of specific interactions with surrounding groups.

Bacterial ketosteroid isomerase (KSI) has provided a powerful model system for probing fundamental aspects of enzymatic catalysis. The basic catalytic mechanism (Fig. 1B) is well established, many high resolution structures are available, good transition state analogs exist, and the importance of electrostatics to catalysis is the subject of extensive and ongoing study (29–32). In earlier work, we demonstrated that vibrational probes could be introduced into the active site in proximity to key catalytic residues (28, 33). We now report detailed studies that use vibrational probes to measure the differences in local electrostatic fields sampled at discrete positions within the active site and to test the sensitivity of local fields to nearby electrostatic perturbations from ligand binding, pH changes, and mutation. Using Eq. 1, we make direct comparisons between observed and computed changes in the electrostatic potential. Discrepancies, even at a qualitative level, between the predicted and observed effects led us to reconsider the ionization states of key residues. Our results provide quantitative insight into the fundamental electrostatic properties of a protein interior, identify unexpected features of the KSI active site, establish a benchmark for future computations, and provide a basis for designing future tests of electrostatic contributions to catalysis.

**Results**

**Incorporation of Nitrile Electric Field Probes into Ketosteroid Isomerase.** Bacterial KSI from *Pseudomonas putida* catalyzes double-bond isomerization in steroids using a general base, D40, to deprotonate the substrate and form a dienolate reaction intermediate. This intermediate is stabilized by hydrogen bonds formed to Y16 and protonated D103 within an active site oxygen hole that is linked via a hydrogen bond network to Y57 and Y32 (Fig. 1B) (34–37). We incorporated thiocyanate vibrational probes into KSI from *P. putida* at positions M116, M105, and F86...
(Fig. 1C). These positions were selected on the basis of their proximity (within 3–11 Å) to the key catalytic residues listed above. Sites M116 and M105 were used in prior investigations of the effects of solvation on active site electrostatic properties (28, 33). In the analysis that follows, we also refer to data for a thioxyanate probe incorporated at residue L61 obtained in a previous study (28). This residue is located near the mouth of the active site and forms part of the solvent-accessible surface of the steroid binding pocket. Thiocyanate labeling of KSI was accomplished by engineering single cysteine mutations at each of the above positions in the cysteine-free variant C69S/C81S/C97S to ensure unique labeling. The Cys residues at positions 69, 81, and 97 are located on the enzyme’s surface, and their removal has a negligible effect on catalysis (38).

The single Cys-SH group introduced in each mutant was converted to Cys-S-CN using previously published methods (33, 39), and the individual probe-labeled enzymes are referred to as M116C-CN, M105C-CN, F86C-CN, and L61C-CN. Labeled KSI variants were prepared with the general base D40 intact or mutated to D40N to mimic the protonated D40 present in the dienolate intermediate complex (Fig. 1B). Functional assays with the thioxyanate-modified enzymes revealed activities (kcat/KM) within three- to sixfold of Cys-free KSI with D40 present or affinity within twofold for binding of the transition state analog within three- to sixfold of Cys-free KSI with D40 present or affinity within twofold for binding of the transition state analog equilenin to the D40N mutants, suggesting minimal perturbation to KSI structure and function.

X-Ray Crystal Structures of KSI-CN Variants. We determined X-ray crystal structures for the D40N variants of M116C-CN, M105C-CN, and F86C-CN with or without the bound steroid transition state analog equilenin (Table 1, complete data collection and refinement statistics for each structure are listed in Table S1). The overall structures obtained for these variants were the same as those determined previously for P. putida KSI bound to equilenin (1OH0) or phenol (2PZV), with average r.m.s.d values of 0.26 Å, 0.31 Å, and 0.31 Å, respectively, for main-chain atoms of the apo M116C-CN, M105C-CN•equilenin, and F86C-CN•equilenin from the previously determined structures (which have a main-chain r.m.s.d of 0.27 Å from each other). Thus, thiocyanate incorporation does not perturb the global structure of KSI.

Focusing on the nitrile probes, the electron-density maps determined for each structure (Fig. S2) supported refinement of the thioxyanate group to a single well-ordered conformation (Fig. 1C). Superposition of active site residues observed in the probe modified versus unmodified KSI structures revealed conformational differences on the same scale as the 0.1–0.2 Å estimated coordinate uncertainty in these structures, indicating minimal structural rearrangements within the active site in response to probe insertion. The thioxyanate group in both the M116C-CN and M105C-CN structures refined to a position very similar to that occupied by the parent methionine residue in the unmodified structures, indicating a sterically conservative substitution. For F86C-CN the thioxyanate group refined to a position distinct from that occupied by the phenyl ring with F86 present. Nevertheless, we did not observe structural rearrangement of nearby groups to fill this space, nor did this change perturb the position of bound equilenin. Finally, comparison of F86C-CN structures with and without bound equilenin revealed identical thioxyanate positions, within error. We conclude that thioxyanate incorporation at each of these three positions minimally perturbs the structure and function of the KSI active site and results in ordered and well-defined nitrile probe conformations.

Each of the three probes is positioned within a unique local environment. The nitrile of M105C-CN is buried well below the solvent-accessible surface of the KSI active site and is located 6 to 7 Å from the polar side-chain hydroxyl moieties of Tyr 16, 32 and 57, Asp 103, and the oxyanion of the bound steroid (Fig. 1C). Its immediate solvation environment is thus composed of a tightly packed shell of hydrophobic aliphatic and aromatic residues (Fig. 1D).

The cyano group in M116C-CN has the most central active site location of the probes, with the nitrile positioned less than 5 Å from the polar side-chain hydroxyl groups of Tyr 16, 32 and 57 and Asp 103, and less than 4 Å from the general base, Asp 40, and the expected position of the steroid oxyanion (Fig. 1C). It forms part of the solvent-accessible active site surface and is thus susceptible to hydrogen bond formation to adventitious water molecules within the active site.

The nitrile of F86C-CN refined to a position orthogonal to that of the original F86 phenyl ring, projecting away from the bound substrate and into a small recess contoured by the backbone groups of D103 and M84 (Fig. 1C). The 3-Å distance and triangular arrangement between the nitrile and amide groups of D103 and M84 suggested formation of a bifurcated hydrogen bond in which lone pairs on both the nitrile nitrogen and M84 carbonyl oxygen overlap with the amide proton of D103. IR and 13C NMR measurements described below support hydrogen bond formation to the probe. Positioned within this recess, the nitrile of F86C-CN is located 4 Å from a side-chain oxygen of D103, 5 Å from the oxyanion of the bound steroid, and 7–10 Å from the active site tyrosines.

These three probes represent distinct orientations relative to the molecular frame of the enzyme (Fig. 1C, Inset), span a range of distances to the polar groups of the active site, and experience varying degrees of solvent sequestration, hydrogen bond formation, and hydrophobic packing interactions. The unique and anisotropic chemical environment observed at each site therefore suggested that each probe samples a discrete electrostatic environment that, via Eq. 1, would lead to dispersion in the vibrational stretching frequency from site to site. In the sections that follow, we first present IR spectra for the three probes that demonstrate substantial -CN frequency dispersion. We then describe 13C NMR measurements for each nitrile that allowed us to rule out hydrogen bonding to the M105C-CN probe and identify and correct for hydrogen bond formation to the M116C-CN, and F86C-CN probes. Then we present in situ calibration of [Δν(CN)] for each nitrile to determine the inherent electric field sensitivity of each probe. With these measurements in hand, we are able to quantitatively determine the differences in electric fields sampled by the three probes in their unique positions and orientations within the active site and the response to changes in the local electrostatic environment.

Tandem Measurement of IR Stretch Frequency and 13C NMR Chemical Shift. FTIR spectra were acquired for each apo KSI-CN variant with either D40 or D40N present. These spectra, shown in Fig. 2A–C and summarized in Table 2, reveal asymmetric peaks that span a 12 cm−1 range in stretching frequency and exhibit a broad range in linewidth. Although linewidths are determined by the complex distribution of inhomogeneous environments and the time scale for interconversion between them and are often difficult to interpret, our structural studies and NMR data provide a guide for understanding the observed linewidth differences. The narrow linewidth of the M105C-CN peak relative to the other probes, or compared to the spectrum for urea-unfolded protein (33), is consistent with the tight packing of hydrophobic groups around the probe within a region of the active site previously suggested to have limited dynamic mobility.

Table 1. X-ray crystallographic data for nitrile-modified KSI variants

<table>
<thead>
<tr>
<th>D40N Variant</th>
<th>Ligand</th>
<th>Resolution (Å)</th>
<th>Rwork/Rfree (%)</th>
<th>PDB ID code</th>
</tr>
</thead>
<tbody>
<tr>
<td>M116C-CN</td>
<td></td>
<td>1.9</td>
<td>24.4/29.0</td>
<td>30XA</td>
</tr>
<tr>
<td>M105C-CN</td>
<td>equilenin</td>
<td>2.3</td>
<td>28.0/34.6</td>
<td>30YW</td>
</tr>
<tr>
<td>F86C-CN</td>
<td></td>
<td>2.0</td>
<td>26.5/34.1</td>
<td>30X9</td>
</tr>
<tr>
<td>F86C-CN</td>
<td>equilenin</td>
<td>1.7</td>
<td>21.6/27.0</td>
<td>30WU</td>
</tr>
</tbody>
</table>
Table 2. Vibrational frequencies and $^{13}$C NMR chemical shifts for nitrile-modified KSI variants at pH 7 in 40 mM potassium phosphate

<table>
<thead>
<tr>
<th></th>
<th>M116C-CN</th>
<th></th>
<th>F86C-CN</th>
<th></th>
<th>M105C-CN</th>
<th></th>
<th>L51C-CN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\nu_{CN}$ [cm$^{-1}$]</td>
<td>$\delta$ [ppm]</td>
<td>$\nu_{CN}$ [cm$^{-1}$]</td>
<td>$\delta$ [ppm]</td>
<td>$\nu_{CN}$ [cm$^{-1}$]</td>
<td>$\delta$ [ppm]</td>
<td>$\nu_{CN}$ [cm$^{-1}$]</td>
<td>$\delta$ [ppm]</td>
</tr>
<tr>
<td>D40N</td>
<td>2,159.6*</td>
<td>115.62</td>
<td>2,165.9</td>
<td>114.10</td>
<td>2,154.0</td>
<td>113.60</td>
<td>2,158.6</td>
<td>117.10</td>
</tr>
<tr>
<td>D40</td>
<td>2,161.3</td>
<td>114.43</td>
<td>2,165.3</td>
<td>113.98</td>
<td>2,152.3</td>
<td>NA</td>
<td>113.55</td>
<td>NA</td>
</tr>
<tr>
<td>3-F,4-NO$_2$-phe</td>
<td>2,163.3</td>
<td>113.92</td>
<td>2,166.4</td>
<td>112.78</td>
<td>2,153.3</td>
<td>NA</td>
<td>113.55</td>
<td>NA</td>
</tr>
<tr>
<td>4-NO$_2$-phe</td>
<td>2,163.8</td>
<td>113.94</td>
<td>2,166.6</td>
<td>112.78</td>
<td>2,153.3</td>
<td>NA</td>
<td>113.55</td>
<td>NA</td>
</tr>
<tr>
<td>equilenin</td>
<td>2,162.4*</td>
<td>114.07</td>
<td>2,167.7</td>
<td>112.78</td>
<td>2,153.9</td>
<td>113.61</td>
<td>2,153</td>
<td>113.61</td>
</tr>
</tbody>
</table>

*Reproduced from ref. 33.
†Reproduced from ref. 28.

(40). The relatively broad F86C-CN spectrum may reflect that its substitution for bulkier phenylalanine results in looser packing around this probe than for the other probes. The observation of the same ordering of linewidths for the $^{13}$C NMR peaks for each probe (Fig. S3) supports the hypothesis that broader linewidth reflects greater conformational inhomogeneity (additional discussion in SI Text).

Prior to any electrostatic analysis, it was necessary to first test and correct for effects from hydrogen bond formation to the nitrile probes. We recently published a method, based on tandem measurement of the nitrile IR frequency and $^{13}$C NMR chemical shift, that can detect the presence of specific probe interactions such as hydrogen bonds and quantitatively separate observed nitrile IR shifts into electrostatic and nonelectrostatic contributions (28). The common, linear electric field sensitivity of the IR and NMR observables leads to a correlation between them in the absence of specific interactions with the probe. Direct hydrogen bond formation to a nitrile leads to a departure from this linear correlation, ascribed to anomalous behavior in the IR but not the NMR dimension (41–43). By comparing tandem IR and NMR measurements for KSI-CN probes to a previously determined correlation of IR stretching frequency versus $^{13}$C NMR chemical shift of ethyl thiocyanate in aprotic solvents, such departures can be identified.

We previously applied this method to identify that the M116C-CN nitrile accepts a hydrogen bond and that the M105C-CN nitrile is free from hydrogen bonding (Table 2 and Fig. 3) (28). Hydrogen bond formation to the nitrile of M116C-CN but not M105C-CN is consistent with the solvent exposure of the former and hydrophobic burial of the latter observed in the X-ray structures (Fig. 1D). When we applied this method to F86C-CN, we observed a substantial offset in the IR frequency of F86C-CN from that expected based on its $^{13}$C NMR chemical shift and the trend line shown in green in Fig. 3, supporting the presence of the bifurcated hydrogen bond suggested from the X-ray structure.

Individual trend lines for the IR and $^{13}$C NMR data were determined by compiling multiple IR and $^{13}$C NMR measurements with mutants and bound ligands for each nitrile probe (Fig. 3). The slopes of all of these lines were the same, within error, strongly suggesting that in none of the examples of ligand binding or mutation studied by this method does the nitrile change from free to hydrogen-bonded or vice versa at any of the three sites. The larger 13 cm$^{-1}$ offset observed for F86C-CN from the IR/NMR correlation line, relative to the 10 cm$^{-1}$ offset observed for M116C-CN, may arise from the different hydrogen bond donors in the two cases: a backbone amide in the case of F86C-CN versus water for M116C-CN. Our prior conclusion (28), that IR shifts due to hydrogen bonding and local electrostatic effects are additive, is reinforced with the contribution here of additional data that strengthen the observed correlations. Based on this information, we can quantitatively correct for hydrogen bond formation to the F86C-CN and M116C-CN probes by subtracting their IR frequency offset from the IR/NMR correlation line (Fig. 3). Upon correction, the total dispersion in nitrile frequency between the three probes was reduced from 12 to 4.4 cm$^{-1}$ (Table 2) with D40N present and 1.6 cm$^{-1}$ with D40 present. This correction also reversed the relative peak positions of M105C-CN and F86C-CN, leaving the M105C-CN nitrile with the highest energy IR frequency.

In Situ Determination of the Stark Tuning Rate for Each Nitrile Probe. While both NMR and IR are sensitive to electrostatic fields, the VSE experiment provides a quantitative experimental calibration of the sensitivity of the vibrational frequency with respect to a known, external electric field. Such a direct calibration is not currently possible for $^{13}$C NMR chemical shifts. Vibrational Stark
Fig. 3. $^{13}$C-NMR Chemical shift versus nitrile stretching frequency for nitrile probes in apo or ligand-bound KSI-CN variants: F86C-CN (blue, $-1.6$ cm$^{-1}$/ppm, $R^2 = 0.9$), M116C-CN (black, $-2.2$ cm$^{-1}$/ppm, $R^2 = 0.89$), and M105C-CN (red), in solid circles (data from Table 2). These results are compared to data for the model compound ethylthiocyanate in different aprotic solvents, taken from ref. 28, in green diamonds ($-1.7$ cm$^{-1}$/ppm, $R^2 = 0.68$). Hydrogen bond corrections of $-13$ and $-10$ cm$^{-1}$, applied to the nitrile stretching frequency of F86C-CN and M116C-CN, respectively, as described in the text, give the values shown with blue and black open circles, respectively. For a combined dataset including ethylthiocyanate, M105C-CN, and the corrected F86C-CN and M116C-CN values, the best fit line gives a slope of $-1.8$ cm$^{-1}$/ppm, $R^2 = 0.83$. Data points are numbered for ethylthiocyanate in dimethylsulfoxide (1), dimethylformamide (2), acetone (3), methane chloride (4), tetrahydrofuran (5), chloroform (6), toluene (7), and cyclohexane (8). Data for KSI-CN variants are numbered 9-19: apo M116C-CN/D40N (9), apo M116C-CN (10), M116C-CN/D40N+equilin (11), M116C-CN/D40N+3-F-4-NO$_2$-phenol (12), M116C-CN/D40N+4-NO$_2$-phenol (13), apo F86C-CN (14), apo F86C-CN/D40N (15), F86C-CN/D40N+equilin (16), apo M105C-CN/D40N (17), apo M105C-CN (18), and M105C-CN/D40N+3-F-4-NO$_2$-phenol (19).

Tuning rates for individual probes are typically obtained using small molecules in frozen organic glasses (44–46). However, it remained untested whether the Stark tuning rate of a specific probe, such as a nitrile, varies substantially upon large perturbations to the local environment of the probe (47).

To address this question, we determined vibrational Stark spectra (field on minus field off) in situ for each KSI-CN variant embedded in a frozen glass at 80 K (Fig. 4). The Stark spectra were fit as a sum of derivatives of the absorption spectrum, and the contribution of the second derivative component was quantified to calculate $|\Delta \nu_{\text{CN}}|$ (see SI Materials and Methods).

$|\Delta \nu_{\text{CN}}| = f$ (where $f$ is the local field correction factor, which for simplicity we assume to be equal to one*) determined for each probe were within 20% of each other and of the model compound ethyl thiocyanate (EtSCN) measured in frozen organic solvent (49) (Fig. 4). Thus, the electric field sensitivity of the three nitrile probes, quantified by the Stark tuning rate, appears to be an intrinsic property of the oscillator that varies minimally across the observed range of measured IR frequencies and unique structural environments, regardless of whether the probe is hydrogen-bonded or not. According to Eq. 1 and using the average calibrated value for $|\Delta \nu_{\text{CN}}|$, a change in the projection of the field along the C-to-N bond vector of $+1$ MV/cm would shift the IR absorption, $\Delta \nu_{\text{obs}}$, by $+0.65$ cm$^{-1}$.

Quantitative Measurement of Electrostatic Field Heterogeneity Within the KSI Active Site. We first consider the data with the D40N mutant, because our results (X-ray, $^{13}$C-NMR, and Stark) are most extensive in this background. After correction for hydrogen bond-

Electrostatic Fields Within the KSI Active Site Are Altered by Restoring Asp at Position 40. Upon changing residue 40 from Asn back to Asp, IR shifts of different magnitude and direction were registered by each probe (Fig. 2 D–F). The peak widths in either background were similar, suggesting that mutations at residue 40 do not affect the dynamics or flexibility at any of the probe sites. Relatively small shifts, defined as $\Delta \nu_{\text{D40N}} = \nu_{\text{D40N}} - \nu_{\text{D40}}$, of $+0.6$ and $+0.8$ cm$^{-1}$ were observed for the nitriles of F86C-CN and M105C-CN, respectively, and a larger $-1.7$ cm$^{-1}$ shift was observed for the M116C-CN nitrile. These shifts by each probe, none of which involves a change in hydrogen bonding state (Fig. 3 and Table 2), indicate that each nitrile experiences the electrostatic effects of this mutation differently. On the basis of these shifts, we determined that the 7 MV/cm field dispersion sampled in D40N by probes at positions 86, 105, and 116 is reduced to 3 MV/cm with the negatively charged Asp a position 40.

**The local field correction factor describes the enhancement of the local field in the vicinity of the probe, with respect to the applied field in a Stark experiment, due to differences in the solvent and solute dielectric properties. In general, it is a tensor property but can be approximated with a scalar and has been calculated (48) to fall within a range of $1 < c < 1.3$, with the larger value for larger dielectric differences.
To understand the electrostatic effects of the D40N mutation, we first had to establish the protonation states of the ionizable residues in the active site before and after the mutation. Prior functional studies of catalysis by wild-type KSI from *P. putida* revealed a pKₐ of approximately 4 that was assigned to the general base, D40 (50). Upon mutation of D40 to Asn, phenolate and equilin binding studies suggested an enzymatic pKₐ of approximately 6 that was provisionally assigned to the oxyanion hole residue, D103 (Fig. 1B) (32). This assignment was based on work with the homologous KSI from *Comamonas testosteroni* that suggested that D103 has an elevated pKₐ of approximately 9 in WT, due to electrostatic repulsion from negative charge on D40 and burial in a hydrophobic pocket, but a lower pKₐ in D40N in the absence of charge repulsion from D40 (51). We tested these assignments using state-of-the-art algorithms to predict pKₐ in the absence of charge repulsion from D40 (51). We tested methods predicted that the D40N mutation, which ablates the cm⁻² of the homologous KSI from *P. putida* (32). This assignment was based on work on residue, D103 (Fig. 1), were predicted by computation to titrate with pKₐ values greater than 11 (Table S2).

**Electrostatic Fields Within the KSI Active Site Are Altered by pH Titration.** The functional results described above provide evidence that the D40N mutation is accompanied by ionization of a nearby residue whose pKₐ has been lowered to a value near 7 due to this mutation. To study the effects of this ionization alone and thus to understand its contribution to the net electrostatic effect of the D40N mutation, we compared the changes sensed by the probes in response to raising the buffer pH from 5.6 to 7.8, values at which we expected this nearby residue to be neutral or ionized, respectively. Shifts, defined as Δν = ν - ν₀, of +1, 0, and -2 cm⁻¹ were observed for F86CN-CN, M105C-CN, and M116C-CN, respectively (Fig. 2 J–L). These values recapitulate the approximate size and direction of the IR shifts observed above (Fig. 2 D–F) due to mutation of residue 40 from Asp to Asn. This suggests that the ionization of this nearby residue, affected either by pH or as an adventitious consequence of the D40N mutation, makes a larger contribution to the observed IR shifts of our probes than the ablation of charge at position 40 due to the D40N mutation itself.

**Electrostatic Modeling Suggests Ionization of an Active Site Tyr in D40N.** Literature precedent (32), pKₐ prediction calculations, and simple chemical reasoning all suggested that the D40N mutation resulted in charge transfer from D40 to D103, with a resulting pKₐ for D103 near 7 in D40N. However, the following analysis motivated us to question this assumption. We compared the observed frequency shifts to those predicted from continuum electrostatic modeling with DelPhi (55, 56), to test whether these predicted charge configurations were consistent with the electrostatic field changes reported by the IR probes above (Fig. 5). As summarized in Fig. 5, the predicted shifts assuming that D103 was the ionizing residue in D40N disagreed with the experimental results for each probe position. Even though we would not expect quantitative agreement between calculation and experiment, we were surprised by the enormity of the discrepancies at each probe position.

The disagreement between measured field changes and the predicted effects due to D103 ionization can be readily appreciated using Coulomb’s law, Eq. 1, and the distances and angles between each probe and D103 shown in Fig. 2 G and H (for a list of angles, distances and Coulomb’s law calculations, see Table S3). Based on the 5.7- and 3.5-Å proximity of D103 (circled in red) to M105C-CN and F86C-CN, respectively, and the 107° and 68° angles formed by the nitrile bond vector and the vector between the probe and the carboxylate group, the ionization of D103 would be expected to result in large ~15 and ~22 MV/cm fields at these probe positions. In contrast to this expectation, only small shifts are observed for these probes (Fig. 2 J and K). The failure of electrostatic modeling based on ionization of D103 to account for our observed IR shifts motivated us to consider whether a residue other than D103 might be ionizing in the D40N mutant.

The only other titratable residues within the active site are Y16, Y32, and Y57, which form an extended hydrogen bond network in the oxyanion hole (Fig. 1B). Although computational pKₐ predictions for these Tyr residues all gave values greater than 11 (see above), we nonetheless used Delphi to predict IR shifts based on ionization of each of these residues, as done above for D103, and compared these predictions to the observed shifts for mutation of D40 to Asn (Fig. 5). Surprisingly, ionization at any of these residues did a better job of accounting for the observed frequency shifts than ionization of D103, with residue Y57 giving the closest congruence. We therefore turned to experiments that could directly assess whether an ionized tyrosine was present in the enzyme at neutral pH.

**UV and Tyr 13C NMR Spectra Identify an Ionized Tyr in the D40N Active Site.** UV absorbance provides a sensitive probe of possible tyrosine ionization within KSI, as the absorbance peak centered at 280 nm for neutral tyrosine shifts to 295 nm upon ionization (57). As shown in Fig. 6 there is an increase in absorbance at 295 nm in the UV spectrum of the D40N mutant that titrated with an apparent pKₐ of 6.3 ± 0.1 as the solution pH was raised from 4.7 to 10.5, suggesting ionization of a Tyr residue within unliganded KSI D40N with this pKₐ.†

†Similar results are obtained for the probe modified KSI-CN variants (SI Text and Fig. S5).

**Fig. 5.** Comparison of experimental electrostatic field changes (in MV/cm) observed at each probe site and predicted field changes calculated for each probe site based on ionization of different active site residues using DelPhi and the experimental Stark tuning rate. Comparison of experimental versus calculated field changes are made for D40 versus D40N (A–B) and for raising the pH from 6 to 8 (C–D). A and C show the sign and magnitude of the observed (black) and predicted shifts side by side, with predictions for the scenarios in which D103, Y16, Y32, or Y57 (red, green, blue, and magenta, respectively) lose a proton due to mutation or raising the pH. B and D show the magnitude of the error in the predicted shift depending on the choice for the group undergoing ionization at each of the probe sites, 105, 86, and 116 (orange, light blue, and gray, respectively).
To directly test for the presence of an ionized Tyr in pKSI at neutral pH, we turned to $^{13}$C NMR studies of KSI containing Tyr residues labeled with $^{13}$C at position C$_z$, the carbon adjacent to the hydroxyl group, as the chemical shift of this carbon is exquisitely sensitive to the ionization state of Tyr. For tyrosine-containing peptides in neutral water, C$_z$ has a $^{13}$C chemical shift of 155.5 ppm for a protonated Tyr that shifts downfield nearly 11 ppm to 166.3 ppm upon ionization (58). Wild-type and the D40N mutant of pKSI were expressed in cells grown in a minimal medium containing $^{13}$C-labeled tyrosine, and $^{13}$C-NMR spectra were obtained. KSI from P. putida contains four tyrosines: Y16, Y32, and Y57 form an active site hydrogen bond network with each other, and Y119 is located on the protein surface near the dimer interface and far from the active site. For WT KSI, D40 is expected to be the only ionized active site residue. Consistent with this expectation, we observed a cluster of peaks in the $^{13}$C NMR spectrum of WT KSI with chemical shifts around 158 ppm, very similar to the expected position for a protonated Tyr (Fig. 7A). The peak at 157.5 ppm is approximately twice as intense as the flanking peaks at 157.0 and 158.5 ppm, suggesting that two of the Tyr residues in WT KSI have identical chemical shifts.

In contrast, the $^{13}$C NMR spectrum of D40N at pH 7 displays four well-resolved peaks (Fig. 7B). The most upfield peak at 157.5 ppm is very similar to the expected chemical shift for a neutral tyrosine in water. This peak can be assigned to the surface residue Y119 based on its selective disappearance in the $^{13}$C spectrum of the Y119F/D40N mutant (Fig. S4). The three remaining peaks are observed at 159.7 ppm, 161.6 ppm, and 165.4 ppm, all downfield of the chemical shift expected for neutral Tyr. The peak at 165.4 ppm is deshielded by nearly 8 ppm relative to the peak at 157.5 ppm for neutral Y119 and is within 1 ppm of the chemical shift expected for a fully ionized Tyr in aqueous solution. On the basis of our UV and $^{13}$C NMR results, we conclude that an active site Tyr is indeed ionized in unliganded D40N at neutral pH, consistent with our prediction above based on quantitative modeling of the IR peak shifts for our nitrile probes.

Electrostatic Fields Within the KSI Active Site Are Altered by Binding a Charged Ligand. As a final test of how fields in the KSI active site are affected by a local charge perturbation, we measured the electric field changes associated with binding of a charged ligand. The ligands 3-fluoro-4-nitrophenol (3F,4-NO$_2$-phe), and 4-nitrophenol (4NO$_2$-phe) were previously shown to bind KSI D40N as anions and have been studied as transition state analogs (32, 59). As shown in Fig. 7C, binding of 4NO$_2$-phe to D40N shifts the observed $^{13}$C resonances for KSI Tyr residues into the chemical shift range expected for neutral tyrosine. Thus, binding of 4NO$_2$-phe to D40N results in migration of negative charge from an active site Tyr to the phenolate oxygen. Using the IR frequency difference upon ligand binding, defined as $\Delta \nu_{\text{CN}} = \nu_{\text{CN}}^{\text{apo}} - \nu_{\text{CN}}^{\text{bound}}$, the average Stark tuning rate of 0.65 cm$^{-1}$/MV/cm, and Eq. 1, we determined that binding of 4-nitrophenol results in field shifts of +6.5, −1.1, and +1.1 MV/cm for probes at position 116, 105, and 86, respectively. Similar values are observed for 3F,4-NO$_2$-phe (Table 2). The large +6.5 MV/cm shift observed for M116C-CN upon 4-NO$_2$-phe binding is consistent with the close proximity (<4 Å) of this probe to the phenolate oxygen and is nearly as large as the 8 MV/cm site-to-site field variation observed in apo D40N.

Discussion and Conclusions

Electrostatic Heterogeneity Within an Enzyme Active Site. X-ray crystallography has contributed enormously to our understanding of protein structure. Nevertheless, translating the pictures of the chemically heterogeneous milieu of charged, polar, and hydrophobic groups yielded by X-ray crystallography into a quantitative understanding of the highly heterogeneous and anisotropic electrostatic environment has remained one of the most challenging and elusive goals of protein biophysics. Using thioceyanate vibrational probes incorporated at multiple positions and orientations within the active site of bacterial ketosteroid isomerase, we have exploited the directional nature of the nitrile stretching vibration and its exquisite electric field sensitivity to determine that electric fields within the KSI active site vary up to 8 MV/cm, and future studies of additional probe positions may give a larger range.

To put this magnitude of site-to-site field heterogeneity in context, consider that a positive and a negative elementary unit of charge separated from each other by 1 Å represents a dipole of 5 Debye. The polar covalent bonds present within amino acids or the KSI steroid substrate typically have bond dipoles in the range of 1–4 Debye. The difference in electrostatic stabilization of a 5 Debye dipole situated with the position and orientation of the nitrile of L61C-CN versus that of M105C-CN can be calculated.
from $\Delta E = -e \mu_{lipid} \cdot \Delta F$ and equals 2 kcal/mol. Thus, measured differences in the local electrostatic fields within the KSI active site due to differences in location or orientation are of sufficient magnitude to result in significant energetic effects on individual bond dipoles, if properly aligned and if the surrounding dipoles are sufficiently restricted from rearrangement. Similar arguments apply to differential stabilization of dipolar transition states.

Prior reports of electrostatic variation sampled within proteins generally explored either changes in response to charge perturbation, or site-to-site differences, whereas this study encompassed both features. In the first case, the stretching vibration of CO bound to myoglobin (17, 60) and a nitrile-containing inhibitor bound at the active site of human aldose reductase (46) exhibited a range of frequencies in response to mutations in the vicinity of the probe. The vibrational Stark model presented for these band shifts translated into electrostatic changes of 15 and 13 MV/cm, respectively. However, the largest shifts coincided with cases that likely involved the making or breaking of hydrogen bonds to the probe, as hypothesized on structural grounds (46, 61), which would convolute the observed effects (28). In the second case, the chemical shift dispersion for $^{19}$F-labeled amino acids incorporated at multiple locations within galactose binding protein were calibrated using a computational model to suggest that site-to-site variations in the projection of the electrostatic field on the C–F bond vector as large as 40 MV/cm occur (20). However, conflicting computational arguments have been made that short-range chemical interactions (i.e., van der Waals interactions) between fluorine atoms and surrounding groups can dominate the chemical shift dispersion rather than local electrostatics (19, 62, 63).

While it may not be possible to completely isolate long-range electrostatic effects from these examples of specific local interactions, our work suggests that, in select cases, such effects can be controlled and corrected for by determining Stark tuning rates (Fig. 4), vibrational frequencies, and $^{13}$C NMR chemical shifts (Fig. 3 and Table 2) for the nitrile probes. In particular, the relationship between the nitrile’s NMR chemical shift and vibrational frequency allows for a separate accounting for hydrogen bond contributions, in contrast to electric field observables such as those provided by modeling pKa shifts and for which no such straightforward corrections are currently possible. Furthermore, use of a spectrometer that experiences field changes remotely, rather than participating in a chemical change as part of the observation, greatly simplifies the interpretation. Thus, our measurements of both charge perturbations and site-to-site variations, based entirely on experimental data, represent the cleanest isolation of electrostatic field effects in proteins obtained to date. Nevertheless, future experimental tests and computational comparisons will be needed to further define the capabilities and limitations of this approach.1

**Local Versus Global Determinants of the Electrostatic Environment.** The large size of enzymes relative to simple chemical catalysts has prompted many questions regarding the role of the global protein structure in shaping the catalytic properties of active sites. Unlike the significant, collective contribution of distal or nonactive site residues to the positioning of binding and catalytic groups within active sites, a role for these distal residues in shaping electrostatic properties of active sites has remained uncertain. Electrostatic effects are additive and therefore always a collective property of all residues. However, the relative contributions of discrete charges and dipoles to the collective sum is unknown.

**Limited Electrostatic Rearrangement Within Active Sites.** Two related observations in this report support the long-held view that structural organization of proteins during folding restricts electrostatic rearrangement and place quantitative limits on the extent of any such rearrangements. Our observation that charge perturbations separated by several angstroms from the probes can lead to substantial electrostatic changes (Fig. 2) provides evidence for limited electrostatic attenuation due to restricted motion of protein groups and limited water penetration into the protein interior. The observation of qualitative agreement between measured field changes and those calculated with continuum electrostatics, assuming a value of 2 for the internal dielectric, also argues for low electrostatic screening. However, we stress that computational models based on a uniform “protein dielectric” are an inadequate description of the heterogeneous protein environment, as noted by others (6, 11).

While the charge transfers explored in this manuscript represent changes that have equilibrated with the surrounding protein matrix, understanding nonequilibrium effects due to transient charge perturbations remains an important challenge. In a separate study, we have carried out time-resolved measurements of the changes in vibrational frequency of these same nitrogen-modified KSI constructs following photoexcitation of a bound chromophore. These experiments provide a measure of the electrostatic changes within the protein matrix in response to the near-instantaneous creation of an excited state dipole in the KSI active site (65). The results of this study, especially the observation that the magnitude of the initial electric field-induced shift and the orientation of the nitrile remain constant over the pico-second time-scale of the experiment suggest an effectively rigid and structurally electrostatic environment on this time scale. This observation complements the finding in the current study of limited steady-state electrostatic rearrangement in response to charge perturbation.

**Observation of an Ionized Tyr in the KSI D40N Active Site.** An apparent pK$_a$ of 6.3, tentatively assigned to Y57 based on electrostatic modeling, represents a 3.7 unit perturbation from the solution pK$_a$ value of 10 for the tyrosine side chain in aqueous peptides (58) and occurs in the absence of a nearby stabilizing counter-ion. In contrast, ionized tyrosine residues in calmodulin, dTDP-glucose-4,6-dehydratase, UDP-galactose-4-epimerase and alanine racemase, the only prior literature reports to our knowledge for tyrosine residues with pK$_a$ values near 6, appear to be stabilized by proximal positively charged side-chains or metal atoms (66–69). The role played by the unique and extended hydrogen bond network, that includes Y16, Y57, Y32 and multiple water molecules, in shaping the unusual acidity of the tyrosine residue in D40N is the subject of an upcoming study.

**Computational Prediction Versus Experimental Measurement of Protein Properties.** We have leveraged the strengths of computational modeling to extend and deepen our understanding of electrostatic fields and features within KSI, while also highlighting critical limitations in current computational approaches, as have been raised on theoretical grounds (7, 8). The algorithms we tested universally erred in predicting the relative pK$_a$ values for

---

1While it is theoretically possible to also extract the absolute value of the electric field projection with this model by comparing the measured frequency with a gas-phase (i.e., no external field) frequency measurement, there are multiple features of the gas-phase to condensed-phase frequency shift that are not captured by electrostatics alone. For an introduction to this topic see ref. 64.
D103 and the experimentally identified titratable active site tyrosine in the D40N mutant of *P. putida* KSI. This failure is not a subtle effect due to ionization of the first of two residues of nearly equal pKₐ creating a charge-repulsion interaction that raises the pKₐ of the second residue; even for the D40N/D103N double mutant, the active site tyrosines are predicted to have pKₐ values above 11 (additional discussion in SI Text). In light of the complex hydrogen bond network that appears to underpin the unusually low tyrosine pKₐ observed, we speculate that simple continuum electrostatic models for the protein and solvent are inadequate to have predicted this observation.

Without experimental knowledge of the ionization states for all residues in the active site in all scenarios of mutation, ligand binding and pH (a problem common to all but the most exhaustively characterized proteins), we found irreconcilable discrepancies between the electrostatic field differences measured by our nitrile probes and continuum electrostatic calculations or even simple arguments based on Coulomb's law. However, computational models for the change in electrostatic field felt by our nitrile probes gave qualitative agreement with the experimental results provided that an ionization at neutral pH was assumed for an active site tyrosine residue instead of D103, with the models giving the best agreement for ionization of Y57. These computational results inspired UV and 13C NMR experiments in this work that tested and confirmed ionization of a Tyr in apo D40N with a pKₐ of 6.3. Combined molecular mechanics sampling and that tested and confirmed ionization of a Tyr in apo D40N with a pKₐ of 6.3. Combined molecular mechanics sampling and continuum electrostatic models for the protein and solvent are inadequate to have predicted this observation.

This synergism between directional IR field measurements, computational modeling, and spectroscopic probes of side-chain ionization states highlights the productive interplay between experiment and theory and the power of utilizing directional probes at multiple positions to triangulate local electrostatic effects. The directional sensitivity of our electrostatic probes was indispensable in modeling and distinguishing the effects of ionization at Y57 versus other positions, and contrasts with common electrostatic characterizations of protein interiors in terms of nondirectional or uniform “effective dielectric” descriptions. Finally, our study directly measuring field differences from multiple positions and orientations provides important benchmarks for improving computation modeling of proteins and can serve as a guide for designing future incisive tests of the functional roles and contributions of protein electrostatics.

**Materials and Methods.** See SI Materials and Methods for complete experimental methods including the procedure for thiocyanate and 13C-tyrosine labeling of KSI, the conditions of crystal growth for KSI-CN variants, the details of X-ray acquisition and structure solving, the parameters for continuum electrostatic calculations with the program DelPhi and the description of instrumentation and sample preparation for UV-vis, IR, Stark-effect and NMR spectroscopies.

**ACKNOWLEDGMENTS.** We thank Corey Liu and Steve Lynch for assistance with NMR experiments and Axel Brunger for helpful discussions. P.A.S. was supported in part by NIH and G. Lieberman predoctoral fellowships. Funding was provided by grants to S.G.B. (National Institutes of Health Grant GM27738) and D.H. (National Science Foundation Grant MCB-1121778). Portions of this research were conducted at the Advanced Light Source, a national user facility operated by the Lawrence Berkeley National Laboratory, and at the Stanford Magnetic Resonance Laboratory, which is supported in part by the Stanford University Medical School.


NMR Spectroscopy of KSI Variants Containing 13C-Tyrosine or a 13C-Nitride. One-dimensional 13C NMR spectra were acquired on Varian UNITY INOVA NMR spectrometers (500 or 600 MHz, 1 H frequency) running VNMR v6.1C and equipped with a 5 mm PFG switchable probe or a 10 mm broadband probe operating at ambient temperature (20 °C). Samples contained 1 mM KSI, 2 mM ligand (bound samples only), 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% D2O as the lock solvent. Spectra were acquired in a 5 mm or 10 mm Shigemi symmetric microtube for approximately 5,000 scans and processed using a 10 Hz line broadening. Peak positions were determined using a second-derivative-based algorithm (3, 4). Site-directed mutagenesis of KSI was performed as previously described (5), and sequences of all constructs were confirmed by sequencing miniprep plasmid DNA. Protein masses were confirmed by KSI LC/MS and were within 5 Da of the expected values in all cases.

FSK X-ray Crystallography. Crystals of KSI-CN variants in space group P21 were obtained using hanging drop vapor diffusion by mixing 2 μL of 30 mg/mL KSI (also containing equimolar equilenin in the cases of the F86C-CN/D40N•equilenin and M105C-CN/D40N•equilenin structures) with 2 μL of reservoir solution (1.1 M ammonium sulfate, 40 mM potassium phosphate (pH 7.2), 1 mM EDTA, 2 mM DTT, and 5% D2O) as the lock solvent. Spectra were acquired in a 5 mm or 10 mm Shigemi symmetrical microtube for approximately 5,000 scans and processed using a 10 Hz line broadening. Peak positions are reported as the chemical shift value corresponding to the position of maximum intensity of each peak. 13C chemical shifts were referenced to an external standard of sodium 3-trimethylsilylpropionate-2,2,3,3-d4 (0 ppm) under identical solvent and temperature conditions.

UV-vis Absorption Measurements. UV-vis spectra were obtained for the nitrile-modified enzymes as a function of pH using a Perkin Elmer Lambda 4 spectrophotometer with a 1-cm path length quartz cell. Difference spectra were acquired by subtraction of the absorption spectrum at pH 4.7 for wild type and D40N (Fig. 6), and at pH 5.3 for KSI-CN variants (Fig. S5). The higher pH value for the KSI-CN variants was necessary due to reduced solubility at lower pH. The spectra were baseline-corrected to a value of zero at 320 nm, and the lowest pH spectrum was subtracted from subsequent spectra to calculate the difference spec-
tra. For D40N and WT enzymes, sodium acetate was used for pH 4.7, potassium phosphate was used in the range of 4.7 to 9.1 and glycine was used in the range of 9.0 to 10, with a buffer concentration of 10 mM in all cases. For the KSI-CN variants, 20 mM MES buffer was used and pH-adjusted by adding aliquots of 0.1 M HCl while monitoring the pH with a microelectrode in situ. The enzyme concentration was 10–20 μM.

Poisson–Boltzmann Modeling. DelPhi, (14, 15) a finite difference algorithm for solving the Poisson–Boltzmann equation, was employed for each different charge-perturbation modeled above. Charge and radii parameters were taken from the PARSE parameter set (14), which treats only heavy atoms and hydrogens attached to heteroatoms explicitly. Charge and radii parameters for the nonnatural Cys-CN amino acid were taken from a previous study (12). A dielectric of value, ε, of 80 was used for the protein exterior. The value for the ε assigned to the protein interior presents an important choice: ε = 2 is justified based on the intrinsic polarizability of the chemical groups the protein is composed of (16), whereas it has been recognized that higher dielectric values can be employed in continuum electrostatic calculations to compensate for unaccounted for charge and dipole rearrangements (17,18). Through X-ray crystallography and NMR spectroscopy we have attempted to minimize the structural and ionization state changes, respectively, that are not explicitly accounted for, therefore a lower value of ε is called for. Calculations were performed for ε = 2 and ε = 4, with ε = 2 providing a lower rmsd from the experimental dataset; all calculated values reported were therefore taken from the ε = 2 calculation. Heavy atoms positions were determined from the X-ray structures, with hydrogen atoms automatically added to heteroatoms using the program PDB2GMX, (19) based on database values for bond distances and angles, H-bond-satisfaction criteria and unperturbed pK∞ values. To avoid artifacts from the finite grid treatment, a seven tier focusing routine was used. A calculation of the potential on a 50 × 50 × 5 grid of 0.5 grid lines per Ångstrom, centered around the CN-bond and encompassing the whole protein provided the initial the boundary conditions for a subsequent calculation at two times finer grid spacing. This process was repeated until the grid spacing was 32 lines per Ångstrom. The gradient in the potential along the CN-bond axis was calculated at each tier, and monitored for convergence; all values reported are from the final tier.

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

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

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

II. Discussion of IR and NMR transition linewidths. As discussed in the main text, the linewidths of the nitrile IR stretching transitions for the apo-D40N KSI variants follow the order M165C-CN < M116C-CN < F86C-CN, both at room temperature (Fig. 2 D–F) and at 80 K (Fig. 4). The overall linewidth of a spectroscopic transition results from the complex interplay between inhomogeneous broadening (due to distinct subpopulations with unique transition frequencies within the ensemble), motional narrowing (due to rapid interconversion between these subpopulations), and system-dependent excited-state lifetime-limited broadening. The observation of identical hierarchies at temperatures low enough to freeze out many motional narrowing mechanisms suggests that the observed difference in linewidths at the three probe-sites is primarily due to inhomogeneous broadening. Additional evidence that linewidth differences are due to differential contributions from inhomogeneous broadening is provided by the observation of an identical hierarchy of linewidths in the 13C NMR spectra of the nitrile probes (Fig. S3). The time scale defined by the full-width-at-half-maxima (FWHM) for the NMR peaks, approximately 100 Hz, is nine orders of magnitude slower than the time scale defined by the IR linewidths of approximately 10 cm⁻¹ or 300 GHz. Inhomogeneity that persists on the long time scale defined by NMR will necessarily be detected on the much faster timescale of IR. We propose that the observed differences in IR and NMR linewidth for the individual nitriles arise from differences in conformational heterogeneity for each nitrile arising from differences in local packing interactions, as discussed in the main text.

III. Discussion of Monitoring Tyr ionization by NMR and UV-vis. We attempted to use NMR to ascertain the pK∞ of the most downfield tyrosine resonance in Fig. 7. Increasing the pH to 8 resulted in no detectable change in the NMR spectrum. Lowering the pH to 6.2 resulted in peak broadening and a diminished signal-to-noise (Fig. S6). The broadening increased until the signal was obscured but before new resonances from the protonated tyrosine network were observed. This observation is common for widely separated chemical shifts in dynamic exchange whose rate of exchange depends on pH (21). This effect limits the utility of NMR for determining a pK∞ for this ionization. Furthermore, the broadened resonances do not appear to disappear due to changes in relaxation time with pH. At pH 6.4, where the intensity of the peak at 165.5 ppm is diminished relative to pH 7.1, increasing the recycle delay 4x (more than doubling the spacing between pulses) had no effect on the peak height. Fitting the downfield peak with
a Lorentzian lineshape allows determination of the linewidth at each pH. Loss in peak height from pH 7.1 to pH 6.5 is exactly compensated for by increase in peak width, so that net area is unchanged.

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

IV Discussion of pKa Prediction Algorithms.

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

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

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

Fig. S1. Origins of the linear Stark effect. (A) An anharmonic potential energy function describing the CN stretching mode. As a consequence of the anharmonicity, different average interatomic separation distances in the ground and vibrationally excited states, \( x_{gs} \) and \( x_{es} \), respectively, lead to an increase in the dipole moment, \( \Delta \mu_{\text{anh}} \), in the excited state. The increase in the dipole moment leads to a different relative effect of an electric field on the two states, and thus a field-dependent change in the energy of excitation. The total influence of the field on the CN stretching transition also includes an effect on the force constant of the bond, \( \Delta \mu_{\text{bond}} \). For details see ref. 20. (B) Spectra for three variants of a hypothetical protein with a fixed nitrile probe. The field projection \( F \) in the three separate constructs—wild-type, mutant 1, and mutant 2—is different in magnitude and/or angle, \( \theta \) as a consequence of the mutation, leading to different projections \( \Delta F \), on the fixed position of \( \Delta \mu \). This leads to changes in the transition frequency, \( \Delta \nu \), relative to WT.
Fig. S2. 2F₂-F₁ Electron density maps KSI-CN variants. The electron density, contoured at 1.5 sigma, is displayed as a wire mesh superimposed upon the structural model shown in stick format. Carbon, oxygen, nitrogen, and sulfur atoms are shown in green, blue, red, and gold, respectively.
Fig. S3. NMR spectra of $^{13}$C-nitrile labeled pKSI D40N variants. Spectral parameters are also shown for M116C-CN/D40N unfolded in 6 M urea.

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

Fig. S5. UV-Vis difference-absorption spectra of probe-modified D40N KSI variants. (A–C) Difference spectra for pH values 5.3 to 9 (red to black) are calculated relative to the spectrum at pH 5.3, for M105C-CN, F86C-CN and M116C-CN, respectively. (D–F) The change in absorbance at the absorption maximum of tyrosinate, 293 nm, as a function of measured pH, for M105C-CN, F86C-CN and M116C-CN, respectively.
Fig. S6. ¹³C NMR spectra of KSI D40N with labeled tyrosines. (A) The effect of decreasing the pH. All spectra were taken on the same sample (buffer = 40 mM phosphate, no other additives; pH decreased with dropwise addition of 1 M KH₂PO₄). (B) Expanded view of downfield region of A, height-normalized to 1.

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>D40N/M116C-CN</th>
<th>D40N/F86C-CN</th>
<th>D40N/F86C-CN</th>
<th>D40N/M105C-CN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand</td>
<td>–</td>
<td>Equilenin</td>
<td>Equilenin</td>
<td></td>
</tr>
<tr>
<td>PDB ID code</td>
<td>3OXA</td>
<td>3OX9</td>
<td>3OWU</td>
<td>3OWY</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>41.67–1.89</td>
<td>41.13–2.00</td>
<td>36.14–1.70</td>
<td>36.36–2.30</td>
</tr>
<tr>
<td>Space group</td>
<td>P₂₁</td>
<td>P₂₁</td>
<td>P₂₁</td>
<td>P₂₁</td>
</tr>
<tr>
<td>a, Å</td>
<td>42.9</td>
<td>44.1</td>
<td>45.2</td>
<td>45.3</td>
</tr>
<tr>
<td>b, Å</td>
<td>93.6</td>
<td>129.8</td>
<td>130.3</td>
<td>129.4</td>
</tr>
<tr>
<td>c, Å</td>
<td>75.4</td>
<td>75.4</td>
<td>45.1</td>
<td>84.0</td>
</tr>
<tr>
<td>α, °</td>
<td>103.9</td>
<td>114.5</td>
<td>112.7</td>
<td>96.5</td>
</tr>
<tr>
<td>β, °</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
</tr>
<tr>
<td>γ, °</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
<td>90.0</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>39,383</td>
<td>23,898</td>
<td>32,425</td>
<td>39,308</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>85.0</td>
<td>76.4</td>
<td>61.6</td>
<td>92.4</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.4</td>
<td>3.8</td>
<td>4.4</td>
<td>4.8</td>
</tr>
<tr>
<td>R_{merge}, %</td>
<td>9.4</td>
<td>18.9</td>
<td>11.7</td>
<td>25.5</td>
</tr>
<tr>
<td>I/σ_{average} (I/σ_{highres})</td>
<td>14.6 (1.5)</td>
<td>6.7 (1.1)</td>
<td>11.0 (1.0)</td>
<td>6.0 (1.2)</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<table>
<thead>
<tr>
<th>No. residues</th>
<th>130</th>
<th>129</th>
<th>129</th>
<th>129</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. waters</td>
<td>44</td>
<td>77</td>
<td>218</td>
<td>132</td>
</tr>
<tr>
<td>R_{work}, %</td>
<td>24.4</td>
<td>26.5</td>
<td>21.6</td>
<td>28.0</td>
</tr>
<tr>
<td>R_{free}, %</td>
<td>29.0</td>
<td>29.0</td>
<td>27.0</td>
<td>34.6</td>
</tr>
<tr>
<td>rmsd bond, Å</td>
<td>0.005</td>
<td>0.009</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>rmsd angle, °</td>
<td>0.62</td>
<td>1.08</td>
<td>1.06</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*R_{merge} = \sum_{hkl} \sum_{i} |I(hkl)| - |\langle I(hkl) \rangle| / \sum_{hkl} \sum_{i} I(hkl),

*R_{work} = \sum_{hkl} \sum_{i} F(hkl) - |F(hkl)| / \sum_{hkl} \sum_{i} F(hkl),

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

### Table S2. pK_a predictions from web-based continuum electrostatic calculation engines

<table>
<thead>
<tr>
<th></th>
<th>M116A/D40N</th>
<th>M116A/D40N/D103N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asp103</td>
<td>Tyr16</td>
</tr>
<tr>
<td>H++</td>
<td>6.0</td>
<td>27</td>
</tr>
<tr>
<td>PROPKA</td>
<td>6.4</td>
<td>20</td>
</tr>
<tr>
<td>Karlsberg</td>
<td>7.8</td>
<td>&gt;20</td>
</tr>
</tbody>
</table>

Fafarman et al. www.pnas.org/cgi/doi/10.1073/pnas.1111566109
Table S3. Distances and angles between midpoint of nitrile bond and side chains that could undergo ionization changes

<table>
<thead>
<tr>
<th>Atom position</th>
<th>Dist [Å]</th>
<th>Angle [°]</th>
<th>$\cos \theta$</th>
<th>Projected field [MV/cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP 103OD1 §</td>
<td>M116C-CN</td>
<td>4.1</td>
<td>110.7</td>
<td>0.353</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>3.5</td>
<td>67.5</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>5.7</td>
<td>106.8</td>
<td>0.289</td>
</tr>
<tr>
<td>ASN 40OD1 §</td>
<td>M116C-CN</td>
<td>5.7</td>
<td>27.7</td>
<td>0.885</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>9.2</td>
<td>90.7</td>
<td>−0.012</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>11.1</td>
<td>117.5</td>
<td>0.462</td>
</tr>
<tr>
<td>EquileninO1 §</td>
<td>M116C-CN</td>
<td>4.1</td>
<td>80.1</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>5.2</td>
<td>93.1</td>
<td>−0.054</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>7.1</td>
<td>96.6</td>
<td>−0.115</td>
</tr>
<tr>
<td>Tyr 16OH §</td>
<td>M116C-CN</td>
<td>4.5</td>
<td>89.1</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>7.3</td>
<td>86.8</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>6.5</td>
<td>84.9</td>
<td>0.089</td>
</tr>
<tr>
<td>Tyr 57OH §</td>
<td>M116C-CN</td>
<td>4.1</td>
<td>80.7</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>9.5</td>
<td>83.1</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>7.6</td>
<td>91.4</td>
<td>−0.024</td>
</tr>
<tr>
<td>Tyr 32OH §</td>
<td>M116C-CN</td>
<td>4.3</td>
<td>110.2</td>
<td>−0.345</td>
</tr>
<tr>
<td></td>
<td>F86C-CN</td>
<td>10.2</td>
<td>69.2</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td>M105C-CN</td>
<td>6.3</td>
<td>100.7</td>
<td>−0.186</td>
</tr>
</tbody>
</table>

*The distance is calculated between the location of the closest atom of the side-chain to the and the midpoint of the nitrile bond.
†The angle between the CN-bond axis and the location of the ionizing atom, X, measured from the midpoint of the nitrile bond (N-midpoint-X).
‡Fields according to Coulomb’s law: $\frac{q^2}{4\pi\varepsilon\varepsilon_0 r^2}$, with $\varepsilon = 2$, projected on the nitrile bond axis.
§The residue and atom names refer to the names given in the PDB structures.