Supporting Information

Hansen and Kay 10.1073/pnas.1400577111

SI Materials and Methods

Sample Preparation

The products 1-methyl histidine and 3-methyl histidine were purchased from Sigma Aldrich and suspended in 50 mM potassium phosphate buffer, pH 8.2. Isotopically enriched protein samples [wild-type Im7 and phospholipase C-γ_1 (PLCCγ) SH2 domain] were produced with 3 g/L U-^{13}C glucose and 1 g/L $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively, in protonated M9 minimal media. Wild-type Im7 was expressed and purified as described previously (1) with 6–7 mg protein fractions lyophilized in water for storage and subsequently resuspended in Im7 NMR buffer (50 mM potassium phosphate, pH 5.5–8.0, 0.02% azide, 95% H_2O/5% D_2O) before recording the NMR experiments. The PLCCγ SH2 domain was produced by protein overexpression in *BL21 Escherichia coli* cells, 37 °C, to an OD of 0.8–1.0 before induction with 1 mM isopropyl β-D-1-thiogalactopyranoside. Protein expression was allowed to continue overnight at 16 °C with a final OD of 1.5 immediately before harvesting. Cells were pelleted at 6,000 × *g* for 20 min and resuspended in 25 mL lysis buffer (50 mM Mes pH 6.0, 50 mM NaCl, 1 mM TCEP, 1 mM EDTA, 1 mM benzamidine, DNase I, lysozyme) before storing at −20 °C overnight. Subsequently the cells were thawed, lysed using a homogenizer, and pelleted at 30,000 × *g* for 20 min. PLCCγ SH2 was purified from the soluble fraction by cation exchange chromatography (GE Healthcare; HiTrap SP XL) followed by buffer exchange into 50 mM sodium citrate pH 5.5 buffer, 100 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP) for gel filtration (GE Healthcare; Hiload 10/60 Superdex, 75 pg). The collected fractions were concentrated and exchanged into NMR buffer (25 mM sodium arsenate, 25 mM sodium citrate, pH 4–8, 1 mM TCEP). SH2 pH titrations were performed by exchanging NMR samples into fresh buffer and carefully measuring the pH before and after the experiments. All samples used in the present studies were ∼1.0–1.5 mM in protein.

NMR Spectroscopy

All NMR experiments were recorded at 10 °C (Im7) or 20 °C (SH2) on Varian Inova spectrometers equipped with room-temperature triple-resonance probes and pulsed-field gradient capabilities. At each pH measured (both for Im7 and the PLCCγ SH2 domain) the following datasets were recorded at 11.7 T: ^{1}H-^{13}C heteronuclear single-quantum coherence (2) (for assignment of $^{13}C_{\varepsilon 1}$ and $^{13}C_{\delta 2}$ chemical shifts in the ground state), ^{1}H-^{13}C IPAP (3) (for measurement of $^1J_{C\varepsilon H\varepsilon}$ couplings of the ground state), ^{1}H-^{15}N heteronuclear multiple-bond correlation (4) (for measurement of $^{15}N_{\varepsilon 1}$ and $^{15}N_{\delta 2}$ chemical shifts, ground state), and a $^1H_{\delta 2}$-$^{13}C_{\gamma}$ plane (measured using the scheme of Fig. S2*C* with $T_{rlx} = 0$). $^1H_{\varepsilon 1}$ and $^1H_{\delta 2}$ chemical shifts, measured in the various experiments, were averaged (separately).

Details of the various chemical exchange saturation transfer (CEST) experiments used in this study are provided in Table S2. All experiments include a single reference plane, where $T_{rlx} = 0$, to determine the longitudinal relaxation rate of the ^{13}C spin in question, that is required for data analysis (5). For each type of CEST experiment, a pair of datasets with different weak B_1 field strengths (applied during T_{rlx}) was acquired with the fields differing typically by a factor of 2 to 4 (Table S2). We have found that recording multiple CEST profiles in this manner helps ensure that accurate chemical exchange parameters are obtained, even in the presence of additional fast exchange processes involving the excited state (6). The specific B_1 field strengths and T_{rlx} delays were chosen based on computer simulations of CEST profiles using rough estimates of the expected chemical exchange properties. Optimal parameters were those (B_1,T_{rlx}) pairs that result in the largest change in CEST profiles from a small change in (p_I,k_{ex}). Experimental B_1 field strengths were calibrated by monitoring the intensity of an on-resonance signal as a function T_{rlx}, as described previously (7). To have an accurate, independent measure of the chemical exchange parameters, backbone amide ^{15}N CEST experiments (5) were acquired at pH 6.25 and 6.60. At lower pH values, the backbone spectra of Im7 become prohibitively exchange-broadened so that ^{15}N CEST datasets cannot be analyzed.

An obvious issue relates to the signal-to-noise (S/N) ratio required for the accurate measurement of scalar coupling and chemical shift values. In general, it is difficult to come up with a single S/N number because this depends critically on the exchange parameters (p_I,k_{ex}), chemical shift differences, effective line widths of major and minor state dips in CEST profiles, number and strengths of the B_1 fields used, spacing of each B_1 that defines the CEST profile, and, of course, the accuracy that is needed. In the present case where $p_I \approx 2\%$, $k_{ex} \approx 100\ s^{-1}$ (pH 6.3, 10 °C), simulations show that a S/N of at least 150 is required in CEST profiles for the measurement of accurate $^1J_{C\varepsilon H\varepsilon}$ couplings (corresponding to errors in CEST points on the order of 0.6% of the intensity of the CEST baseline). We have therefore recorded all of our datasets with a S/N ≥ 150 (between ∼150 and 300). Because a 1D experiment (Fig. S2*A*) was used, the acquisition times for H40 and H47 were optimized separately to achieve the desired sensitivity, as indicated in Table S2. In general, the S/N required for measurement of chemical shifts is not as demanding, however, here too we have recorded spectra (in this case using the 2D schemes of Fig. S2 *B* and *C*) with S/N values on the order of or greater than 100, that required ∼20 h of measurement time for each complete CEST dataset.

Data Analysis

NMR spectra were processed with the NMRPipe suite of programs (8). Basic peak-picking of the pH titration data and visualization were carried out with the Sparky program (9). Peak intensities were extracted from the pseudo-3D CEST datasets by fitting resonances to a Gaussian model using the nlinLS function provided with NMRPipe, which uses common fitting parameters of peak positions and line widths across the pseudodimension corresponding to the position of the ^{13}C or ^{15}N carrier.

Numerical fitting of the CEST intensity profiles to a model of two-site chemical exchange was carried out with an in-house Python program, ChemEx (available from the authors upon request) (10). In the case of proton-decoupled experiments, intensities are calculated by evolving the density matrix of a single exchanging spin in Liouville space, as described previously (11, 12). In this case only the evolution of in-phase X, Y, and Z magnetization of the ground and excited states need be considered. In the case of U-^{13}C–labeled proteins, the presence of multiplet structure that arises from homonuclear ^{13}C-^{13}C couplings can be accounted for by simulating each spectral line separately, assuming that all lines relax identically (13), thus contributing no additional parameters to the fitting protocol. Fitting parameters include five nucleus-specific values ($\{\varpi_N, \Delta\varpi = \varpi_I - \varpi_N, R_1, R_{2,N}, \Delta R_2 = R_{2,I} - R_{2,N}\}$), two global parameters ($\{k_{ex} = k_{NI} + k_{IN}, p_I\}$), and a scaling factor (I_0) for each profile. Here, $\varpi_{N(I)}$ is the native (intermediate) state chemical shift, R_1 is the longitudinal relaxation rate for the exchanging spin that is assumed to be the

same in each of the exchanging states (5, 6, 12), $R_{2,N(I)}$ is the transverse relaxation rate in the native (intermediate) state, and p_I is the fractional population of the intermediate state. In the analysis of $^{13}C_{\delta 2}$ CEST profiles, $^1J_{C\delta 2C\gamma}$ was set to 72 Hz (14, 15) (see below) whereas in the analysis of $^{13}C_\gamma$ datasets, the magnetization was split into four components to additionally account for the $^1J_{C\beta C\gamma}$ coupling of 51 Hz (14). Note that the $^{13}C_{\varepsilon 1}$ nucleus is "isolated" by the two imidazole nitrogens; one-bond ^{13}C-^{13}C couplings are not relevant, while the relatively small ^{13}C-^{15}N couplings have not been taken into consideration. Long-range couplings are also not considered in the analysis. Similarly, backbone amide ^{15}N CEST profiles are analyzed without inclusion of the one- and two-bond couplings to neighboring backbone ^{13}CO and $^{13}C_\alpha$ nuclei.

Nitrogen CEST experiments are preferred over carbon-based methods for estimating exchange parameters because of their generally higher sensitivity, ease of analysis, and the abundance of probes. However, for Im7 at low pH values the intermediate population increases significantly (Fig. 4 *A* and *B*) and the quality of ^{15}N-^1H spectra become severely degraded. Fortunately, histidine $^{13}C_{\varepsilon 1}$-$^1H_{\varepsilon 1}$ correlations for Im7 are well resolved and can be quantified over the complete pH range studied. Here we have used the scheme of Fig. S2*A* with the addition of ^1H decoupling during the CEST interval to obtain p_I values. To evaluate the accuracy of the extracted parameters from the ^{13}C-based approach, we have measured both ^{13}C and ^{15}N CEST profiles at pH 6.6, where p_I is sufficiently small so that ^{15}N datasets can still be quantified. Extracted chemical exchange parameters from fits of the ^{13}C dataset ($k_{ex} = 321 \pm 12\ s^{-1}$, $p_I = 1.55 \pm 0.02\%$) agree well with those obtained from the backbone amide CEST experiment ($k_{ex} = 323 \pm 4\ s^{-1}$, $p_I = 1.54 \pm 0.01\%$), albeit with slightly elevated errors that reflect the limited number of probes used in the fits. It is clear that robust measures can be obtained from the $^{13}C_{\varepsilon 1}$-$^1H_{\varepsilon 1}$ spin pair.

As described above, the analysis of ^1H decoupled CEST profiles has been carried out using a procedure that is detailed in the literature (5, 6, 13). In the experiment of Fig. S2*A* that is used to obtain $^1J_{C\varepsilon H\varepsilon}$, there is no ^1H decoupling during the CEST interval T_{rlx} and a proper fit of the resulting CEST profiles must naturally take into account magnetization evolution in a two-spin ^{13}C-^1H system. In this case evolution of density elements, **m**, can be expressed as

$$\mathbf{m}(T_{rlx}) = e^{i\mathbf{L}T_{rlx}} \cdot \mathbf{m}(0) \qquad \text{[S1]}$$

where $\mathbf{m}(0)$ is evaluated at the beginning of the CEST period, and **L** is the Liouvillian matrix describing the evolution of the ^1H-^{13}C spin system. In principle, the Cartesian basis set for a two-spin system undergoing two-site chemical exchange contains 30 elements, neglecting the unity element (10, 16, 17). However, in the absence of ^1H pulses (as in the present case) terms related to antiphase proton magnetization and multiple-quantum coherences can be eliminated, leaving a much simpler 12-term basis set

$$\mathbf{m} = \left[C_X^N, C_Y^N, C_Z^N, 2H_ZC_X^N, 2H_ZC_Y^N, 2H_ZC_Z^N, C_X^I, \ldots, 2H_ZC_Z^I\right]^T \qquad \text{[S2]}$$

where A_B^D is the B component of A magnetization derived from state D and superscript T denotes transpose. Note that the unity element is not included as terms that are related to equilibrium polarization are subtracted by the phase cycle of each of the experiments. The Liouvillian can thus be written as

$$\mathbf{L} = \begin{bmatrix} \mathbf{R_N} & \mathbf{0_6} \\ \mathbf{0_6} & \mathbf{R_I} \end{bmatrix} + \begin{bmatrix} -k_{NI} & k_{IN} \\ k_{NI} & -k_{IN} \end{bmatrix} \otimes \mathbf{1_6} \qquad \text{[S3]}$$

where $\mathbf{0_6}$ ($\mathbf{1_6}$) is the 6D null (identity) matrix, and the relaxation submatrix, $\mathbf{R}_j$, describing the density elements in state $j = \{N,I\}$ is given by

$$\mathbf{R}_j = -\begin{bmatrix} R_{TR}^j & \omega_C^j & 0 & \eta_{TR}^j & \pi J_{C\varepsilon H\varepsilon}^j & 0 \\ -\omega_C^j & R_{TR}^j & \omega_1 & -\pi J_{C\varepsilon H\varepsilon}^j & \eta_{TR}^j & 0 \\ 0 & -\omega_1 & R_Z & 0 & 0 & \eta_Z \\ \eta_{TR}^j & \pi J_{C\varepsilon H\varepsilon}^j & 0 & R_{AP}^j & \omega_C^j & 0 \\ -\pi J_{C\varepsilon H\varepsilon}^j & \eta_{TR}^j & 0 & -\omega_C^j & R_{AP}^j & \omega_1 \\ 0 & 0 & \eta_Z & 0 & -\omega_1 & R_{ZZ} \end{bmatrix}. \qquad \text{[S4]}$$

R_{TR}, R_Z, R_{AP}, and R_{ZZ} are the autorelaxation rates of ^{13}C transverse, ^{13}C longitudinal, ^{13}C antiphase, and ^{13}C-^1H two-spin order magnetization, respectively. ^{13}C-^1H dipole-^{13}C chemical shift anisotropy cross-correlated relaxation that connects longitudinal (transverse) and two-spin order (antiphase) elements is accounted for by η_Z (η_{TR}). The carbon chemical shift offset from the carrier frequency is denoted by ω_C (rad/s) and ω_1 is the strength of the B_1 field (rad/s) applied during T_{rlx}. The final fitting parameters includes up to 11 nucleus-specific parameters, $\{\varpi_N, \Delta\varpi, {}^1J_{C\varepsilon H\varepsilon}^N, \Delta J = {}^1J_{C\varepsilon H\varepsilon}^I - {}^1J_{C\varepsilon H\varepsilon}^N, R_Z, R_{ZZ}, \eta_Z, R_{2,N}, \Delta R_2 = R_{2,I} - R_{2,N}, \eta_{TR}^N, \Delta\eta = \eta_{TR}^I - \eta_{TR}^N\}$. Here we have assumed that terms such as η_Z and R_{ZZ}, that are not well constrained by CEST profiles, do not change between ground and excited states. An additional simplification can be made whereby $R_{AP}^j = R_{TR}^j + R_{ZZ} - R_Z$. Global parameters, $\{k_{ex}, p_I\}$, are held fixed to values determined from fits of the ^{15}N CEST experiments. In practice, the inclusion of ΔR_2 and $\Delta\eta$ is determined by fits of the data. In the present application inclusion of these two parameters results in only a modest improvement in the quality of the fit, and has therefore been neglected. Importantly, the addition or removal of either of these parameters has little effect on the extracted ΔJ values and their uncertainties.

Error Analysis

Uncertainties in reported I state pK_a values were estimated from 1,000 Monte Carlo calculations (18) that included experimental errors in measured $^1J_{C\varepsilon H\varepsilon}$ values, errors in measured pH values (0.1 of a unit) and uncertainties in the J-coupling endpoints (Eq. **1**). The SD of the distribution of pK_a values determined in this manner is defined as the error. The 68% and 95% confidence limits shown in Fig. 3 *A* and *B* were calculated in a similar manner by fitting $^1J_{C\varepsilon H\varepsilon}$ values taking into account errors as described above. Confidence limits were calculated from the 1,000 fitted $^1J_{C\varepsilon H\varepsilon}$ vs. pH profiles so produced. The distributions shown in Fig. 4 *B*–*D* were obtained similarly, by including errors in measured pH values and input I state pK_as (Fig. 4 *B* and *C*) or rates [Fig. 4*D*, minimum error of 0.1 kJ/mol was applied to the values of $-RT\ln(k_{ij})$]. In Fig. 4*C*, however, the experimental data were not fitted; as described in the text, $\Delta\Delta G_{N\to I}$ was obtained in two different ways and the agreement shown as a means of cross-validation.

Fits of k_{IN}/k_{NI} Rate–pH Profiles to Extract pK_a Values of Histidine Residues in the Transition State

The rate of a reaction from A to B can be expressed using transition state (TS) theory (19) as,

$$k_{A\to B} = \frac{\kappa k_B T}{h} \exp\left(-\frac{\Delta G_{A\to TS}}{RT}\right) \qquad \text{[S5]}$$

where κ is a transmission coefficient, k_B and T are the Boltzmann constant and absolute temperature, respectively, h is Plank's

constant, R is the gas constant, and $\Delta G_{A\to TS}$ is the difference in free energy between state A and the TS. Combining Eq. **S5** with Eq. **2** of the main text gives

$$-RT\ln k_{IN} = -RT\ln\left(\frac{\kappa k_B T}{h}\right) + \Delta G^{o}_{I\to TS} - RT\sum_{m=1}^{i}\ln\left(\frac{K_a^{TS,m}+[H^+]}{K_a^{I,m}+[H^+]}\right) \quad \textbf{[S6]}$$

where $K_a^{P,m}$ is the acid dissociation constant of group m in state P and the summation includes both H40 and H47 (i.e., $i = 2$). Eq. **S6** can be written as

$$-RT\ln k_{IN} = C_{IN} - RT\sum_{m=1}^{i}\ln\left(\frac{K_a^{TS,m}+[H^+]}{K_a^{I,m}+[H^+]}\right) \quad \textbf{[S7]}$$

where C_{IN} is independent of pH. In a similar manner we can write

$$-RT\ln k_{NI} = -RT\ln\left(\frac{\kappa k_B T}{h}\right) + \Delta G^{o}_{N\to TS} - RT\sum_{m=1}^{i}\ln\left(\frac{K_a^{TS,m}+[H^+]}{K_a^{N,m}+[H^+]}\right). \quad \textbf{[S8]}$$

As discussed in the text, $pK_a^{N,m} < pH$ over the range of pH values where the measurements are made. Thus, Eq. **S8** becomes

$$-RT\ln k_{NI} = C_{NI} - RT\sum_{m=1}^{i}\ln\left(K_a^{TS,m}+[H^+]\right) \quad \textbf{[S9]}$$

where C_{NI} collects all pH-independent terms. Values of k_{IN}, k_{NI} were fit simultaneously to expressions given by Eqs. **S7** and **S9**, however the error surface was shallow and a large range of $pK_a^{TS,m}$ were obtained. We therefore fit to a model that assumes a single averageTS pK_a value for both histidines, with the result shown in Fig. 4*D*.

1. Hansen AL, Kay LE (2011) Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. *J Biomol NMR* 50(4):347–355.
2. Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69(1):185–189.
3. Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 131(2):373–378.
4. Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci* 2(4):543–558.
5. Vallurupalli P, Bouvignies G, Kay LE (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* 134(19):8148–8161.
6. Hansen AL, Bouvignies G, Kay LE (2013) Probing slowly exchanging protein systems via ^{13}Cα-CEST: Monitoring folding of the Im7 protein. *J Biomol NMR* 55(3):279–289.
7. Guenneugues M, Berthault P, Desvaux H (1999) A method for determining B1 field inhomogeneity. Are the biases assumed in heteronuclear relaxation experiments usually underestimated? *J Magn Reson* 136(1):118–126.
8. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
9. Kneller DG, Kuntz ID (1993) UCSF Sparky: An NMR display, annotation and assignment tool. *J Cell Biochem* 53(S17C):254.
10. Bouvignies G, Kay LE (2012) Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J Phys Chem B* 116(49):14311–14317.
11. Vallurupalli P, Bouvignies G, Kay LE (2013) A computational study of the effects of (13) C-(13) C scalar couplings on (13) C CEST NMR spectra: Towards studies on a uniformly (13) C-labeled protein. *ChemBioChem* 14(14):1709–1713.
12. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM (2011) Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* 480(7376):268–272.
13. Bouvignies G, Vallurupalli P, Kay LE (2014) Visualizing side chains of invisible protein conformers by solution NMR. *J Mol Biol* 426(3):763–774.
14. Bystrov VF (1976) Spin—spin coupling and the conformational states of peptide systems. *Prog Nucl Magn Reson Spectrosc* 10(2):41–82.
15. Sudmeier JL, et al. (1996) HCN, a triple-resonance NMR technique for selective observation of histidine and tryptophan side chains in 13C/15N-labeled proteins. *J Magn Reson B* 113(3):236–247.
16. Allard P, Helgstrand M, Härd T (1998) The complete homogeneous master equation for a heteronuclear two-spin system in the basis of cartesian product operators. *J Magn Reson* 134(1):7–16.
17. Helgstrand M, Härd T, Allard P (2000) Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J Biomol NMR* 18(1): 49–63.
18. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing*, ed Press WH (Cambridge Univ Press, Cambridge, UK), 3rd Ed.
19. Fersht A (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, New York).

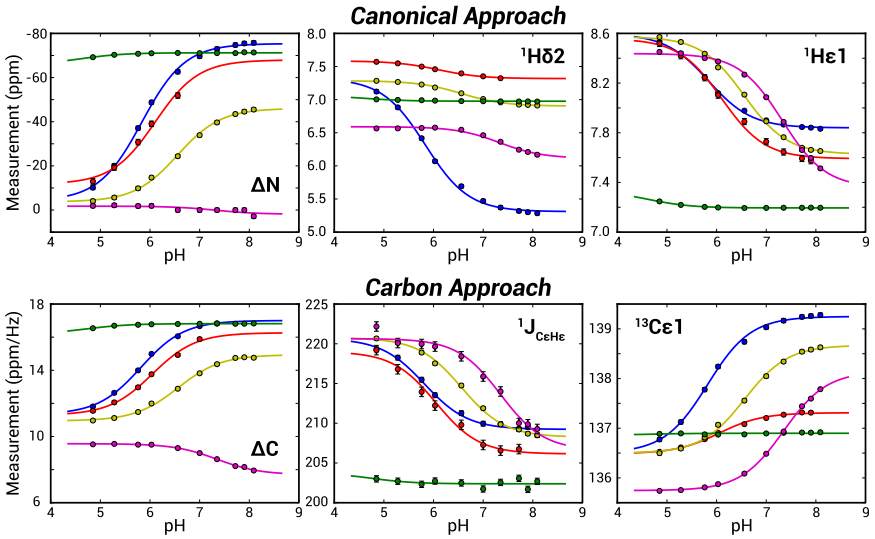

Fig. S1. pH titration profiles of H6 (blue), H13 (red), H23 (yellow), H57 (green), and H87 (purple) of the PLCγ SH2 domain. Solid lines in *Upper* are the result of per-residue fits of $\Delta^{N,meas} = \varpi_{N\varepsilon 2} - \varpi_{N\delta 1}$, $\varpi_{H\delta 2}$, and $\varpi_{H\varepsilon 1}$ as a function of pH to a single protonation event (single pK_a for each histidine). Profiles are fit to Eq. **1** of the main text. The limiting $\Delta^{N,meas}$ value at high pH, $\Delta^{N,meas}_{high}$, (Q_{high} in Eq. **1**) can be recast in terms of the populations of δ and ε tautomeric states in the limit that only the neutral form of the imidazole ring is present, p^{high}_{δ} and p^{high}_{ε}, according to the relation $\Delta^{N,meas}_{high} = 82\ \text{ppm} \times p^{high}_{\delta} - 82\ \text{ppm} \times p^{high}_{\varepsilon}$, $p^{high}_{\delta} + p^{high}_{\varepsilon} = 1$, where $\Delta^{N} = 82$ ppm, −82 ppm are the canonical values for the δ, ε tautomers, respectively (1). Resulting pK_a (p^{high}_{ε}) values from the combined ^{15}N and ^{1}H fits are 5.84 ± 0.03 (95.9%), 6.10 ± 0.03 (91.5%), 6.55 ± 0.03 (78.0%), 4.63 ± 0.20 (93.4%), and 7.32 ± 0.03 (48.8%) for residues 6, 13, 23, 57, and 87, respectively. *Lower* are fit analogously using the three carbon datasets, $\Delta^{C,meas} = \varpi_{C\gamma} - \varpi_{C\delta 2}$, $\varpi_{C\varepsilon 1}$, and $^{1}J_{C\varepsilon H\varepsilon}$ vs. pH. Values of p^{high}_{δ} and p^{high}_{ε} were obtained from the relation $\Delta^{C,meas}_{high} = 18.8\ \text{ppm} \times p^{high}_{\varepsilon} + 0.1\ \text{ppm} \times p^{high}_{\delta}$ (Table S1). Resulting pK_a (p^{high}_{ε}) values are 5.82 ± 0.03 (90.0%), 6.05 ± 0.03 (86.1%), 6.56 ± 0.03 (79.0%), 4.74 ± 0.46 (93.4%), and 7.35 ± 0.03 (48.8%) for H6, H13, H23, H57 and H87, respectively.

1. Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci* 2(4):543–558.

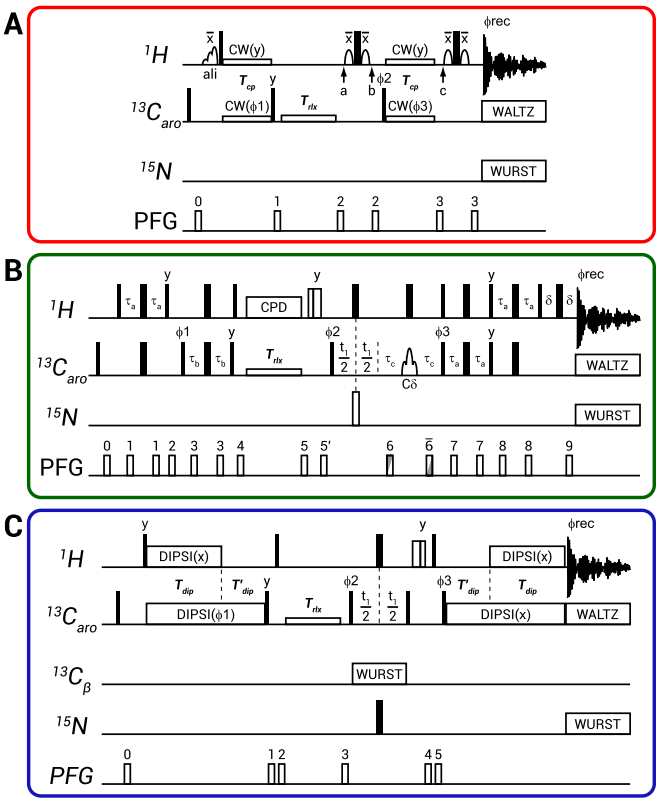

Fig. S2. ^{13}C CEST-based pulse sequences for the measurement of histidine pK_a values and tautomer populations in invisible protein states. Uniform ^{13}C labeling is assumed. In all sequences, ^{1}H polarization is transferred from aromatic protons to the desired carbon nucleus to generate ^{13}C *Z* magnetization. During the relaxation period, T_{rlx}, a weak B_1 field (Table S2) is applied at an offset frequency that is arrayed to produce the characteristic CEST saturation profiles shown in Fig. 2 *B–E*. Magnetization is then encoded with the carbon chemical shift (*B* and *C* only) and transferred to the aromatic proton for detection. Simultaneous carbon and nitrogen broadband decoupling is used during acquisition in all sequences, to suppress one- and two-bond ^{1}H-^{13}C and ^{1}H-^{15}N couplings. Broadband carbon decoupling is achieved with a 1.6-kHz WALTZ-16 sequence (1) at 11.7 and 18.8 T, whereas adiabatic ^{15}N WURST decoupling (2) with $B_{1,max}$ of 0.37 (0.49) kHz is used at 11.7 T (18.8 T). The latter decoupling scheme uses a 10-ms constant-adiabaticity WURST-20 shape with a 100-ppm bandwidth centered at 205 ppm. Radiofrequency fields during T_{rlx}, offset frequency ranges and additional experimental parameters are listed in Table S2. (*A*) One-dimensional CEST pulse sequence for measurement of $^{1}J_{C\varepsilon H\varepsilon}$. Solid narrow (wide) rectangles indicate maximum power 90° (180°) pulses. All pulses are *x* phase, unless otherwise noted. Short, open rectangles indicate weak, continuous wave (CW) irradiation used for magnetization transfer (T_{cp}) or CEST (T_{rlx}). Selective magnetization transfer is achieved through the Hartmann–Hahn effect (3), with matching proton and carbon weak B_1 fields of $\sqrt{3/4}\cdot{}^{1}J_{C\varepsilon H\varepsilon}$ ~95 Hz (4) for a duration $T_{cp} = 1/{}^{1}J_{C\varepsilon H\varepsilon} = 4.75$ ms. During this time both ^{1}H and ^{13}C carriers are positioned at the resonance frequency of the spin in question. At points a and c the ^{1}H carrier is shifted to the water signal with the pulses on either side of the high power ^{1}H 180°s applied for a duration of ~1.5 ms with a rectangular shape, optimized to suppress the residual water signal. The proton carrier is returned to the aromatic proton resonance position at point b and after acquisition. The carbon carrier frequency is set on resonance with the histidine $^{13}C_{\varepsilon 1}$ signal throughout the experiment, except during the T_{rlx} period where the frequency is arrayed to produce the CEST saturation profile (one dataset for each frequency). The Eburp-1 pulse (5) (90°) marked "ali" at the beginning of the sequence has a length 1.5 (0.9) ms at 11.7 T (18.8 T) and is centered at ~1.85 ppm so as to excite the aliphatic region and water signal, leaving the aromatic spins unaffected. The net result is that aliphatic magnetization is preserved along *Z* during the interscan delay, enhancing the recovery rate of the aromatic proton signal of interest. An eight-step phase cycle was used with $\phi1 = 2(x)\ 2(-x)$, $\phi2 = 4(y)\ 4(-y)$, $\phi3 = x\ -x$, and $\phi rec = x\ -x\ -x\ x\ -x\ x\ x\ -x$. Gradient strengths (lengths) in units of G/cm (ms) are g0 = 16 (1.0), g1 = 12 (0.5), g2 = −30 (0.5), and g3 = 20 (0.6). (*B*) Pulse scheme for measurement of $^{13}C_{\delta 2}$ chemical shifts. Solid narrow (wide) rectangles indicate maximum power 90° (180°) pulses, with all pulses along the *X* axis, unless otherwise noted. ^{1}H, ^{13}C pulses are applied at 4.7 and 126 ppm, respectively; for each ^{13}C plane the carbon carrier is positioned at a different frequency during T_{rlx}, as described above. ^{1}H decoupling is achieved during the CEST period using a 3.77-kHz 90_x-240_y-90_x element (6) (11.7 T), as described previously (7). A proton purge element (for water suppression) is applied after the CEST block with *x* and *y* pulse durations of 2 and 3 ms and a field of 7 kHz. The nitrogen pulse in the middle of the t_1 period is of the composite variety (8), centered at 180 ppm, wherea the ^{13}C Reburp pulse (5) (0.75 ms, 11.7 T) denoted "Cδ" refocuses evolution from the $^{1}J_{C\gamma C\delta 2}$ coupling, preventing loss of magnetization. Owing to the isolated nature of the histidine resonances studied here, magnetization evolution from the $^{1}J_{C\gamma C\delta 2}$ coupling is allowed to occur during the short t_1 acquisition time (typically ~5–7 ms < $0.5/{}^{1}J_{C\gamma C\delta 2}$ so that multiplet structure is not observed). Delays τ_a, $\tau_b = 1/(4J_{CH})$ and δ are set to 1.1, 1.25, and 0.45 ms, respectively, while $\tau_c = \tau_b - 0.5\times$pwc, where pwc is the pulse width of the Cδ pulse. A four-step phase cycle was used with $\phi1 = x\ -x$, $\phi2 = 2(y)\ 2(-y)$, $\phi3 = x$, and $\phi rec = x\ -x\ -x\ x$. Gradient strengths (lengths) in units of G/cm (ms) used are g0 = 15 (1.0), g1 = 20 (0.5), g2 = 30 (0.8), g3 = −8 (0.6), g4 = −32 (1.0), g5 = −20 (0.7), g5′ = −20 (0.91), g6 = 30 (0.2), g7 = 12 (0.2), g8 = 18 (0.5), and g9 = 29.6 (0.1). Quadrature detection is achieved in the indirect dimension by recording a pair of spectra for each t_1 value with (ϕ3, g6) and (ϕ3 + 180, −g6) (9, 10). Axial peaks are shifted to the edge of the spectrum by incrementing ϕ2 and ϕrec by 180° for each successive complex t_1 point (11). (*C*) Pulse scheme for the measurement of $^{13}C_{\gamma}$ chemical shifts. Solid narrow (wide) rectangles indicate maximum power 90° (180°) pulses (along the *X* axis, unless otherwise noted). ^{1}H and ^{13}C pulses are centered at 4.7 and 126 ppm respectively. Magnetization transfer is achieved using a scheme where both hetero- ($^{1}H_{\delta 2}$ to $^{13}C_{\delta 2}$) and homonuclear ($^{13}C_{\delta 2}$ to $^{13}C_{\gamma}$) transfers occur simultaneously during T_{dip}, with a subsequent homonuclear transfer time of T'_{dip}. Simulations and experiment have established that either DIPSI-1 or DIPSI-2 schemes (12) can be used, with similar net transfer times (9 ms). In the case of DIPSI-1, three cycles (T_{dip}) and two cycles (T_{dip}) are used along with 8.5-kHz ^{1}H, ^{13}C B_1 fields, whereas the DIPSI-2 transfer uses two (T_{dip}) and one cycles (T'_{dip}) and a 10-kHz field. Transfer efficiencies on the order of 85–90% are obtained, neglecting relaxation. $^{13}C_{\beta}$ decoupling is achieved during t_1 by the application of a pair of constant adiabaticity WURST-8 fields (2, 13, 14) (combined maximum field strength of 1 kHz, 11.7 T). The first is centered at 35.5 ppm with a bandwidth of 25 ppm, whereas a second symmetric decoupling element is centered at 234.5 ppm, with a reverse sweep to minimize Bloch–Siegert artifacts (15, 16). Because the sequence also generates ($^{13}C_{\varepsilon 1}$, $^{1}H_{\varepsilon 1}$) correlations, a 180° proton pulse is applied at the center of t_1

Legend continued on following page

to remove splittings from $^{1}J_{C\varepsilon H\varepsilon}$ couplings. A four-step phase cycle was used with $\phi1 = 2(x)\ 2(-x)$, $\phi2 = x$, $\phi3 = y\ -y$, and $\phi rec = x\ -x\ -x\ x$. Gradient strengths (lengths) in units of G/cm (ms) are g0 = 15 (1.0), g1 = 30 (1.0), g2 = 20 (1.0), g3 = 16 (1.0), g4 = 24 (1.0), and g5 = −18 (1.0). Quadrature detection is achieved in the indirect dimension using the States–time proportional phase incrementation approach (11), by incrementing $\phi2$.

1. Shaka AJ, Keeler J, Frenkiel T, Freeman R (1983) An improved sequence for broadband decoupling: WALTZ-16. *J Magn Reson* 52(2):335–338.
2. Kupče Ē, Freeman R (1996) Optimized adiabatic pulses for wideband spin inversion. *J Magn Reson A* 118(2):299–303.
3. Pelupessy P, Chiarparin E, Bodenhausen G (1999) Excitation of selected proton signals in NMR of isotopically labeled macromolecules. *J Magn Reson* 138(1):178–181.
4. Pelupessy P, Chiarparin E (2000) Hartmann-Hahn polarization transfer in liquids: An ideal tool for selective experiments. *Concepts Magn Reson* 12(3):103–124.
5. Geen H, Freeman R (1991) Band-selective radiofrequency pulses. *J Magn Reson* 93(1):93–141.
6. Levitt MH (1982) Symmetrical composite pulse sequences for NMR population inversion. II. Compensation of resonance offset. *J Magn Reson* 50(1):95–110.
7. Vallurupalli P, Bouvignies G, Kay LE (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* 134(19):8148–8161.
8. Levitt MH, Freeman R (1979) NMR population inversion using a composite pulse. *J Magn Reson* 33(2):473–476.
9. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114(26): 10663–10665.
10. Schleucher J, Sattler M, Griesinger C (1993) Coherence selection by gradients without signal attenuation: Application to the three-dimensional HNCO experiment. *Angew Chem Int Ed Engl* 32(10):1489–1491.
11. Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J Magn Reson* 85(2):393–399.
12. Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. *J Magn Reson* 77(2):274–293.
13. Matsuo H, Kupce E, Li H, Wagner G (1996) Increased sensitivity in HNCA and HN(CO)CA experiments by selective C beta decoupling. *J Magn Reson B* 113(1):91–96.
14. Kupče Ē, Wagner G (1996) Multisite band-selective decoupling in proteins. *J Magn Reson B* 110(3):309–312.
15. Zhang S, Gorenstein DG (1998) Bloch-Siegert shift compensated and cyclic irradiation sidebands eliminated, double-adiabatic homonuclear decoupling for 13C- and 15N-double-labeled proteins. *J Magn Reson* 132(1):81–87.
16. Kupce Ē, Freeman R (1995) Stretched Adiabatic Pulses for Broadband Spin Inversion. *J Magn Reson A* 117:246–256.

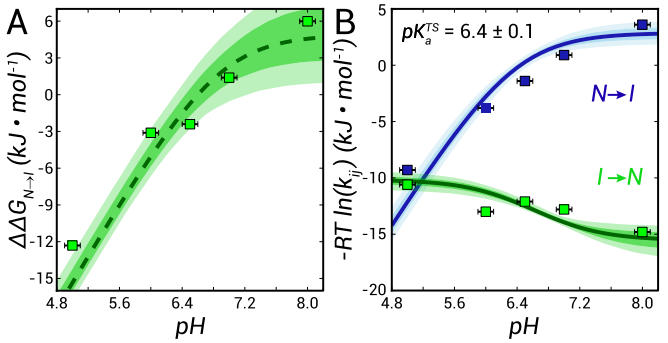

Fig. S3. Further validation of NMR methodology. (*A*) The pH dependence of the relative stabilities of *I* and *N* have been calculated using thermodynamic values determined by Gorski et al. (1). For comparison purposes, the experimental values from table 2 of Gorski et al. (1) are recast as $\Delta G_{N\rightarrow I,pH} = \Delta G_{U\rightarrow I,pH} - \Delta G_{U\rightarrow N,pH}$, and $\Delta\Delta G_{N\rightarrow I,pH} = \Delta G_{N\rightarrow I,pH} - \Delta G_{N\rightarrow I,ref}$, where $\Delta G_{N\rightarrow I,ref}$ = 12.3 kJ/mol (pH 6.6). The dashed line denotes the expected pH dependence of stability calculated from Eq. **3** and the pK_a values of H40, H47 of *I*, measured as described in the main text. (*B*) The pH dependence of k_{IN} and k_{NI} from fluorescence-based literature values (1). Solid lines are best-fits of the data using Eqs. **S7** and **S9** assuming *I* state pK_a values determined in the present work. Shaded regions in *A* and *B* represent the 68% (dark) and 95% (light) confidence range. The extracted average pK_a(TS) value for H40/H47 is in excellent agreement with that obtained from fits to the CEST data. It is worth noting that although stability or rate profiles vs. pH can be obtained for Im7 through careful fluorescence studies, only the NMR methodology described here provides site-specific pK_a values and tautomer states for histidine residues of the invisible *I* state.

1. Gorski SA, Capaldi AP, Kleanthous C, Radford SE (2001) Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J Mol Biol* 312(4):849–863.

Table S1. Literature values for the histidine ring carbon chemical shifts of interest along with those from the present work

Frequency, ppm	State	Reynolds et al.*	Cheng et al.†	Li and Hong‡	This work§
$\varpi_{C\gamma}$	δ	129.7	127.5 (129.6)	131.9	125.7 ± 2.2
	+	130.6	131.7	131.4	131.2 ± 0.7
	ε	137.7	140.2 (136.7)	140.0	135.7 ± 2.2
$\varpi_{C\delta2}$	δ	129.4	131.1 (127.6)	128.1	125.6 ± 2.1
	+	121.1	123.4	122.1	120.6 ± 1.3
	ε	122.8	117.3 (119.3)	116.0	116.9 ± 2.1
$\Delta^C = \varpi_{C\gamma} - \varpi_{C\delta2}$	δ	0.3	−3.6 (2.0)	3.8	0.1 ± 3.0
	+	9.5	8.3	9.3	10.6 ± 1.5
	ε	14.9	22.9 (17.4)	24.0	18.8 ± 3.0

All chemical shifts are rereferenced to 2-2-dimethyl-2-silapentane-5-sulfonate from their original values.
*Histidine (amino acid) and methyl-histidine values from Reynolds et al. (1).
†Histidine-containing dipeptide values from Cheng et al. (2). Listed are the extreme (average) chemical shifts from the set of dipeptides attributed to a particular conformational state.
‡Solid-state NMR pH titration values taken from Li and Hong (3). Chemical shifts for the ε tautomer are the average of the neutral and anionic ε tautomers reported.
§Values obtained from the present study using Biological Magnetic Resonance Data Bank data, as described in *A Strategy for Studies of Histidine Side Chains in Excited Protein States* (main text).

1. Reynolds WF, Peat IR, Freedman MH, Lyerla JR, Jr. (1973) Determination of the tautomeric form of the imidazole ring of L-histidine in basic solution by carbon-13 magnetic resonance spectroscopy. *J Am Chem Soc* 95(2):328–331.
2. Cheng F, Sun H, Zhang Y, Mukkamala D, Oldfield E (2005) A solid state 13C NMR, crystallographic, and quantum chemical investigation of chemical shifts and hydrogen bonding in histidine dipeptides. *J Am Chem Soc* 127(36):12544–12554.
3. Li S, Hong M (2011) Protonation, tautomerization, and rotameric structure of histidine: A comprehensive study by magic-angle-spinning solid-state NMR. *J Am Chem Soc* 133(5): 1534–1544.

Table S2. Acquisition parameters used to record data

Experiment	B_0, T	nt*	pts†	sw, ppm‡	at, ms§	B_1, Hz	T_{rlx}, ms	Carrier, ppm¶	Offsets/range, Hz
pH 6.60, 10 °C, $k_{ex} = 323 \pm 4$ s^{-1}, $p_l = 1.5 \pm 0.1$%									
2D ^{15}N CEST‖	11.7	4	48/512	17.6/16.0	53.4/64.0	20	600	118.8	59/−1,000→750
						40	400	118.8	59/−1,000→750
1D $^{13}C_{\varepsilon 1}$ - H40 (Fig. S2*A*)	18.8	400 (×2)	384	15.0	32.0	15	1,000	140.8	88/−2,000→1,500
						45	400	140.8	88/−2,000→1,500
1D $^{13}C_{\varepsilon 1}$ (decoupled) - H40**	11.7	256	512	24.0	42.6	20	600	140.8	56/−1,500→1,000
						40	400	140.8	56/−1,500→1,000
1D $^{13}C_{\varepsilon 1}$ - H47 (Fig. S2*A*)	18.8	96 (×3)	512	15.0	42.6	15	1,000	139.4	74/−1,500→1,500
						45	400	139.4	74/−1,500→1,500
1D $^{13}C_{\varepsilon 1}$ (decoupled) - H47**	11.7	64	512	24.0	42.6	20	600	139.4	56/−1,500→1,000
						40	400	139.4	56/−1,500→1,000
2D $^{13}C_{\delta 2}$ (Fig. S2*B*)	11.7	16	20/512	28.0/16.0	5.5/64.0	30	500	125.5	42/−2,000→0
						60	400	125.5	42/−2,000→0
2D $^{13}C_{\gamma}$ (Fig. S2*C*)	11.7	8	32/512	32.0/16.0	7.8/64.0	20	600	128.2	55/−550→2,750
						40	400	128.2	55/−550→2,750
pH 6.25, 10 °C, $k_{ex} = 256 \pm 4$ s^{-1}, $p_l = 3.5 \pm 0.1$%									
2D ^{15}N CEST‖	11.7	4	48/512	17.6/16.0	53.4/64.0	25	125	118.8	59/−1,000→750
						100	100	118.8	59/−1,000→750
1D $^{13}C_{\varepsilon 1}$ - H40 (Fig S2*A*)	18.8	256 (×2)	384	15.0	32.0	10	1,000	140.8	88/−2,000→1,500
						30	250	140.8	88/−2,000→1,500
1D $^{13}C_{\varepsilon 1}$ - H47 (Fig. S2*A*)	18.8	96 (×2)	512	15.0	42.6	10	1,000	139.4	74/−1,500→1,500
						30	250	139.4	74/−1,500→1,500
pH 6.07, 10 °C, $k_{ex} = 167 \pm 9$ s^{-1}, $p_l = 17.4 \pm 0.4$%									
1D $^{13}C_{\varepsilon 1}$ (decoupled) - H40**	11.7	256	512	24.0	42.7	20	200	140.7	46/−1,000→1,000
						60	125	140.7	30/−1,000→1,000
1D $^{13}C_{\varepsilon 1}$ (decoupled) - H47**	11.7	256	512	24.0	42.7	20	200	139.4	46/−1,000→1,000
						60	125	139.4	30/−1,000→1,000
pH 5.73, 10 °C, $k_{ex} = 156 \pm 11$ s^{-1}, $p_l = 46.5 \pm 3.7$%									
2D $^{13}C_{\varepsilon 1}$ (decoupled)††	11.7	32	24/512	16.0/16.0	11.7/64.0	20	150	138.5	24/−700→700
						60	100	138.5	24/−700→700

*Number of transients (nt) per free induction decay. In parentheses are the number of duplicate measurements recorded of each experiment.
†Number of complex points (pts) acquired in t_1/t_2.
‡Spectral widths (sw) in the indirect/direct dimensions.
§Total acquisition time (at) for the indirect/direct dimensions.
¶Each CEST profile is generated from N experiments, with the ^{13}C carrier positioned over a range of frequencies extending between Ω1 to Ω2 Hz from the carrier position listed. For example, first line: 59 spectra were recorded with a 20-Hz B_1 field placed at offsets of −1,000 to 750 Hz from 118.8 ppm.
‖Acquired as described in Vallurupalli et al. (1).
**The pulse sequence in Fig. S2*A* was modified to include a 90_x-240_y-90_x proton decoupling element during the T_{rlx} period, as described for the experiment of Fig. S2*B*.
††Data acquired using the scheme of Fig. S2*C* with $T_{dip}' = 0$ and with the 90_x-240_y-90_x proton decoupling sequence added to the T_{rlx} period.

1. Vallurupalli P, Bouvignies G, Kay LE (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* 134(19):8148–8161.

Table S3. Im7 histidine pK_a values

	State		
Histidine	Native	<Transition>	Intermediate
H40	<5	6.3 ± 0.1	6.9 ± 0.3
H47	<5	6.3 ± 0.1	6.9 ± 0.2

Im7 histidine side-chain pK_a values as determined in the present work. Owing to the disappearance of signals in the NMR spectra with decreasing pH, only an upper limit of the native state pK_a values is determined. A model (Eqs. **S7** and **S9**) was used to determine an average pK_a of the histidines in the TS. Intermediate state pK_a values are determined using the $^1J_{C\varepsilon H\varepsilon}$ couplings shown in Fig. 3 *A* and *B*.