Measurement of histidine pK_a values and tautomer populations in invisible protein states

Alexandar L. Hansen[a] and Lewis E. Kay[a,b,1]

[a]Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8; and [b]Molecular Structure and Function Program, Hospital for Sick Children, Toronto, ON, Canada M5G 1X8

Edited by Peter E. Wright, The Scripps Research Institute, La Jolla, CA, and approved March 10, 2014 (received for review January 10, 2014)

The histidine imidazole side chain plays a critical role in protein function and stability. Its importance for catalysis is underscored by the fact that histidines are localized to active sites in ∼50% of all enzymes. NMR spectroscopy has become an important tool for studies of histidine side chains through the measurement of site-specific pK_as and tautomer populations. To date, such studies have been confined to observable protein ground states; however, a complete understanding of the role of histidine electrostatics in protein function and stability requires that similar investigations be extended to rare, transiently formed conformers that populate the energy landscape, yet are often "invisible" in standard NMR spectra. Here we present NMR experiments and a simple strategy for studies of such conformationally excited states based on measurement of histidine $^{13}C_\gamma$, $^{13}C_{\delta 2}$ chemical shifts and $^{1}H_\varepsilon$-$^{13}C_\varepsilon$ one-bond scalar couplings. The methodology is first validated and then used to obtain pK_a values and tautomer distributions for histidine residues of an invisible on-pathway folding intermediate of the colicin E7 immunity protein. Our results imply that the side chains of H40 and H47 are exposed in the intermediate state and undergo significant conformational rearrangements during folding to the native structure. Further, the pK_a values explain the pH-dependent stability differences between native and intermediate states over the pH range 5.5–6.5 and they suggest that imidazole deprotonation is not a barrier to the folding of this protein.

conformationally excited protein states | Im7 protein-folding intermediate | chemical exchange | pH stability | CEST

Electrostatic interactions can be critically important to the stability and function of biological molecules (1, 2). Such interactions in proteins manifest most often through side chains of amino acids that can assume different protonation states and hence different charged forms that, for example, guide an enzyme reaction to completion (3–5). Histidine side chains play a particularly important role in the protein structure–function paradigm because they can exist in one of a pair of neutral tautomeric states or as a charged conformer, they serve as multiple hydrogen bond acceptors and donors, they are often localized to active sites of enzymes where they are integral to catalysis (5, 6), and their properties can be modified through metal binding and phosphorylation (7, 8). Moreover, the pK_a of the histidine imidazole moiety is usually close to physiological pH allowing it to serve as either an acid or as a base, which nature has exploited for a variety of different protein functions (3, 9).

It is well appreciated that solution NMR spectroscopy is a powerful technique for studying conformational equilibria of histidine residues in proteins and for measuring pK_a values that, in turn, can provide insight into protein function, especially in studies of enzyme mechanism (6, 10, 11). Initial studies of histidine in proteins date to the mid-1960s and for many years focused on recording 1D ^{1}H and ^{13}C NMR spectra (12–14). With the advent of multidimensional NMR spectroscopy and the development of ^{15}N-labeling methods many studies now rely on recording long-range ^{1}H-^{15}N correlation maps that connect histidine $^{15}N_{\varepsilon 2}$ and $^{15}N_{\delta 1}$ chemical shifts with nonexchangeable ring protons. Analysis of such spectra, using canonical ^{15}N chemical shifts for the tautomeric and cationic states of the side chain, provides an estimate of the relative populations of each conformer, with pK_a values fit from the pH dependence of the shifts (7, 15, 16).

To date, the great majority of NMR studies of protein electrostatics have focused on the native state that is highly populated under physiological conditions and often the only conformer that can be observed in standard experiments. In addition, a number of NMR investigations have been reported on denatured proteins, showing distinct differences in pK_a values from those that would be predicted on the basis of model compounds (17). Tollinger et al. have exploited the fact that the N-terminal SH3 domain from the *drk* protein exists in an approximate 1:1 equilibrium between native and unfolded conformations, at neutral pH, and in the absence of any denaturants, to measure pK_a values directly for both states from a single set of experiments (18). Notably, pK_as for ionizable side-chain groups of the unfolded state were found to vary in a sequence-dependent manner and again were not consistent with predictions based on model systems.

Although studies of electrostatics in the native and unfolded states of proteins provide a first step toward understanding aspects of function, they are inherently incomplete. The energy landscape of a protein molecule is rugged, with additional states populated at positions other than the minima corresponding to native and unfolded conformations (19–21). These additional, unexplored conformers can have different structures and functions relative to the native state and in some cases the overall function of the protein is derived from these higher energy states, despite being sparsely populated and transiently formed (22–24). A complete description of the role of electrostatics in function, therefore, must consider the ionization properties of side-chain

Significance

Electrostatic interactions in proteins play significant roles in conferring stability and in dictating function. Histidine residues are particularly important because their side chains can serve as both acids and bases over the physiological pH range and as both hydrogen bond donors and acceptors. Solution NMR spectroscopy is a powerful method for studying these residues in highly populated ground-state conformers. Here we develop a strategy for extending such studies to sparsely populated, short-lived protein states that can also play a significant role in defining protein function. An application to an invisible folding intermediate of the Im7 protein is provided, where site-specific pK_a values of histidine residues have been determined that explain the strong pH-dependent stability differences between native and intermediate states.

Author contributions: A.L.H. and L.E.K. designed research; A.L.H. and L.E.K. performed research; A.L.H. contributed new reagents/analytic tools; A.L.H. analyzed data; and A.L.H. and L.E.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: kay@pound.med.utoronto.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400577111/-/DCSupplemental.

moieties in these rare conformers as well. One example is α-lytic protease where the dynamics of the catalytic residue H57 have been shown to be critical for turnover (25). Quantification of the pK_a value(s) of the imidazole side chain of this residue in its different positions during catalysis is thus of considerable interest.

In addition to the fact that rare states often play important functional roles, sparsely populated conformers can also be important for directing protein-folding pathways. It is increasingly clear that folding, even of single-domain proteins, often proceeds through the establishment of low-populated and transiently formed intermediates (19, 26–30). A complete description of the mechanism of protein folding, therefore, is predicated on a characterization of these intermediate states, including the role that electrostatic interactions play in providing stability to them. As described below in detail, we consider here a protein-folding reaction involving an on-pathway intermediate whose relative stability with respect to the native conformer is dictated by the ionization properties of a pair of highly conserved histidine residues.

Solution-based NMR spectroscopy has recently emerged as a powerful technique for characterizing such conformationally rare states. The underlying physics was described a half century ago, leading to the development of Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion (31–33) and saturation transfer techniques (34, 35). In the past 15 y a wide range of different experiments that are based on the seminal ideas proposed then have been developed for studies of suitably labeled biomolecules. It is now possible to measure ^{1}H, ^{13}C, and ^{15}N backbone chemical shifts of "invisible" protein conformers (excited states) so long as they exchange with the NMR visible state (ground conformation) on a millisecond time-scale and where the fractional population of the rare conformer is at least 0.5% (36–38). Such chemical shifts can, in turn, be used as input into computational database programs to generate atomic resolution models of excited conformers, providing a first glimpse of the different structures that populate the energy landscape for the protein in question (39–43). Recently, the methodology has been extended to the measurement of side-chain ^{13}C chemical shifts in excited states of proteins, with the development of chemical exchange saturation transfer (CEST) experiments (44).

Herein we present an experimental approach based on CEST for quantifying the ionization/tautomeric states and pK_a values of histidine side chains in conformationally excited protein states. Although these states do not give rise to peaks in NMR spectra, the chemical shifts of the histidine C_γ, $C_{\delta 2}$, and $C_{\varepsilon 1}$ carbons can nevertheless be obtained accurately by amplification of their signals through measurements involving the corresponding nuclei in the ground state. The methodology is used to study the folding reaction of the 87-residue four-helix bundle Im7 protein, which proceeds through the formation of an on-pathway intermediate (45, 46). The kinetics and thermodynamics of folding have been studied extensively with tryptophan fluorescence, equilibrium denaturation and φ value analysis (29, 47–49) and it has been established that the stability of the intermediate relative to the native state increases with decreasing pH over a range where histidine side chains typically titrate (50). However, the inability to measure site-specific electrostatic interactions has limited an atomic-level understanding of the role(s) of the histidine residues in the folding process. We show that H40 and H47 have pK_a values of 6.9 ± 0.3 in the invisible folding intermediate, with an ε/δ tautomer distribution of 3:1 for H47, whereas the pK_a values of both histidines in the ground state are below 5. The pH-dependent stability difference of the intermediate and native states over the pH range 5.5–6.5, where our measurements were obtained, could be explained quantitatively in terms of the histidine pK_a values. An average pK_a value for H40 and H47 in the folding transition state (TS) was calculated from the pH-dependent folding and unfolding rates that is similar to the histidine pK_as of the folding intermediate (6.3 ± 0.1 vs. 6.9 ± 0.3). This suggests that histidine deprotonation has little effect on the rate of Im7 folding from the intermediate state.

Results

A Strategy for Studies of Histidine Side Chains in Excited Protein States. The imidazole ring of histidine can exist in three conformations that almost always exchange rapidly in aqueous buffer solutions (51). These include a pair of neutral tautomers, denoted as δ ($N_{\delta 1}$-H) and ε ($N_{\varepsilon 2}$-H) that are able to act as either hydrogen bond donors or acceptors and a protonated, cationic state (denoted as + in Fig. 1, *Upper*). The cationic form is of course dominant at low pH, whereas the ε form is favored ∼4:1 over the δ form at high pH, according to observations of the free amino acid in solution (52, 53). These conformations are usually indistinguishable in crystal structures, but can be readily identified by heteronuclear multiple-bond correlation (HMBC)-type NMR experiments which measure the imidazole ring ^{15}N chemical shifts (7). Studies of the free amino acid, model compounds, and proteins have led to a canonical set of ^{15}N imidazole chemical shifts whereby the protonated and deprotonated nitrogen in either the δ or ε form resonates at 167.5 or

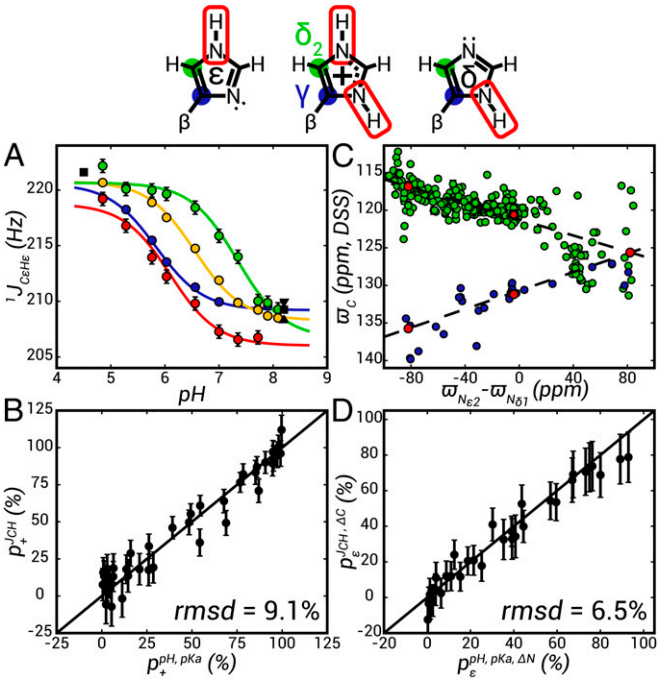

Fig. 1. Validation of an approach for the measurement of histidine pK_a values and tautomer populations in invisible protein conformers. (*A*) pH dependence of $^1J_{C\varepsilon H\varepsilon}$ couplings for H6 (blue), H13 (red), H23 (yellow), and H87 (green) of the PLCCγ SH2 domain (circles) fit to Eq. **1** (solid lines). In black are the couplings measured for model compounds; the squares represent L-histidine (59), the up triangle represents 1-methyl L-histidine, and the down triangle represents 3-methyl L-histidine. (*B*) Linear correlation plot of p_+ calculated from either $^1J_{C\varepsilon H\varepsilon}$ (*Y* axis) or pK_a values obtained from fits of ^{15}N and ^{1}H chemical shift titration curves (*X* axis) of the histidine residues of the PLCCγ SH2 domain, as described in the text. (*C*) Correlation of histidine $^{13}C_\gamma$ (blue) and $^{13}C_{\delta 2}$ (green) chemical shifts [referenced to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS)] from the BMRB database, with the difference in $^{15}N_{\varepsilon 2}$ and $^{15}N_{\delta 1}$ chemical shifts, as described in the text. Using canonical ^{15}N chemical shifts (7), $\Delta^N = 82$ ppm, −82 ppm, and −4 ppm for the δ, ε, and + conformers (structures above plots; *Upper*) the corresponding ^{13}C shifts for each state (δ, ε, +) can be determined assuming a linear correlation between ^{13}C shifts and Δ^N (Table S1). (*D*) Linear correlation plot of ε tautomer populations, p_ε, determined from $\Delta^{C,meas}$ (*Y* axis) or $\Delta^{N,meas}$(*X* axis) values of histidines in the PLCCγ SH2 domain, as described in the text. In *C* and *D* the line $y = x$ is drawn in black. Error bars in *A*, *B*, and *D* correspond to ±1 SD in $^1J_{C\varepsilon H\varepsilon}$, $p_+^{J_{CH}}$, and $p_\varepsilon^{J_{CH},\Delta C}$, respectively.

249.5 ppm (7), respectively, whereas in the cationic form the resonance frequencies of the $^{15}N_{\varepsilon 2}$ and $^{15}N_{\delta 1}$ nuclei are at ~174 and 178 ppm, respectively (54, 55). These reference shifts provide a basis for interpreting measured $^{15}N_{\varepsilon 2}$ and $^{15}N_{\delta 1}$ resonances in terms of fractional populations of δ, ε, and + (p_δ, p_ε, p_+), and the HMBC experiment has therefore become the method of choice for studies of histidine side chains in proteins. Unfortunately, the HMBC is relatively insensitive due to the long magnetization transfer delays that are required to connect the nitrogens and nonexchangeable protons of the imidazole ring so that the corresponding experiment for measuring chemical shifts in an excited state is not feasible.

A promising alternative for studies of rare conformers is to monitor the chemical shifts of the $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ imidazole carbons. Theoretical and experimental studies of model compounds show that the $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ nuclei, but not $^{13}C_{\varepsilon 1}$, are sensitive to the histidine conformational state, with shift differences as large as 12 and 8 ppm, respectively (7, 52, 56, 57). Although the $^{13}C_{\varepsilon 1}$ chemical shift is insensitive to the imidazole conformational equilibrium, the one-bond $^{13}C_{\varepsilon 1}$-$^{1}H_{\varepsilon 1}$ scalar coupling ($^1J_{C\varepsilon H\varepsilon}$) shows a relatively large variation of ~12 Hz between neutral and cationic states with little or no difference between neutral tautomers (58, 59) (Fig. 1*A*, black symbols). In principle, the combination of $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ shifts and $^1J_{C\varepsilon H\varepsilon}$ allows the extraction of pK_as as well as side-chain tautomer and ionization states even from measurements at a single pH value. This is particularly important because applications to excited states involve experiments that are less sensitive than for ground conformers, restricting the number of measurements possible. Moreover, the pH-dependent stability of the excited state may be such that experiments can only be recorded over a narrow pH range.

To establish the feasibility of the approach described above, we first considered an application to the C-terminal SH2 domain of phospholipase C-γ_1 (PLCCγ), focusing on the native state of the protein. This protein contains five histidines, four of which titrate with pK_a values ranging between 5.5 and 7 and all five side chains have p_ε values between 50% and 100% at high pH (55). Fig. 1 shows $^1J_{C\varepsilon H\varepsilon}$ vs. pH profiles for H6, H13, H23, and H87 (blue, red, yellow, and green, respectively in Fig. 1*A*), fit to a standard titration model involving a single protonation event (solid lines in Fig. 1*A*),

$$Q_{obs} = \frac{Q_{low} - Q_{high}}{1 + 10^{pH - pKa}} + Q_{high} \quad [1]$$

where Q is the measured parameter (in this case $^1J_{C\varepsilon H\varepsilon}$) and Q_{high} and Q_{low} are the individually fit limiting values at high and low pH, respectively. The pK_a values so obtained agree well with a previous study that used the ^{15}N HMBC approach (55) and virtually identical pK_as and extrema were obtained when the scalar coupling data were supplemented by measured ^{13}C, ^{15}N, and ^{1}H shifts for each histidine. Although there is some variation in the exact limiting values on a per-residue basis, the cationic and neutral extrema converge to 220.6 ± 1.0 Hz and 207.5 ± 1.5 Hz, respectively (Fig. 1*A*).

Notably, $^1J_{C\varepsilon H\varepsilon}$ = 202.5 ± 0.5 Hz was measured for the nontitrating histidine (H57) over pHs extending from 4 to 8, well outside the proposed range of scalar coupling values discussed above. This value indicates that the imidazole is neutral and indeed both carbon and nitrogen chemical shifts confirm that the side chain is almost entirely in the ε tautomer form. H57 $N_{\delta 1}$ and $N_{\varepsilon 2}$ hydrogen bond with the guanidinium group of R37 and the side-chain carboxyl group of E22, respectively, in structures of the SH2 domain in complex with target peptides and these interactions are expected to persist in solution in the absence of a binding partner (55, 60). It is likely that the close proximity to the cationic arginine lowers the $^1J_{C\varepsilon H\varepsilon}$ value below that typically expected. A low value for $^1J_{C\varepsilon H\varepsilon}$ (208 Hz) has also been observed for H57 of the catalytic triad of α-lytic protease that hydrogen bonds with S195, S214, and D102 in the neutral state (61). Additionally, nontitrating (pK_a < 3) histidines in subtilisin BPN′ have coupling values ranging from 205 to 209 Hz in either the δ or ε form (62). All of these residues are buried in polar or neutral environments, but lack a nearby R or K side chain.

Using the *J*-coupling endpoints of 220.6 ± 1.0 Hz and 207.5 ± 1.5 Hz (see above), p_+ can be calculated easily from a single $^1J_{C\varepsilon H\varepsilon}$ measurement [$p_+ = (Q_{obs} - Q_{high})/(Q_{low} - Q_{high})$]. Values of p_+ were obtained for the four titratable residues in the SH2 domain at each pH in this manner ($p_+^{J_{CH}}$) and compared with corresponding values calculated from pK_as extracted from fits of nitrogen and proton titration curves (p_+^{pH,pK_a}) (Fig. 1*B*). Errors in p_+ from 7% to 13% are obtained from propagation of uncertainties in the measurement values and in Q_{high}, Q_{low}, that are in good agreement with the rmsd between $p_+^{J_{CH}}$ and p_+^{pH,pK_a} (9.1%). Notably, for pH values where $^1J_{C\varepsilon H\varepsilon}$ varies between 211 and 218 Hz the uncertainties in Q_{high} and Q_{low} translate into errors in derived pK_a values of no more than 0.2 pH units. It is worth emphasizing that the validity of this approach is predicated on the assumption that $^1J_{C\varepsilon H\varepsilon}$ is insensitive to the distribution of tautomeric states (δ, ε), as suggested in previous studies (58, 59). We have further verified this by measuring $^1J_{C\varepsilon H\varepsilon}$ for L-histidine (Fig. 1*A*, black square), 1-methyl L-histidine (Fig. 1*A*, up triangle), and 3-methyl L-histidine (Fig. 1*A*, down triangle), pH 8.2, the latter two compounds serving as mimics of the ε and δ tautomeric states (52, 58), respectively, with little difference obtained in the coupling values.

Having established that measurement of $^1J_{C\varepsilon H\varepsilon}$ provides a robust estimate of p_+ and hence also of the histidine pK_a value (via Eq. **1**), we next sought an approach for obtaining p_δ and p_ε. Very early work on model compounds clearly demonstrated the sensitivity of ^{13}C NMR to the two neutral imidazole forms (52). More recently, Oldfield and coworkers used a combination of solid-state NMR, crystallography, and quantum calculations to analyze histidine ring ^{13}C chemical shifts of eight histidine-containing dipeptides (56). They found that $^{13}C_{\varepsilon 1}$ chemical shift perturbations are due primarily to changes in hydrogen bond length and intermolecular interactions. On the other hand, $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ shifts show a strong dependence on the tautomeric state. Further, these two shifts are highly anticorrelated, suggesting that their difference could potentially remove local effects arising from interactions with neighboring residues, and therefore provide an excellent probe of the tautomeric state of the imidazole ring. Although a number of studies, both experimental and computational, have provided estimates of ^{13}C shifts in each of the δ, ε, and + states (7, 52, 56, 57, 63), the majority have been performed on small peptides or amino acids, and we were interested in obtaining consensus values that would be appropriate in the context of a protein environment. Fig. 1 plots $^{13}C_{\delta 2}$ and $^{13}C_\gamma$ (green and blue, respectively in Fig. 1*C*) chemical shifts from the Biological Magnetic Resonance Data Bank (BMRB) (ϖ_C, *Y* axis in Fig. 1*C*) vs. the difference in $N_{\varepsilon 2}$ and $N_{\delta 1}$ chemical shifts ($\Delta^N = \varpi_{N\varepsilon 2} - \varpi_{N\delta 1}$, *X* axis in Fig. 1*C*). The canonical ^{15}N chemical shifts discussed above, Δ^N = 82 ppm, −82 ppm, and −4 ppm for the δ, ε, and + conformers were used to evaluate the corresponding ^{13}C shifts for each state (δ, ε, +) assuming a linear correlation between ^{13}C shifts and Δ^N in Fig. 1*C*. The values so obtained along with their uncertainties based on the rmsd between the BMRB data points and the best-fit lines correspond to $\Delta^C = \varpi_{C\gamma} - \varpi_{C\delta 2}$ values of 0.1 ± 3.0, 10.6 ± 1.5, and 18.8 ± 3.0 for δ, +, and ε states, respectively, as summarized in Table S1.

We have used titration data from the PLCCγ SH2 domain to cross-validate the limiting values of ^{13}C chemical shifts by comparing extracted p_ε values from ^{13}C and ^{15}N shift data. We prefer to obtain p_δ and p_ε from the difference in $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ chemical shifts ($\Delta^{C,meas}$) because this eliminates, at least partially, the effects

beyond tautomerization that can influence these shifts. Thus, at each point of the titration curve, $\Delta^{C,meas}$ vs. pH (Fig. S1), $p_\varepsilon = 1 - p_\delta - p_+$ is extracted from $\Delta^{C,meas} = p_\delta\Delta^C_\delta + p_\varepsilon\Delta^C_\varepsilon + p_+\Delta^C_+$, where Δ^C_i is the difference in $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ shifts for state $i \in \{\delta, \varepsilon, +\}$, and where the value of p_+ is obtained directly from $^1J_{C\varepsilon H\varepsilon}$. In a similar manner, values of p_δ and p_ε are calculated from $\Delta^{N,meas} = p_\delta\Delta^N_\delta + p_\varepsilon\Delta^N_\varepsilon + p_+\Delta^N_+$ at each pH value and for each histidine in the SH2 domain using the values of Δ^N_i as listed above. In this case p_+ is obtained directly from the pK_a value. A linear correlation plot of p_ε from $\Delta^{C,meas}$ (Y axis) or $\Delta^{N,meas}$ (X axis) is shown in Fig. 1*D*, with an rmsd between the two approaches of 6.5%. Approximate twofold increases in rmsd values are obtained if $^{13}C_\gamma$ or $^{13}C_{\delta 2}$ shifts are used independently to calculate p_ε, likely reflecting the fact that the individual ^{13}C chemical shifts are sensitive to factors outside of the distribution of ionization/tautomeric states. On this note, Oldfield and coworkers found in their studies of histidine dipeptides that density functional theory calculations were only able to reproduce experimental ^{13}C shifts when hydrogen bonding was taken into account (56).

NMR Experiments for Measurement of Histidine pK_a and Tautomers in the Excited State. Having shown above that accurate pK_a and p_i ($i \in \{\delta, \varepsilon, +\}$) can be obtained from fits of $^1J_{C\varepsilon H\varepsilon}$ and Δ^C, we subsequently developed experiments for measuring these parameters in protein-excited states. The set of proposed experiments (see below) requires ^{13}C at all carbon positions of the imidazole ring that is most easily achieved by producing proteins that are uniformly ^{13}C-labeled. This has the advantage in that labeling is relatively inexpensive, however the ^{13}C-^{13}C couplings that are generated through such a scheme (for example $^{13}C_\gamma$-$^{13}C_{\delta 2}$ for histidine) pose certain restrictions on the type of experiments that can be performed. In particular, the most common NMR method for studying protein-excited states uses CPMG elements in which trains of ^{13}C chemical shift refocusing pulses are applied that modulate the effects of exchange. However, these pulse trains will also lead to magnetization transfer between coupled ^{13}C spins, thus severely complicating the extraction of accurate exchange parameters (64, 65). In contrast, it has recently been shown that such effects do not compromise the parameters obtained from the analysis of CEST datasets (44). In this class of experiment a set of spectra is recorded with the position of a very weak radiofrequency (rf) field (typically 20–50 Hz) varied from one spectrum to the next (66, 67). The intensity of each correlation derived from the major state is quantified as a function of the position of the weak field. When the field is coincident with or close to the resonance frequency of the ^{13}C in the ground state, the intensity of the peak from this carbon decreases due to saturation. Additionally, when the weak field is applied at the position of the corresponding peak from the excited state, the ground-state peak intensity also decreases because the effect of the perturbation is transferred via the exchange process. A plot of the intensity of the ground-state (observed) peak as a function of the position of the rf field produces a pair of dips, the larger of the two at the resonance frequency of the carbon in the ground state with the second at the frequency of the corresponding ^{13}C in the excited state (see *An Application to an Invisible Im7-Folding Intermediate*). It is also possible to design an experiment (Fig. S2*A*) whereby each of the dips can be further split into a pair of peaks that are separated by $^1J_{CH}$, as illustrated in Fig. 2, from which both $^{13}C_{\varepsilon 1}$ shifts and $^1J_{C\varepsilon H\varepsilon}$ couplings are obtained.

An Application to an Invisible Im7-Folding Intermediate. The Im7 protein folds via an on-pathway intermediate to the native conformation, a process that has been studied in detail previously using a variety of biophysical methods (29, 45–49), including NMR spectroscopy (48, 68–70). CPMG relaxation dispersion and CEST measurements performed in our laboratory at 10 °C, pH 6.6

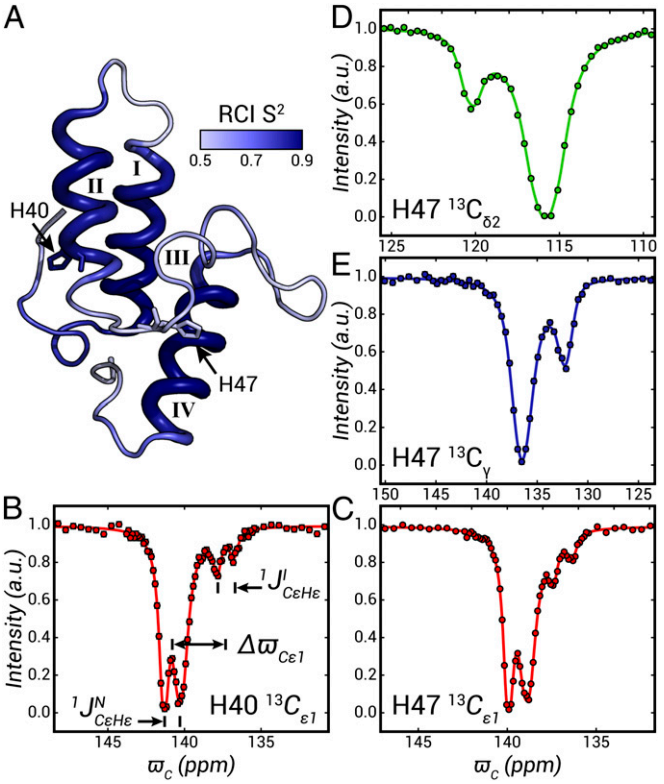

Fig. 2. ^{13}C CEST profiles for H40 and H47 of Im7 (pH 6.6, 10 °C). (*A*) Ribbon diagram of the structure of the native state of Im7 (Protein Data Bank ID code 1CEI) (78). The structure is color coded in shades of blue according to random coil index (RCI) predicted order parameters squared (S^2) (80) of the intermediate state, with the size of the tubes reflecting the extent of α-helicity (79) in *I*. (*B*–*E*) Representative ^{13}C CEST profiles (circles). Major and minor dips in all cases derive from native and intermediate Im7 states, respectively. $^{13}C_\varepsilon$ chemical shifts and $^1J_{C\varepsilon H\varepsilon}$ couplings are extracted from fits of traces (solid lines) in *B* and *C*, whereas $^{13}C_{\delta 2}$ and $^{13}C_\gamma$ shifts are obtained from the analysis of profiles in *D* and *E*. Experimental details and a description of data fitting are provided in *SI Materials and Methods*.

can be well analyzed assuming a two-state exchange mechanism, $I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N$, where I (excited) and N (ground) denote intermediate and native states, respectively (69–71). Under these conditions $k_{ex} = k_{IN} + k_{NI} = 320\ s^{-1}$, the fractional population of I, p_I = 1.5%, and the intermediate state is not directly observed in NMR spectra. Fig. 2*A* shows a ribbon diagram of Im7 (*Discussion*) highlighting the side chains of the two histidines in the protein, H40 and H47, that form the basis of the present study. These histidines are highly conserved in nearly all Im proteins, with over 60% and 90% conservation for H40 and H47, respectively, across 250 sequences identified in a BLAST search query (72). In contrast, other histidine residues in Im proteins have less than 5% conservation. Fig. 2 *B*–*E* shows CEST profiles for these residues (10 °C, pH 6.6) that have been analyzed to extract exchange parameters ($k_{ex}p_I$) and chemical shifts of excited-state nuclei, as described in detail previously, and $^1J_{C\varepsilon H\varepsilon}$ values (*SI Materials and Methods*). The profiles in Fig. 2 show major (native state) and minor (intermediate) dips for $C_{\varepsilon 1}$ of H40 (Fig. 2*B*) and H47 (Fig. 2*C*) that are each split by $^1J_{C\varepsilon H\varepsilon}$ Hz. Values of $^1J_{C\varepsilon H\varepsilon}$ = 217 and 215 Hz are obtained for H40 and H47 of the intermediate state (pH 6.6), respectively, significantly higher than 210 and 208 Hz that are measured for the corresponding side chains in the native conformation.

Histidine $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ chemical shifts were also measured for the invisible Im7-folding intermediate. These experiments (Fig. S2 and Table S2) transfer polarization from $^1H_{\delta 2}$ to $^{13}C_{\delta 2}$, to record $^{13}C_{\delta 2}$ chemical shifts, or further to $^{13}C_\gamma$ to measure the shift of the quaternary aromatic carbon. A prerequisite, therefore, is that high sensitivity ^1H-^{13}C spectra of the ground state be obtained. Unfortunately, the cross-peak connecting $^1H_{\delta 2}$-$^{13}C_{\delta 2}$ in the ground state of H40 was very weak due to exchange broadening and excited-state $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ chemical shifts could not be measured for this residue. In contrast, a strong $^1H_{\delta 2}$-$^{13}C_{\delta 2}$ cross-peak for H47 was observed in heteronuclear single-quantum coherence (HSQC) spectra and high quality CEST profiles were obtained for assignment of *I* state $^{13}C_\gamma$ and $^{13}C_{\delta 2}$ carbons of H47 (Fig. 2 *D* and *E*). A discussion of the signal-to-noise requirements for accurate measurements is provided in *SI Materials and Methods*.

CEST datasets were recorded at pH 6.25 and 6.6 (10 °C) to obtain $^1J_{C\varepsilon H\varepsilon}$ values (H40 and H47) and $^{13}C_\gamma$,$^{13}C_{\delta 2}$ chemical shifts (H47) of the Im7-folding intermediate. Fig. 3 *A* and *B* plots the measured $^1J_{C\varepsilon H\varepsilon}$ values for *I*, from which pK_as can be extracted using 220.6 ± 1.0 Hz and 207.5 ± 1.5 Hz as the endpoints for $^1J_{C\varepsilon H\varepsilon}$, corresponding to the imidazole side chain in the cationic and neutral forms, respectively, as described above. Calculated $^1J_{C\varepsilon H\varepsilon}$ vs. pH profiles are shown in Fig. 3 based on $pK_a = 6.9 \pm 0.3$ and 6.9 ± 0.2 that are obtained for H40 and H47 (solid lines in Fig. 3 *A* and *B*, *Upper*) along with the 68% (solid red in Fig. 3 *A* and *B*, *Upper*) and 95% (light red in Fig. 3 *A* and *B*, *Upper*) confidence limits that take into account uncertainties in the endpoints. In stark contrast to the elevated scalar coupling values obtained for *I*, the corresponding values for both H40 and H47 in the *N* state are much lower and do not vary between pH 6 and 9. This establishes that the pK_a values for these histidines in the ground state are less than 5. Combining the $^1J_{C\varepsilon H\varepsilon}$ and $^{13}C_\gamma$, $^{13}C_{\delta 2}$ chemical shift data as described above, the tautomer/ionization scheme shown in Fig. 3*C* (*Right*) is obtained for H47 of the *I* state (pH 6.6, 10 °C); only $p_\delta + p_\varepsilon$ can be obtained for H40 because $^{13}C_\gamma$, $^{13}C_{\delta 2}$ shifts cannot be measured for this residue in the *I* state. The tautomer distributions for H40 and H47 (pH 6.6) in the native state are shown in Fig. 3*C* (*Left*) based on measurements of $^{13}C_{\delta 2}$ and $^{13}C_\gamma$ chemical shifts.

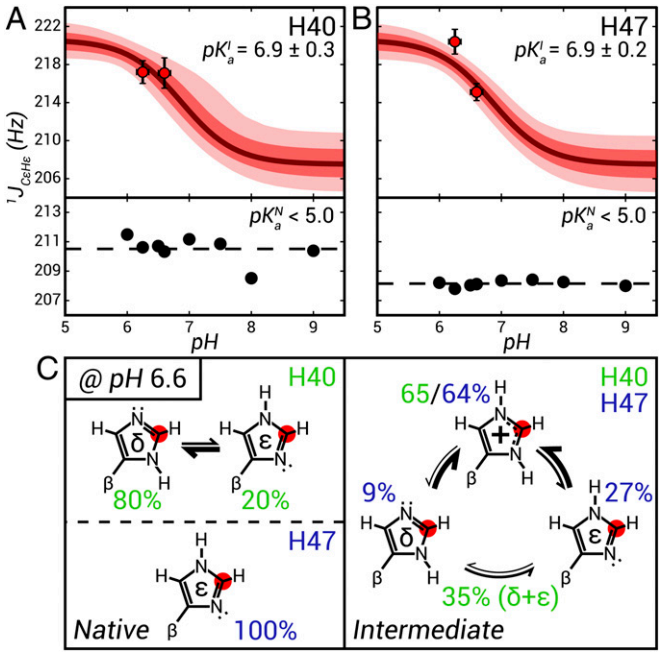

Fig. 3. Determination of Im7 histidine pK_a values and imidazole ring tautomer populations in *N* and *I* states. $^1J_{C\varepsilon H\varepsilon}$ couplings (pH 6.25, 6.6) obtained from fits of ^{13}C CEST profiles for H40 (circles in *A*, *Upper*) and H47 (circles in *B*, *Upper*), along with the $^1J_{C\varepsilon H\varepsilon}$ vs. pH profiles calculated from the fitted pK_a values (solid lines). The 68% and 95% confidence limits of the fits that take into account uncertainties in the endpoints of $^1J_{C\varepsilon H\varepsilon}$ (at high and low pH) are in dark and light red, respectively (*SI Materials and Methods*). Shown in *A* and *B* (*Lower*) are $^1J_{C\varepsilon H\varepsilon}$ couplings for the Im7 native state measured using 2D in-phase/antiphase ^1H-^{13}C HSQC datasets (84), with the average coupling value denoted by the dashed line. (*C*) Schematic representation of the imidazole ring tautomer/ionization distributions of H40 (green) and H47 (blue) in *I* (*Right*) and *N* (*Left*) (pH 6.6, 10 °C). Only $p_\varepsilon + p_\delta = 1 - p_+$ is reported for H40 of *I* because $^{13}C_\gamma$, $^{13}C_{\delta 2}$ chemical shift data of the intermediate could not be obtained for this residue. Errors in $^1J_{C\varepsilon H\varepsilon}$ are based on fits of the 1D CEST profiles. Errors in p_i values are on the order of 10% to 15%.

Discussion

Studies of Histidine Electrostatics in Invisible Protein States. It is becoming increasingly well established that a protein's function depends not only on its low-energy ground-state conformation, but often also on higher energy states that may be sparsely populated and transiently formed (73, 74). Moreover, a large amount of evidence from a variety of different techniques has established that proteins can fold via formation of a series of intermediates that may also be short-lived and low-populated (19, 26–30). Further, excursions from native states to higher energy conformers have been shown to be linked with protein aggregation that may be an important first step in the formation of oligomers that are associated with disease (75). Characterization of these rare conformers, that are involved in a range of different important biological processes, is thus critical. In the past decade the development of a suite of NMR experiments has facilitated structural studies and in some cases atomic level descriptions of these conformationally excited states (26, 27, 76). Herein we extend the methodology to studies of protein electrostatics, focusing on histidine imidazole side chains that can play important functional roles in catalysis, electron transfer, protein stability, and metal binding, as well as in controlling the relative stabilities of transient states that are populated during the protein-folding process. The methodology thus developed has a wide range of applications; we provide one example that focuses on relative pH-dependent stabilities of a native conformer and an on-pathway folding intermediate that can be explained in terms of the pK_a values of a pair of highly conserved histidines.

In the study here, we have first presented a carbon-based approach for extracting reliable estimates of histidine pK_a values as well as protonation and tautomeric states that has been validated against the more common nitrogen-based method (7) using experimental data recorded for the histidine side chains of the ground state of the PLCγ SH2 domain. Subsequently, a set of CEST-based experiments has been developed for measurement of the $^1J_{C\varepsilon H\varepsilon}$ coupling and $^{13}C_\gamma$,$^{13}C_{\delta 2}$ chemical shift values using samples that are uniformly ^{13}C-labeled and hence relatively inexpensive to produce. The methodology circumvents a major problem in applications of the more popular CPMG experiments because magnetization transfer via ^{13}C-^{13}C scalar couplings that occurs during the CPMG pulse train is not an issue with CEST, opening up the possibility for detailed studies of the role of side chains in protein function and stability. The methodology is limited, however, to relatively slow exchange processes ($<\sim 500\ s^{-1}$) because as the exchange rate increases, the line widths of the excited-state peaks (the dips in CEST profiles) grow (44), and the broadening eventually obscures the doublet splitting from which $^1J_{C\varepsilon H\varepsilon}$ values are measured, for example.

We have subsequently applied the methodology to study an invisible on-pathway folding intermediate of Im7, whose steep pH dependence of stability has been suggested to derive from the pair of histidine residues in the protein (50). Unfortunately, only a narrow pH range, between 5.5 and 7.0, can be pursued because at higher pHs the intermediate state is too weakly populated to

generate excited-state dips in CEST profiles, whereas below this pH range the spectral quality and sensitivity degrades as the population of *I* becomes more dominant. We have shown that accurate histidine pK_a values and tautomer populations can be obtained from only a single pH measurement so long as $^1J_{C\varepsilon H\varepsilon}$ and $^{13}C_\gamma$,$^{13}C_{\delta 2}$ chemical shift values are not at extreme ends of the expected distributions. For Im7, pK_a values of 6.9 were obtained for both H40 and H47 in state *I* from fits of data recorded at pH 6.25 and 6.6, very similar to the pK_a of the histidine amino acid in solution (~6.6), suggesting that these residues are in exposed environments in the exited state. This is further confirmed for H47 based on the relative population ratio of ε and δ tautomers (3:1) that is similar to the 4:1 ratio that has been measured for the free amino acid (52, 53). Notably, neither H40 nor H47 exhibit the 4:1 ratio in the native conformation of the protein, with H40 predominantly in the rare δ tautomer form and H47 exclusively in the ε state, suggesting that both histidines are in hydrogen-bonded environments. Indeed, crystal structures of Im7 show that H47 stacks on W75 and is hydrogen bonded to the backbone amide of D49, whereas H40 becomes buried by the N terminus in the native state, potentially forming a hydrogen bond with the backbone carbonyl of K4 (77, 78). Thus, significant rearrangements of the histidine side chains must accompany the *I*-to-*N* transition, leading to burial of the exposed H40 and H47 side chains. Global structural changes have also been noted, largely from analysis of backbone ^1H, ^{15}N, and ^{13}C CPMG relaxation dispersion profiles that provide chemical shifts of the excited state (68–70). These, in turn, can be recast in terms of residue-specific fractional helicity values (79) and order parameters reflecting amplitudes of motion (80), with low values corresponding to disorder. Fig. 2*A* encodes the fractional helicity and order parameters of the *I* state on the structure of *N*, with increasing tube thickness corresponding to higher helical character. It is clear that a region extending from the C-terminal residues of helix II through helix III is not formed in the *I* state, with H47 situated in the intervening loop, consistent with results from other measurements (47).

Relative pH-Dependent Stabilities of *I* and *N* Can Be Understood in Terms of Histidine pK_a Values. We were interested in ascertaining whether the steep pH dependence of the relative stabilities of *I* and *N* over the range 5.5–6.5 could be understood on the basis of the measured pK_a values for H40 and H47. The free-energy difference between states *A* and *B* ($\Delta G_{A\to B} = G_B - G_A$) can be separated into a pH-independent term that contains all contributions that are nonelectrostatic or that involve sites that are fully protonated in both states *A* and *B* ($\Delta G^0_{A\to B}$) and a pH-dependent term ($\Delta G^{pH}_{A\to B}$),

$$\Delta G_{A\to B} = \Delta G^0_{A\to B} + \Delta G^{pH}_{A\to B},$$
$$\Delta G^{pH}_{A\to B} = -RT\sum_{m=1}^{i}\ln\left(\frac{K_a^{B,m}+[H^+]}{K_a^{A,m}+[H^+]}\right) \quad [2]$$

where $K_a^{P,m}$ is the acid dissociation constant for group *m* in state $P \in \{A, B\}$ and RT is the product of the universal gas constant and the absolute temperature (81). Assuming that $\Delta G^{pH}_{A\to B}$ is determined exclusively by the titration of H40 and H47 in the pH range 5.5–7 and noting that pK_a values of H40 and H47 in the native state are less than 5, Eq. **2** can be simplified to

$$\Delta\Delta G_{N\to I} = -RT\sum_{m=1}^{2}\ln\left(\frac{K_a^{I,m}+[H^+]_{pH}}{K_a^{I,m}+[H^+]_{pH_{ref}}}\right) \quad [3]$$

where $\Delta\Delta G_{N\to I} = \Delta G_{N\to I}(pH) - \Delta G_{N\to I}(pH_{ref})$ and the summation is over the two histidines in the protein. The pH dependence of the relative stabilities of *N* and *I* can be calculated from Eq. **3** because the pK_a values are known for H40 and H47 in the intermediate state. Values of $\Delta\Delta G_{N\to I}$ can be obtained in an alternative manner, directly from the populations of each state using the relation $\Delta G_{N\to I} = -RT\ln(p_I/p_N)$, with the populations available from analysis of CEST data recorded as a function of pH. Fig. 4*A* shows CEST profiles for $^{13}C_{\varepsilon 1}$ recorded at a number of pH values that have been analyzed to extract p_I, p_N, with the pH dependence of p_I illustrated in Fig. 4*B*. The $\Delta\Delta G_{N\to I}$ calculated in this manner for the four pH values considered here are plotted in Fig. 4*C* (circles) and the corresponding $\Delta\Delta G_{N\to I}$ profile obtained from Eq. **3** shown in the dashed line along with the 68% and 95% confidence limits in dark and light green, respectively. The close agreement between values obtained using these different methods provides strong evidence that the pH-dependent stability can be explained on the basis of the differential pK_a values for the *I* and *N* states and further provides a cross-validation of the methodology that has been developed.

As a final step we have fitted the pH-dependent exchange rates, k_{IN} and k_{NI}, following a procedure described previously (82) (*SI Materials and Methods*) to extract pK_a values for the histidine residues in the rate-limiting TS (Fig. 4*D*). Attempts to fit the kinetic rates to unique TS pK_a values [pK_a(TS)] for each of H40 and H47 did not produce significantly improved goodness-of-fit scores

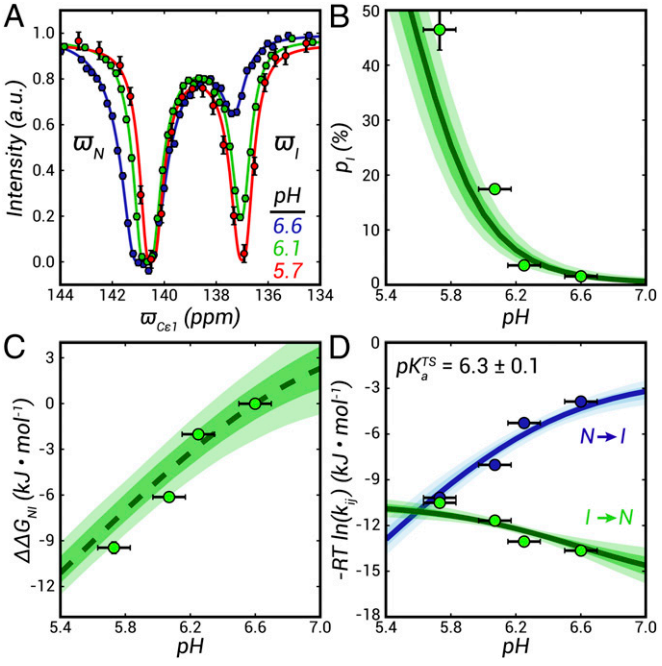

Fig. 4. pH-dependent relative stabilities of *N* and *I* (pH 6.5–5.5) can be explained in terms of H40, H47 pK_a values. (*A*) Representative, ^1H-decoupled $^{13}C_{\varepsilon 1}$ CEST profiles of H40 used to determine *I* state populations as a function of pH (10 °C). Best fits to the experimental data (circles) are shown in solid lines (see *SI Materials and Methods* for details). (*B*) p_I vs. pH profile (10 °C). Green circles are the measured p_I values based on the analysis of CEST profiles. The solid line corresponds to the predicted p_I vs. pH dependence obtained from a fit of the measured p_I values as a function of pH to Eq. **2**. (*C*) $\Delta\Delta G_{N\to I}$ values calculated from populations of states *I* and *N*, as described in the text (circles), as a function of pH. Shown with a dashed line is the $\Delta\Delta G_{N\to I}$ vs. pH profile, Eq. **3**, using the pK_a values of H40 and H47 in the *I* state. $\Delta G_{N\to I}$ at pH 6.6 is taken as the reference. (*D*) Fits (solid lines) of the pH dependence of k_{IN} and k_{NI} (circles), as described in *SI Materials and Methods*, to a model which assumes equivalent pK_as in the TS for H40, H47. Shaded regions in *B*–*D* represent the 68% (dark) and 95% (light) confidence ranges, as described in *SI Materials and Methods*. Vertical error bars correspond to ± 1 SD of the listed values; horizontal error bars are set to ± 0.1 pH unites. See *SI Materials and Methods, Error Analysis*.

relative to the case where only a single, average pK_a(TS) value was considered. Moreover, very large uncertainties in pK_a(TS) values were obtained from fits involving site-specific pK_as. We have therefore used a model where each of the pK_a values for H40 and H47 are assumed to be the same in the TS. This is likely to be a reasonable approximation because pK_a values of the two histidines are nearly identical in the *I* state and are both similarly shifted to low values (<5) in the native conformation (Table S3). It is important to emphasize, however, that our data do not allow us to prove that the pK_as are the same in the TS. An average value of pK_a(TS) = 6.3 ± 0.1 is fit that is similar to the pK_a values measured for the intermediate. This suggests that histidine deprotonation occurs after the rate-limiting TS for the *I*-to-*N* transition. A similar result has been observed for apomyoglobin where analysis of pH-dependent unfolding and refolding kinetics establishes that the pK_a of H24 in the molten globular intermediate and the TS are similar and significantly higher than in the native state (83). Notably, the profiles in Fig. 4 *C* and *D* are fully consistent with results from Radford and coworkers based on equilibrium denaturation and folding/unfolding kinetics of a hexahistidine version of Im7 (50) (Fig. S3), although these authors were not able to measure site-specific pK_a values or histidine tautomer populations in the *I* state.

Previous φ value analysis studies by Radford and coworkers have established that the nonnative interactions stabilizing the intermediate are disrupted in the final, rate-limiting TS, after which helix III and related native contacts form, locking the structure into its native conformation (47). Our work highlights the changes that occur at the level of H40 and H47 that are conserved across a range of Im proteins, showing that the imidazole side chains of these residues undergo an exposed-to-ordered transition that leads to stabilization of the resulting native state. More generally, our work describes an approach for studying histidine electrostatics in sparsely populated and transiently formed proteins states, opening an avenue for a more detailed description of the roles of these critical residues in protein structure and function than, to our knowledge, has been previously possible.

Materials and Methods

Proteins were expressed and purified as described in *SI Materials and Methods*. All NMR experiments were performed at 10 °C on 18.8 or 11.7 T Varian Inova spectrometers equipped with room-temperature pulsed-field gradient triple-resonance probes. Samples were 1–1.5 mM in protein concentration and dissolved in 50 mM potassium phosphate, pH 5.5–8.0, 0.02% azide, 95% H_2O/5% (vol/vol) D_2O (Im7) or 25 mM sodium arsenate, 25 mM sodium citrate, pH 4.0–8.0, 1 mM TCEP, 95% H_2O/5% (vol/vol) D_2O (PLCγ SH2). Details concerning CEST pulse schemes and the analysis of CEST data are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. The authors thank Dr. Sara Whittaker (University of Birmingham) and Profs. Geoff Moore (University of East Anglia) and Sheena Radford (University of Leeds) for the gift of Im7 plasmids and for valuable discussions. A.L.H. acknowledges the National Science Foundation (OISE-0852964) and a Canadian Institutes of Health Research (CIHR) training grant in protein folding and disease for postdoctoral support. This work was supported by a grant from the CIHR. L.E.K. holds a Canada Research Chair in Biochemistry.

1. Perutz MF (1978) Electrostatic effects in proteins. *Science* 201(4362):1187–1191.
2. Warshel A (1981) Electrostatic basis of structure-function correlation in proteins. *Acc Chem Res* 14(9):284–290.
3. Grimsley GR, Scholtz JM, Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* 18(1):247–251.
4. Pace CN, Grimsley GR, Scholtz JM (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* 284(20):13285–13289.
5. Hedstrom L (2002) Serine protease mechanism and specificity. *Chem Rev* 102(12): 4501–4524.
6. Shimba N, et al. (2003) Quantitative identification of the protonation state of histidines in vitro and in vivo. *Biochemistry* 42(30):9227–9234.
7. 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.
8. Puttick J, Baker EN, Delbaere LTJ (2008) Histidine phosphorylation in biological systems. *Biochim Biophys Acta* 1784(1):100–105.
9. Edgcomb SP, Murphy KP (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins* 49(1):1–6.
10. Markley JL (1975) Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy. *Acc Chem Res* 8(2):70–80.
11. Bachovchin WW (2001) Review: Contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites. *Magn Reson Chem* 39:S199–S213.
12. Mandel M (1964) The effect of temperature on the proton magnetic resonance spectra of ribonuclease, oxidized ribonuclease, and lysozyme. *Proc Natl Acad Sci USA* 52(3):736–741.
13. Mandel M (1965) Proton magnetic resonance spectra of some proteins: I. Ribonuclease, oxidized ribonuclease, lysozyme, and cytochrome C. *J Biol Chem* 240:1586–1592.
14. Meadows DH, Markley JL, Cohen JS, Jardetzky O (1967) Nuclear magnetic resonance studies of the structure and binding sites of enzymes. I. Histidine residues. *Proc Natl Acad Sci USA* 58(4):1307–1313.
15. Alei M, Morgan LO, Wageman WE, Whaley TW (1980) The pH dependence of nitrogen-15 NMR shifts and coupling constants in aqueous imidazole and 1-methylimidazole. Comments on estimation of tautomeric equilibrium constants for aqueous histidine. *J Am Chem Soc* 102(9):2881–2887.
16. Farr-Jones S, Wong WYL, Gutheil WG, Bachovchin WW (1993) Direct observation of the tautomeric forms of histidine in nitrogen-15 NMR spectra at low temperatures. Comments on intramolecular hydrogen bonding and on tautomeric equilibrium constants. *J Am Chem Soc* 115(15):6813–6819.
17. Tan YJ, Oliveberg M, Davis B, Fersht AR (1995) Perturbed pKA-values in the denatured states of proteins. *J Mol Biol* 254(5):980–992.
18. Tollinger M, Crowhurst KA, Kay LE, Forman-Kay JD (2003) Site-specific contributions to the pH dependence of protein stability. *Proc Natl Acad Sci USA* 100(8):4545–4550.
19. Brockwell DJ, Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* 17(1):30–37.
20. Roca M, Messer B, Hilvert D, Warshel A (2008) On the relationship between folding and chemical landscapes in enzyme catalysis. *Proc Natl Acad Sci USA* 105(37): 13877–13882.
21. Vendruscolo M, Dobson CM (2005) Towards complete descriptions of the free-energy landscapes of proteins. *Philos Trans A Math Phys Eng Sci* 363(1827):433–450, discussion 450–452.
22. Tsytlonok M, Itzhaki LS (2013) The how's and why's of protein folding intermediates. *Arch Biochem Biophys* 531(1-2):14–23.
23. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
24. Tokuriki N, Tawfik DS (2009) Protein dynamism and evolvability. *Science* 324(5924): 203–207.
25. Haddad KC, Sudmeier JL, Bachovchin DA, Bachovchin WW (2005) alpha-lytic protease can exist in two separately stable conformations with different His57 mobilities and catalytic activities. *Proc Natl Acad Sci USA* 102(4):1006–1011.
26. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein-folding intermediate at atomic resolution. *Science* 329(5997): 1312–1316.
27. Neudecker P, et al. (2012) Structure of an intermediate state in protein folding and aggregation. *Science* 336(6079):362–366.
28. Meinhold DW, Wright PE (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. *Proc Natl Acad Sci USA* 108(22):9078–9083.
29. Capaldi AP, Kleanthous C, Radford SE (2002) Im7 folding mechanism: Misfolding on a path to the native state. *Nat Struct Biol* 9(3):209–216.
30. Rumbley J, Hoang L, Mayne L, Englander SW (2001) An amino acid code for protein folding. *Proc Natl Acad Sci USA* 98(1):105–112.
31. Carr H, Purcell E (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* 94(3):630–638.
32. Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29(8):688.
33. Allerhand A, Gutowsky HS (1964) Spin–echo NMR studies of chemical exchange. I. Some general aspects. *J Chem Phys* 41(7):2115.
34. Forsén S, Hoffman RA (1963) Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. *J Chem Phys* 39(11):2892.
35. Ward KM, Aletras AH, Balaban RS (2000) A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J Magn Reson* 143(1):79–87.
36. Hansen DF, Vallurupalli P, Kay LE (2008) Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR* 41(3): 113–120.
37. Palmer AG, 3rd, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339(1997):204–238.
38. Ishima R, Torchia DA (2000) Protein dynamics from NMR. *Nat Struct Biol* 7(9):740–743.
39. Shen Y, et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105(12):4685–4690.
40. Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J Am Chem Soc* 131(39):13894–13895.

41. Shen Y, Bax A (2010) SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* 48(1):13–22.
42. Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: Significantly improved protein chemical shift prediction. *J Biomol NMR* 50(1):43–57.
43. Camilloni C, Robustelli P, De Simone A, Cavalli A, Vendruscolo M (2012) Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. *J Am Chem Soc* 134(9):3968–3971.
44. Bouvignies G, Vallurupalli P, Kay LE (2014) Visualizing side chains of invisible protein conformers by solution NMR. *J Mol Biol* 426(3):763–774.
45. Capaldi AP, Shastry MC, Kleanthous C, Roder H, Radford SE (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* 8(1): 68–72.
46. Papadakos G, Wojdyla JA, Kleanthous C (2012) Nuclease colicins and their immunity proteins. *Q Rev Biophys* 45(1):57–103.
47. Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J Mol Biol* 326(1):293–305.
48. Gorski SA, et al. (2004) Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J Mol Biol* 337(1): 183–193.
49. Bartlett AI, Radford SE (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *J Mol Biol* 396(5):1329–1345.
50. 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.
51. Hass MAS, Hansen DF, Christensen HEM, Led JJ, Kay LE (2008) Characterization of conformational exchange of a histidine side chain: Protonation, rotamerization, and tautomerization of His61 in plastocyanin from Anabaena variabilis. *J Am Chem Soc* 130(26):8460–8470.
52. 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.
53. Blomberg F, Maurer W, Rüterjans H (1977) Nuclear magnetic resonance investigation of 15N-labeled histidine in aqueous solution. *J Am Chem Soc* 99(25):8149–8159.
54. Shimahara H, et al. (2007) Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. *J Biol Chem* 282(13):9646–9656.
55. Singer AU, Forman-Kay JD (1997) pH titration studies of an SH2 domain-phosphopeptide complex: Unusual histidine and phosphate pKa values. *Protein Sci* 6(9): 1910–1919.
56. 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.
57. 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.
58. Hunkapiller MW, Smallcombe SH, Whitaker DR, Richards JH (1973) Carbon nuclear magnetic resonance studies of the histidine residue in α-lytic protease. Implications for the catalytic mechanism of serine proteases. *Biochemistry* 12(23):4732–4743.
59. Wasylishen RE, Tomlinson G (1975) pH-dependence of 13C chemical shifts and 13C,H coupling constants in imidazole and L-histidine. *Biochem J* 147(3):605–607.
60. Pascal SM, et al. (1994) Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-γ 1 complexed with a high affinity binding peptide. *Cell* 77(3):461–472.
61. Bachovchin WW, Kaiser R, Richards JH, Roberts JD (1981) Catalytic mechanism of serine proteases: Reexamination of the pH dependence of the histidyl 1J13C2-H coupling constant in the catalytic triad of alpha-lytic protease. *Proc Natl Acad Sci USA* 78(12):7323–7326.
62. Day RM, et al. (2003) Tautomerism, acid-base equilibria, and H-bonding of the six histidines in subtilisin BPN′ by NMR. *Protein Sci* 12(4):794–810.
63. Vila JA, Arnautova YA, Vorobjev Y, Scheraga HA (2011) Assessing the fractions of tautomeric forms of the imidazole ring of histidine in proteins as a function of pH. *Proc Natl Acad Sci USA* 108(14):5602–5607.
64. Ishima R, Torchia DA (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J Biomol NMR* 25(3):243–248.
65. Lundström P, Hansen DF, Kay LE (2008) Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: Comparison between uniformly and selectively (13)C labeled samples. *J Biomol NMR* 42(1):35–47.
66. 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.
67. 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.
68. Whittaker SB-M, Clayden NJ, Moore GR (2011) NMR characterisation of the relationship between frustration and the excited state of Im7. *J Mol Biol* 414(4):511–529.
69. 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.
70. Hansen AL, Lundström P, Velyvis A, Kay LE (2012) Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain 1H probes. *J Am Chem Soc* 134(6):3178–3189.
71. 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.
72. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
73. Sekhar A, Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA* 110(32): 12867–12874.
74. Karplus M, Kuriyan J (2005) Molecular dynamics and protein function. *Proc Natl Acad Sci USA* 102(19):6679–6685.
75. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5(1):15–22.
76. Bouvignies G, et al. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111–114.
77. Dennis CA, et al. (1998) A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J* 333(Pt 1):183–191.
78. Chak KF, Safo MK, Ku WY, Hsieh SY, Yuan HS (1996) The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface. *Proc Natl Acad Sci USA* 93(13):6437–6442.
79. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227–241.
80. Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127(43):14970–14971.
81. Wyman J, Jr. (1964) Linked functions and reciprocal effects in hemoglobin: A second look. *Adv Protein Chem* 19:223–286.
82. Tollinger M, Kay LE, Forman-Kay JD (2005) Measuring pK(a) values in protein folding transition state ensembles by NMR spectroscopy. *J Am Chem Soc* 127(25):8904–8905.
83. Jamin M, Geierstanger B, Baldwin RL (2001) The pKa of His-24 in the folding transition state of apomyoglobin. *Proc Natl Acad Sci USA* 98(11):6127–6131.
84. 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.