



Mapping the FF domain folding pathway via structures of transiently populated folding intermediates

Debajyoti De^a, Nemika Thapliyal^a, Ved Prakash Tiwari^a, Yuki Toyama^{b,c,d,e}, D. Flemming Hansen^{f,g}, Lewis E. Kay^{b,c,d,1}, and Pramodh Vallurupalli^{a,1}

Affiliations are included on p. 9.

Edited by Arthur Palmer, Columbia University, New York, NY; received August 22, 2024; accepted November 4, 2024

Despite the tremendous accomplishments of AlphaFold2/3 in predicting biomolecular structure, the protein folding problem remains unsolved in the sense that accurate atomistic models of how protein molecules fold into their native conformations from an unfolded ensemble are still elusive. Here, using chemical exchange saturation transfer (CEST) NMR experiments and a comprehensive four-state kinetic model of the folding trajectory of a 71 residue four-helix bundle FF domain from human HYPB/FBP11 we present an atomic resolution structure of a transiently formed intermediate, I2, that along with the structure of a second intermediate, I1, provides a description of the FF domain folding trajectory. By recording CEST profiles as a function of urea concentration the extent of compaction along the folding pathway is evaluated. Our data establish that unlike the partially disordered I1 state, the I2 intermediate that is also formed before the rate-limiting folding barrier is well ordered and compact like the native conformer, while retaining nonnative interactions similar to those found in I1. The slow-interconversion from I2 to F, involving changes in secondary structure and the breaking of nonnative interactions, proceeds via a compact transition-state. Interestingly, the native state of the FF1 domain from human p190-A Rho GAP resembles the I2 conformation, suggesting that well-ordered folding intermediates can be repurposed by nature in structurally related proteins to assume functional roles. It is anticipated that the strategy for elucidation of sparsely populated and transiently formed structures of intermediates along kinetic pathways described here will be of use in other studies of protein dynamics.

protein folding | sparse folding intermediate structures | urea *m*-values | chemical exchange saturation transfer (CEST) | NMR

It is now well established that even small single-domain proteins often fold via intermediates (1–4) and a detailed description of each folding trajectory requires, therefore, the determination of atomic resolution structures of all the folding intermediates along the kinetic pathway. Determining the folding mechanism from experimental data and subsequently generating atomic resolution models of the relevant intermediates remains a challenge (1, 5). As a result, detailed folding pathways are not often available even for small proteins that have served as model systems and that have been extensively studied using multiple techniques. Detection of intermediates and the determination of an appropriate exchange mechanism for a given folding reaction is challenging. First, intermediates are often sparsely populated at equilibrium. While the populations of folded (F) and unfolded (U) states can be manipulated in a controlled manner by varying experimental conditions, such as temperature and denaturant concentration, intermediates frequently remain only marginally populated. Second, it is often difficult to distinguish intermediates from each other, as changes in secondary structure including helix elongation, register shifts in β sheets, and formation of new tertiary contacts can escape detection by optical probes. Thus, both detecting and, subsequently, distinguishing between the various intermediates requires techniques that are sensitive to residue or atomic level changes in structure, such as solution NMR spectroscopy. Although solution NMR experiments can be used to probe the structure and dynamics at almost every site in a protein molecule, signals from sparsely populated states are not visible in traditional NMR spectra (6). Nonetheless, over the last two decades, a number of NMR relaxation experiments have been developed to detect minor states of proteins in exchange with the dominant visible state, with lifetimes of the minor conformers ranging from ~ 10 μ s to ~ 100 ms and populations that can be as low as $\sim 0.1\%$ in some cases (7–12). Relaxation-based NMR experiments have been used to study biomolecular conformational dynamics associated with processes such as folding, aggregation, and molecular recognition (5, 8, 9, 13–17), and these techniques have been extended (10, 18–20) to determine structures of transiently populated minor

Significance

Many protein molecules fold into well-defined functional structures. Although machine learning approaches enable accurate prediction of many of these structures, the mechanisms by which they fold remain elusive. Using solution NMR spectroscopy, we describe the folding trajectory of a protein domain at atomic resolution. We show that the rapid collapse from the unfolded ensemble results in a folding intermediate with some elements of native structure, but also with nonnative contacts. Many of these are retained in a more mature and highly compact second intermediate which transitions to the native conformer without expansion of the polypeptide chain. The strategy for structure elucidation of sparse intermediates described here is likely to find application in studies of other dynamic systems.

Author contributions: D.D., N.T., V.P.T., Y.T., D.F.H., L.E.K., and P.V. designed research.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: lewis.kay@utoronto.ca or pramodh@tifr.res.in.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2416682121/-/DCSupplemental>.

Published December 4, 2024.

states in favorable cases (3, 18, 20–26). A third, and more subtle, difficulty with detecting short-lived, sparse states on the exchange pathway of interest occurs when both slow and fast processes are involved, with rapidly exchanging intermediates on the opposite side of the major transition state from the dominant conformer (1, 27). In its simplest form, an example would be an exchange process $F \xrightleftharpoons[k_{FI}]{k_{IF}} I \xrightleftharpoons[k_{IU}]{k_{UI}} U$ in which folding proceeds via an intermediate state I, with the populations of all states, p_i , ordered as $p_F \gg p_U \gg p_I$ and where the global folding transition state separates F from I ($k_{ex,IU} \gg k_{ex,FI}$; $k_{ex,JK} = k_{JK} + k_{KJ}$). In this case, the exchange process probed using the dominant F state can appear as two-state, $F \rightleftharpoons U$, missing the crucial folding intermediate I. The inability to detect early folding intermediates imposes a severe limitation on the description of the folding of even small proteins (1).

Recently, we showed that chemical exchange saturation transfer (CEST) NMR experiments (10, 28–30) can be used to inform on exchange processes in which minor states rapidly interconvert with each other (“minor exchange”), exposing the “blind spot” that has challenged other methodologies, as described above (31, 32) (SI Appendix for details of CEST, SI Appendix, Fig. S1). In addition to extracting a folding model and exchange rates, importantly, chemical shifts (ω) of nuclei in the different states can also be obtained via the CEST approach, potentially allowing the determination of structural models of all intermediates along a protein folding pathway.

Here we illustrate that elucidation of multiple structural intermediates is indeed, possible, using as an example the 71 residue four-helix bundle FF domain from human HYPA/FFBP11 (Fig. 1A), a model system to study protein folding. In the F state helices H1, H2, H3, and H4 adopt an α - α - 3_{10} - α topology in which H3 is the 3_{10} helix (33), while the first ten residues (N-terminal tail) are not part of the folded FF domain and are not discussed further. Extensive folding studies of wild-type (WT) and various mutants of the FF domain using stop-flow, temperature-jump, and Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR experiments have shown that the protein folds via a sparsely populated intermediate (4, 34–36). Interestingly, unlike for all other FF domain variants, a folding intermediate could not be detected for the A39G FF mutant from stop-flow data (35, 36), but only an exchange process in which the major F state interconverts slowly with the minor U state ($p_U \sim 1\%$ at 1 °C). However, on the basis of ^{15}N CEST experiments (SI Appendix) in which both the position and linewidths of minor state dips were used to detect other, even more sparsely populated, states in rapid exchange with one another (32) the A39G FF domain was found to fold along two paths via two intermediates, I1 and I2 (Fig. 1B). Here, the U state exchanges rapidly with I1, $k_{ex,I1U} \sim 8,500 \text{ s}^{-1}$, while the exchange between I1 and I2 is somewhat slower with $k_{ex,I1I2} \sim 1,600 \text{ s}^{-1}$. I1 and I2 then exchange slowly with F ($k_{ex,FI1} \sim 150 \text{ s}^{-1}$ & $k_{ex,FI2} \sim 350 \text{ s}^{-1}$). As the CEST derived I1 state chemical shifts (ω_{I1}) (31, 32) are in excellent agreement with those obtained previously for the folding intermediate of WT FF (3), it can be concluded that both the WT

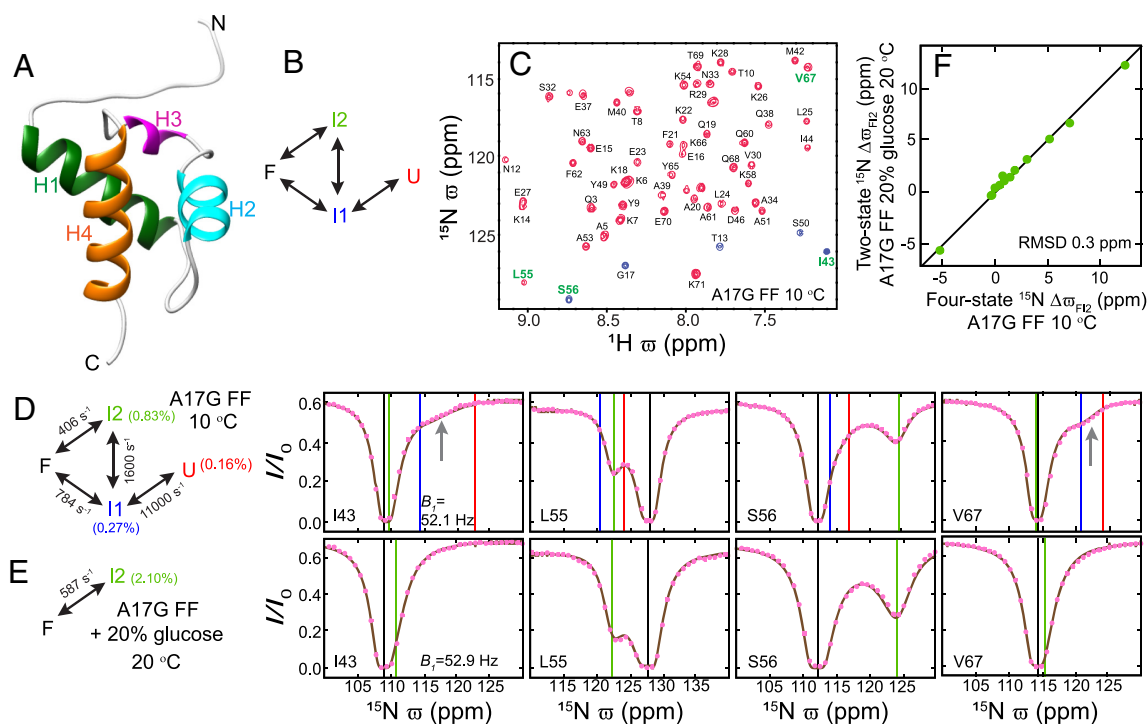


Fig. 1. Simplifying FF domain exchange to obtain I2 state chemical shifts. (A) The native (F) state structure [PDB: 1UZC] (33) of the WT FF domain contains a 3_{10} helix, H3 (P47–S50; magenta), and three α helices, H1 (K14–E27; dark green), H2 (W36–I43; cyan), and H4 (L55–Q68; orange). Helix boundaries were obtained using TALOS-N (37). (B) Schematic illustration of the ^{15}N CEST derived FF domain folding model showing that FF folds via two intermediates I1 and I2 along two pathways. (C) Amide ^1H - ^{15}N correlation map of A17G FF (16.4 T; 10 °C). Peaks aliased in the ^{15}N dimension are shown in blue. Peak assignments are indicated. (D) Amide ^{15}N CEST profiles ($B_1 = 52.1 \text{ Hz}$, $T_{\text{ex}} = 400 \text{ ms}$, 16.4 T, 10 °C) from I43, L55, S56, and V67 in A17G FF. The data were analyzed using the four-state exchange model shown on the left, to produce best-fit exchange parameters (also indicated; SI Appendix, Table S4). The vertical black, blue, green, and red lines correspond to the fitted ω_F , ω_{I1} , ω_{I2} , and ω_U values, respectively. The gray arrow points to the dip that arises due to rapid $I1 \rightleftharpoons U$ exchange. (E) Amide ^{15}N CEST profiles ($B_1 = 52.9 \text{ Hz}$, $T_{\text{ex}} = 400 \text{ ms}$, 16.4 T, 20 °C) from I43, L55, S56, and V67 recorded after adding 20% (w/v) glucose to the sample. The data were analyzed using the two-state exchange model shown, generating the best-fit exchange parameters as indicated. The vertical black and green lines correspond to the fitted ω_F and ω_{I2} values, respectively. In (D) and (E) the experimental data are shown using pink circles while the brown line is calculated using the best-fit parameters. (F) The $\Delta\omega_{FI2}$ values obtained by analyzing the amide ^{15}N CEST data using a four-state model in the absence of glucose (D) and a two-state model in the presence of 20% glucose (E) are very similar, showing that accurate ω_{I2} values can be obtained from a simple two-state analysis of A17G FF CEST data recorded in the presence of 20% glucose.

and A39G FF domains fold via a similar I1 intermediate. Further CEST studies, in which the folding of A17G and WT FF were studied using ^{15}N CEST, confirmed that they also fold via both I1 and I2, providing strong evidence that all the FF domain variants fold via similar pathways that can be described in terms of the same four states (31, 32).

Herein we provide structural models for the complete FF domain folding pathway. Analysis of ^{15}N , $^1\text{H}^{\text{N}}$, $^1\text{H}^{\alpha}$, $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\text{O}}$, $^{13}\text{C}^{\beta}$, and methyl ^1H and ^{13}C CEST profiles allowed us to obtain a large number of chemical shifts for the I2 state which form the basis for determination of the structure of this intermediate using the CS-ROSETTA program (38). The structure of I2 is validated via experiments carried out under conditions where its population is significantly enhanced (~25%). We further recorded CEST experiments as a function of urea concentration, and interpret the resulting urea-based m -values for all sparsely populated states and the four transition-states in terms of compaction along the folding trajectory. Together with the structure of the I1 state that has been determined using CPMG experiments recorded on the WT FF domain under conditions in which the major F state interconverts with the minor I1 conformer (3), the presented I2 structure allows a detailed description of FF domain folding at atomic-resolution. Unlike the “less mature” I1 state where helix H4 is disordered and the overall structure is loosely packed, the CEST derived I2 state structure is fully ordered and as compact as the folded conformer, F, while still retaining some of the nonnative features present in the I1 state. The m -value of the transition-state between F and I2 shows that the interconversion between these compact fully ordered states, involving changes in secondary structure and breaking of either native (F to I2) or nonnative (I2 to F) contacts, proceeds without expansion of the protein. Our study highlights that, at least in some cases, it is possible to obtain structural descriptions of multiple “invisible” intermediates along a reaction, potentially opening up possibilities for a detailed description of the underlying biological processes involved and providing insights into how one might manipulate these processes in cases where they can be associated with function or, alternatively, malfunction.

Results

The determination of structural models for “invisible” protein conformers (so-called excited conformational states) requires chemical shift measurements at a large number of backbone sites (38–40). These shifts must be obtained through relaxation dispersion- or CEST-based experiments, as crosspeaks for excited states are not observed in standard ^1H - ^{15}N or ^1H - ^{13}C experiments. As the FF domain folds via two sparse and transiently formed intermediates along a pair of kinetic pathways (32), the resulting relaxation data must be analyzed in a four-state manner, involving in the most general case twelve different rate constants to reliably obtain chemical shifts of the rare conformers. In the case of the FF domain, it could be established, based on analysis of ^{15}N CEST data, that a simplified four-state model was sufficient (Fig. 1B) (32), allowing extraction of ^{15}N chemical shifts of each conformer (U, I1, and I2). However, it is unlikely that similar four-state fits of less sensitive CEST data (relative to ^{15}N) recorded at backbone ^1H and ^{13}C sites and sidechain $^{13}\text{C}^{\beta}$ carbons would be of sufficient quality to allow robust estimates of these additional chemical shifts. An additional complicating factor relative to ^{15}N data is that dips in ^{13}C CEST profiles recorded using uniformly ^{13}C labeled samples are broadened from $^1\text{J}_{\text{CC}}$ couplings (41–43), while dip linewidths in ^1H CEST profiles are increased due to ^1H - ^1H scalar and dipolar interactions (44). Hence there is a need to “simplify” the exchange process to

robustly obtain the complete set of I2 state chemical shifts necessary for structural studies.

Simplifying Exchange to Determine Chemical Shifts of the Sparsely Populated Folding Intermediate I2. Earlier we had noticed that dips derived from the I2 state are visible in ^{15}N CEST profiles recorded on A17G FF because this variant populates the I2 state to a greater extent than is the case for A39G FF or WT FF domains (32). The amide ^1H - ^{15}N correlation map of A17G FF is well resolved (Fig. 1C) and representative amide ^{15}N CEST profiles from I43, L55, S56, and V67 are shown in Fig. 1D. The ^{15}N CEST data ($B_1 = 26, 52.1, 104.1,$ and 208.3 Hz) were analyzed using a four-state exchange model (Fig. 1B; $\chi_{\text{red}}^2 \sim 0.7$, $k_{\text{ex},F\text{I}1} = 784 \pm 67 \text{ s}^{-1}$, $k_{\text{ex},F\text{I}2} = 406 \pm 5 \text{ s}^{-1}$, $k_{\text{ex},\text{I}1\text{I}2} = 1,600 \pm 113 \text{ s}^{-1}$, $k_{\text{ex},\text{I}1\text{U}} = 11,000 \pm 1,064 \text{ s}^{-1}$, $p_{\text{I}1} = 0.27 \pm 0.01\%$, $p_{\text{I}2} = 0.83 \pm 0.01\%$ and $p_{\text{U}} = 0.16 \pm 0.02\%$). In Fig. 1D the CEST derived ^{15}N ϖ_{F} , $\varpi_{\text{I}2}$, $\varpi_{\text{I}1}$, and ϖ_{U} chemical shift values (ppm) are shown on the CEST profiles using black, green, blue, and red vertical lines, respectively. The shifts agree well with those obtained previously for A39G FF (SI Appendix, Fig. S2), once again confirming that the different FF variants fold via the same intermediates. Notably, the ^{15}N CEST profiles for I43 and V67 contain a broad dip (indicated by the gray arrows in Fig. 1D) between the blue and red lines, arising due to the fast exchange between the I1 and U states. On the other hand, CEST profiles from L55 and S56 contain a clear dip near the green line due to the I2 state. As additives such as glucose and glycerol stabilize folded protein states (F) (45) we added glucose to the buffer in the hopes of reducing the populations of both U and the partially disordered I1 state relative to I2. Indeed, the addition of 20% (w/v) glucose reduces both $p_{\text{I}1}$ and p_{U} as can be seen from the disappearance of the dips arising from the $\text{I}1 \rightleftharpoons \text{U}$ interconversion (compare CEST profiles from I43 and V67 in Fig. 1D and E), while the dips arising from the I2 state remain intact (compare L55 and S56 in Fig. 1D and E). The 20% glucose A17G FF ^{15}N CEST data is well fit using a two-site exchange model ($\chi_{\text{red}}^2 \sim 1.1$; $B_1 = 26.4$ & 52.9 Hz) with $k_{\text{ex},F\text{I}2} = 587 \pm 12 \text{ s}^{-1}$ and $p_{\text{I}2} = 2.1 \pm 0.01\%$, and, importantly, the I2 state chemical shifts obtained from the two-state analysis agree very well with the corresponding values from four-state fits of CEST data recorded in the absence of glucose (RMSD 0.3 ppm; Fig. 1F). Thus, the 20% glucose buffer stabilizes the F and I2 states, simplifying the four-state exchange process to one which can be described in terms of F and I2, so that the I2 state chemical shifts can be obtained accurately. Further, in the presence of 20% glucose, the population of I2 increases from ~1% at 10 °C to ~2% at 20 °C, which is why the higher temperature was chosen for the two-state analysis above, as opposed to our initial study in the absence of glucose which was conducted at 10 °C.

Assembling a Near-Complete Set of I2 State Backbone Chemical Shifts. Having established that accurate $\varpi_{\text{I}2}$ values can be obtained from a two-state analysis of CEST data measured on A17G FF samples in 20% glucose buffer, we next recorded ^1H and ^{13}C CEST profiles to obtain minor state chemical shifts at nearly all backbone and $^{13}\text{C}^{\beta}$ sites using CEST experiments that have been developed over the last decade (30, 42, 43, 46–48). Analysis of these CEST profiles in a two-state manner resulted in an extensive set of backbone I2 state chemical shifts (Fig. 2 and SI Appendix, Tables S1–S3). Amide $^1\text{H}^{\text{N}}$ and ^{15}N shifts were obtained for 57 out of 59 nonproline sites, $^1\text{H}^{\alpha}$ shifts at all 62 sites including one Gly residue, $^{13}\text{C}^{\alpha}$ shifts at all 61 sites, $^{13}\text{C}^{\text{O}}$ shifts at 58 out of 61 sites, and $^{13}\text{C}^{\beta}$ shifts at 55 out of 60 sites. To the best of our knowledge, such an extensive set of chemical shifts has not been previously reported for a transiently populated

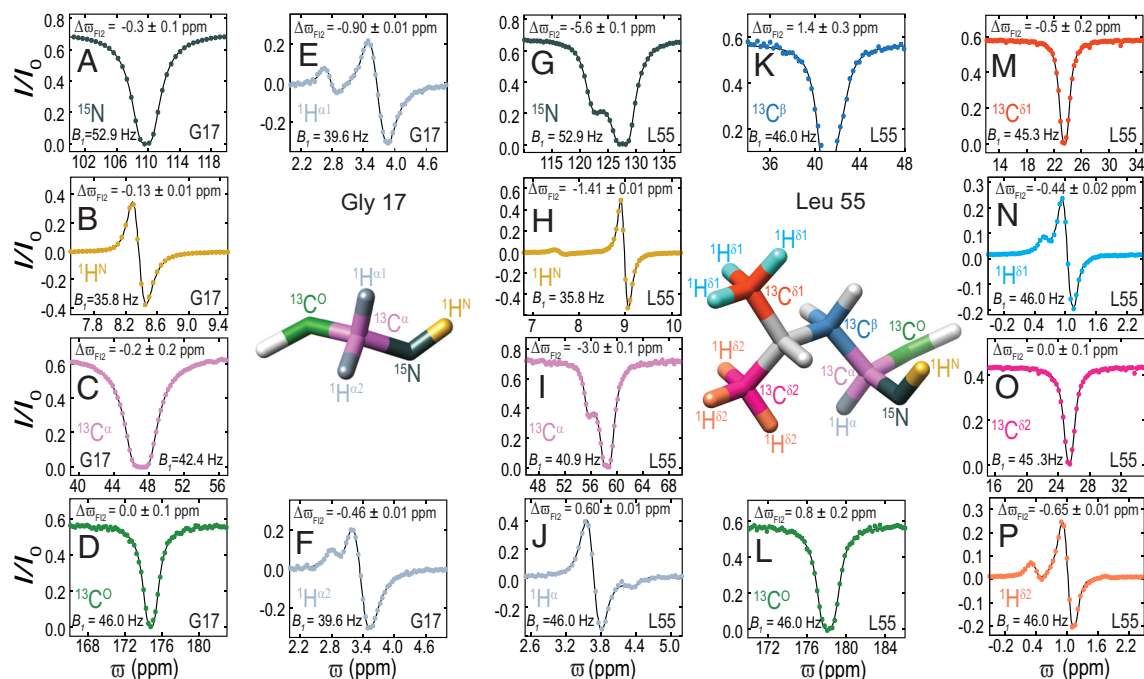


Fig. 2. Reconstructing NMR spectra of the sparsely populated I2 state of the FF domain. ^1H , ^{15}N , and ^{13}C CEST profiles from various sites in G17 (A–F) and L55 (G–P) recorded using A17G FF samples in 20% [^2H] glucose buffer (16.4 T, 20 °C). The B_1 value used to record the CEST data and the best-fit $\Delta\varpi_{FI2}$ value are indicated for each site. Experimental data are shown using colored circles, while the black line is calculated using best fit two-state parameters. Experimental details are given in *SI Appendix*.

minor protein state. ^1H and ^{13}C CEST experiments (43, 46, 49) were also performed to obtain ϖ_{I2} values for Ala, Ile, Leu, and Val methyl sites (*SI Appendix, Table S1–S3*).

Folding Intermediate I2 Is Well Ordered but Contains Nonnative Structural Elements and Contacts. To evaluate, initially qualitatively, the structure of the I2 state, we first compared its chemical shifts to those of the ground state (Fig. 3A). It is clear that significant shift changes between the two states ($\Delta\varpi_{RMS,FI2} > 0.5$ ppm) are largely localized to the stretch of residues between S50 and Q60, starting at the C terminal end of H3 and extending into the N-terminal end of H4 in the folded structure (F). The chemical shift changes between F and I1 are more extensive, including some residues at the C terminal end of H1, in addition to the stretch between S50 and Q68 encompassing the whole of H4 (Fig. 3B). Residue specific S^2 values calculated from the backbone chemical shifts (RCI S^2) (50) provide a measure of the protein backbone flexibility, with S^2 values > 0.8 indicative of ordered structure, as is observed throughout the F state with the exception of the terminal few residues (Fig. 3C). Notably, the S^2 vs residue correlation for the I2 state is very similar to that of the F conformer (Fig. 3C), indicating that I2 is also well ordered. However, I1 is less ordered, with the intervening residues between H1 and H2 as well as H4 less stable (3). Estimates of residue-specific helix propensities derived from backbone chemical shifts establish that I2 also contains four helices, however, a comparison of I2 and F helix propensities shows that H3 extends to K54 in I2 instead of ending at S50 as in F, while H4 starts at K58 in I2 rather than at L55 as in F (Fig. 3D).

The extensive set of CEST derived I2 backbone chemical shifts were then used as input into the CS-ROSETTA program (38) to calculate the structure of the A17G FF I2 state (Fig. 3E). The CS-ROSETTA protocol converged, with a C^α RMSD of 0.8 ± 0.2 Å for the ten lowest energy structures relative to the lowest energy I2 conformer (*SI Appendix, Fig. S3*). In an analogous calculation, serving as a control, and performed using F state chemical shifts

for the same sites as for I2, a C^α RMSD of 0.9 ± 0.2 Å was obtained; more importantly, a C^α RMSD of 1.4 ± 0.2 Å was calculated for the lowest ten energy F state structures relative to the WT FF domain F state structure that was generated using standard NMR (i.e., NOE-based) techniques (33) (*SI Appendix, Fig. S3*). This control calculation provides confidence that the set of CEST-based chemical shifts used in this study is sufficient to define the FF domain I2 structure with CS-ROSETTA.

The CEST derived (A17G FF) I2 state structure along with the (WT FF) F, and (WT FF) I1 state structures are shown in a pair of orientations in Fig. 3E, F, and G, respectively. As expected, based on the analysis of chemical shifts in Fig. 3A–D, the I2 state (Fig. 3E) is ordered, compact, and consists of four helices as in the F state (Fig. 3F), with H3 elongated at its C-terminal end relative to F. As a consequence of this change in the secondary structure of the H3/H4 region, L55 at the start of H4 and interacting with H2/H3 residues (W36/Y49) in the F state (Fig. 3F), is repositioned to within the H3–H4 loop of I2 at the H1/H3 interface (Fig. 3E). Thus, in the I2 state L55 occupies the position of L52 in the F state (compare Fig. 3E and F). As a consequence, in the I2 state of A17G FF (Fig. 3E) the L55 $\delta 1$ and $\delta 2$ methyl groups are proximal to H^α of G17 (H1), while in F of WT FF the L52 $\delta 2$ methyl group is proximal to A17 (Fig. 3F). Further, repositioning of L52 in the I2 state (Fig. 3E) brings it proximal to A20 (H1) and Y49 (H3), in contrast to L52 in F which contacts residue 17. Similar changes in secondary structure and sidechain orientation/packing are seen in the WT I1 state structure (Fig. 3G), obtained from CPMG derived chemical shifts (3). In I1, as in I2, H3 is extended at its C terminal end placing L55 in proximity of A17. Unlike I2 and F, however, H4 is largely disordered in I1. Thus, the structure of I2 shares similarities with those of both F and I1. Finally, the loop connecting H1 and H2 (K28 to S35) is rigid in I2 and F but flexible in I1 (Fig. 3C). Insight into why this is the case is obtained from the structures which show that this loop docks onto the C-terminal end of H4 in F and I2 (Fig. 3E and F) with the sidechain of S32 from the H1–H2 loop in close contact with the sidechains of K66 from

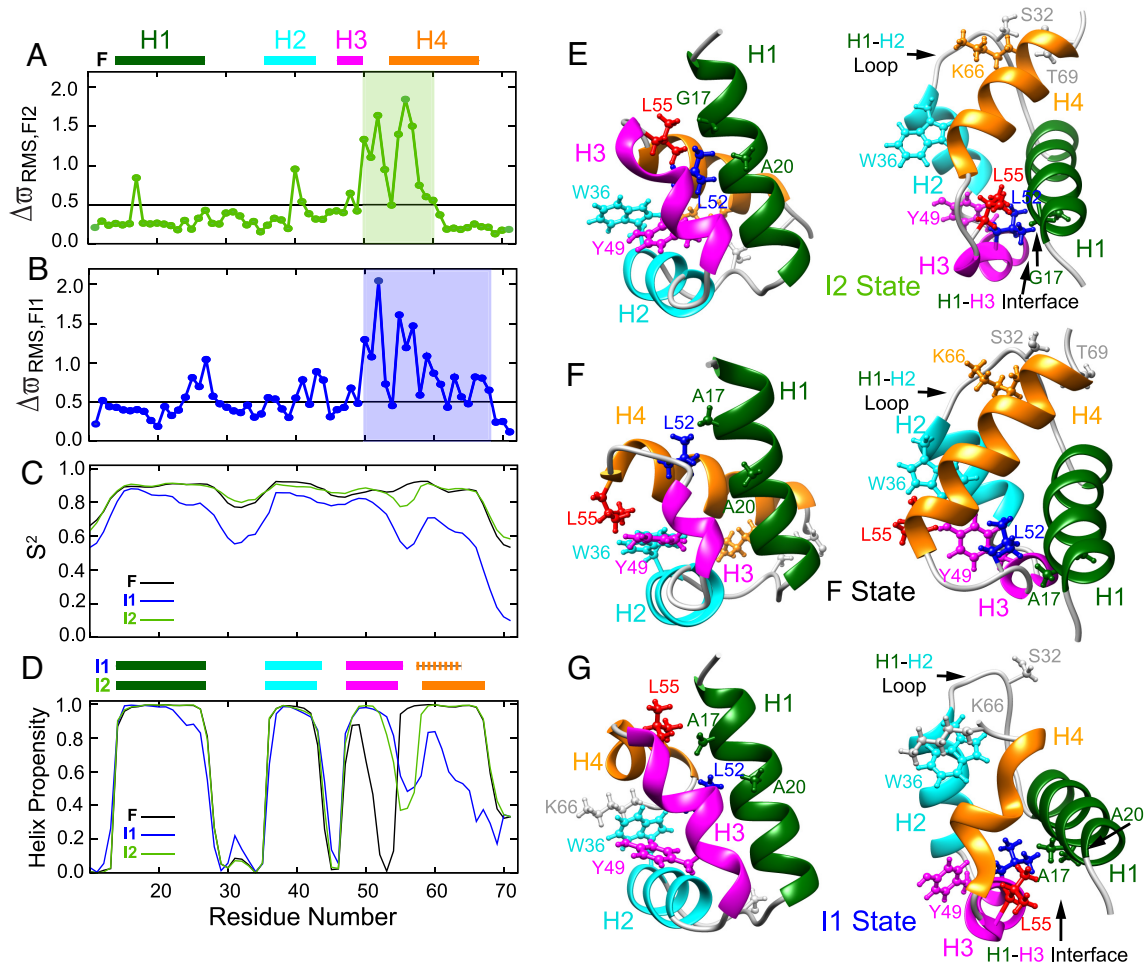


Fig. 3. Insights into the structure and dynamics of the I2 state from CEST derived chemical shifts. Plots of $\Delta\sigma_{RMS,F12}$ (A) and $\Delta\sigma_{RMS,F11}$ (B) as a function of residue. Here, $\Delta\sigma_{RMS,KJ} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta\sigma_{KJ}^i / \sigma_{std}^i)^2}$ is calculated for each residue. The summation over i extends over all the backbone nuclei (^{15}N , $^1\text{H}^\alpha$, $^1\text{H}^\beta$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$) and sidechain $^{13}\text{C}^\beta$ carbons in the residue for which the $\Delta\sigma_{KJ}$ values are available, σ_{std}^i is the SD in reported chemical shifts (51) for nucleus i . In (A) and (B) the horizontal black line is drawn at $\Delta\sigma_{RMS,F12/I1} = 0.5$ ppm. $\Delta\sigma_{F12}$ values (A) are from the CEST experiments performed in this study, while $\Delta\sigma_{F11}$ values (B) were obtained previously using CPMG experiments (3). (C and D) Residue specific S^2 /helix propensity values estimated using TALOS-N from the chemical shifts of the F state (black), CEST derived chemical shifts of the I2 state (green), and CPMG derived chemical shifts of the I1 state (blue). Helix boundaries obtained using TALOS-N are shown for the F state above panel A (H1 (K14–E27), H2 (W36–I43), H3 (P47–S50), and H4 (L55–Q68)) and for the I2/I1 states above panel (D). Helix boundaries in the I2 state are H1 (K14–E27), H2 (W36–I43), H3 (P47–K54), and H4 (K58–Q68) and in the I1 state: H1 (K14–E27), H2 (W36–I44), H3 (P47–L55), and H4 (E57–A64). Two views (orientations) of the A17G FF CEST derived I2 state (E), WT FF F state (F) [PDB: 1UZC] (33), and WT FF CPMG derived I1 state [PDB: 2KZG] (3) (G) structures with key residues shown using the ball and stick representation. In (E–G) L55 is colored in red and L52 is colored in blue. All other residues have the same color as the helix in which they are situated or are in gray if they are in loops. Residues from N12 to E70 are shown in the I2 and F state structures (E and F) while residues from N12 to K66 are shown in the I1 state structure (G). In (E–G) only the lowest energy structure is displayed.

H4 and T69 that is adjacent to H4. In contrast, the C-terminal end of H4 is disordered in I1 and consequently unavailable for docking to the H1–H2 loop, which therefore remains flexible.

Validating the CEST-Derived Structure of I2. Having calculated the structure of the I2 folding intermediate using a CS-ROSETTA protocol based on an extensive collection of chemical shifts, we next sought to validate it. The distance between the G17 H^α and L55 H^δ protons is large in the F state (G17 C^α –L55 $\text{C}^{\delta 1}/\text{C}^{\delta 2} \sim 15$ Å) but is significantly reduced in the I2 state (G17 C^α –L55 $\text{C}^{\delta 1}/\text{C}^{\delta 2} \sim 4.5$ Å; *SI Appendix, Fig. S4*), despite the fact that the structure in the stretch around G17, E15–A20, does not change significantly between F and I2 (C^α RMSD of ~ 0.3 Å in this region). Focusing on G17, $|\Delta\sigma_{F12}|$ is small for $^1\text{H}^\alpha$ (0.13 ppm), ^{15}N (0.31 ppm), $^{13}\text{C}^\alpha$ (0 ppm), and $^{13}\text{C}^\beta$ (0.22 ppm) nuclei (Fig. 2 A–D), but large for the two G17 H^α sites (0.90 and 0.46 ppm; Fig. 2 E and F). Further, large $|\Delta\sigma_{F12}|$ values are also noted for the L55 methyl $\delta 1/\delta 2$ protons (0.44 ppm and 0.65 ppm; Fig. 2 N and P). Taken

together, the large shift changes for G17 H^α and L55 H^δ protons between F and I2 conformers are consistent with a structural rearrangement involving these pair of residues, with G17–L55 proximal in one conformation (such as I2) and more distal in a second state (such as F), for example.

More conclusive validation of the CEST derived structure of the I2 state can be obtained by recording NOEs, as the small distance between G17 $\text{H}^{\alpha 1/\alpha 2}$ and L55 methyl $\text{H}^{\delta 1,\delta 2}$ in the I2 state favors cross relaxation between the two sets of protons. However, NOEs between G17 and L55 protons were not observed in a ^{13}C – ^{13}C – ^1H HSQC–NOESY–HSQC dataset recorded with an A17G FF sample at 20 °C using $T_{MX} = 100$ ms (Fig. 4 A and B) probably because p_{I2} is only $\sim 2\%$ with rapid exchange to state F ($k_{ex,F12} = 474 \pm 8 \text{ s}^{-1}$) where the distance between these protons is large (~ 15 Å). As S56, which is part of H4 in the F state, repositions to the loop between H3 and H4 in the I2 state structure (Fig. 3D), we reasoned that replacing S56 with a proline might increase p_{I2} by destabilizing F. A high-quality spectrum of A17G

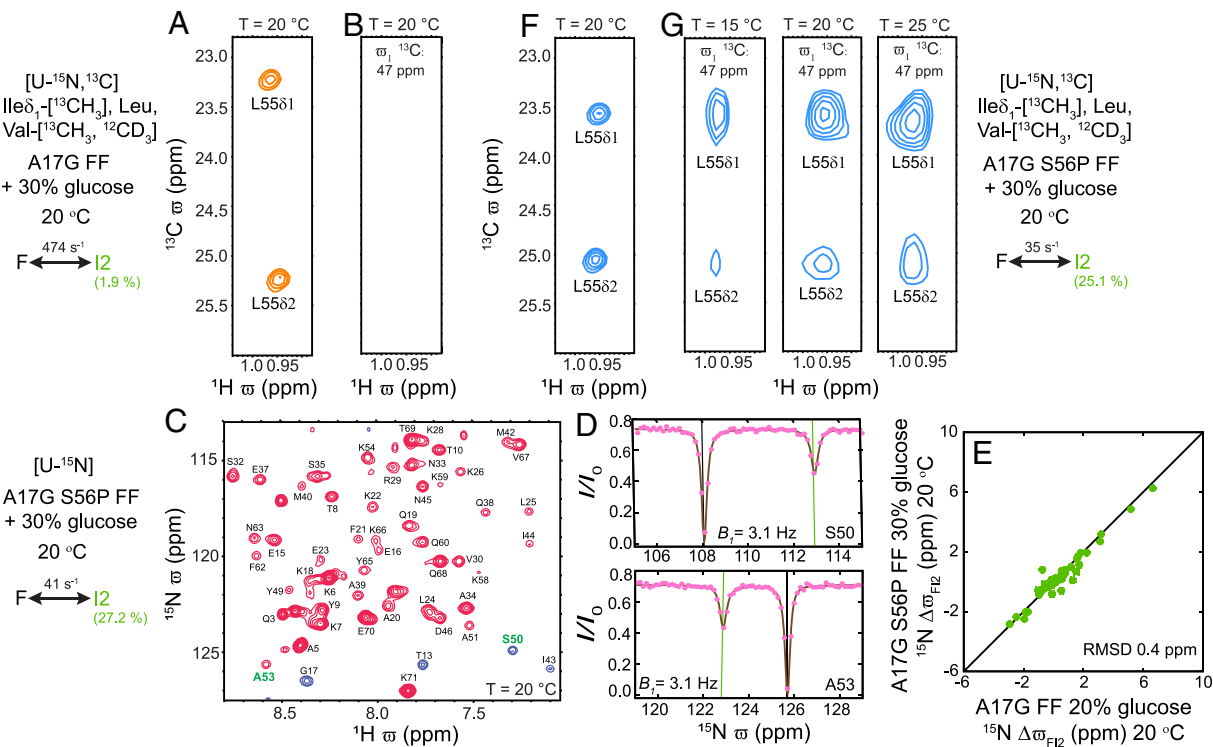


Fig. 4. Validating the I2 state structure. (A) Selected region of the A17G FF ^1H - ^{13}C HSQC map showing L55 $^1\text{H}^{\delta 1}$ - $^{13}\text{C}^{\delta 1}$ and $^1\text{H}^{\delta 2}$ - $^{13}\text{C}^{\delta 2}$ correlations. (B) Strip at the G17 $^{13}\text{C}^{\alpha}$ chemical shift (47 ppm) from the A17G FF 3D HSQC-NOESY-HSQC spectrum does not contain correlations that would be expected for an NOE between G17 $^1\text{H}^{\alpha 1/\alpha 2}$ and L55 $^1\text{H}^{\delta 1/\delta 2}$ protons (distances $< 5 \text{ \AA}$). (C) Amide ^1H - ^{15}N correlation map of [U- ^{15}N] A17G S56P FF (16.4 T; 20 °C). (D) Representative amide ^{15}N CEST profiles. Green and black lines are used to indicate ω_{I2} and ω_F values, respectively. (E) Correlation between ^{15}N CEST derived $\Delta\omega_{FI2}$ values for A17G FF and A17G S56P FF (from D). (F) Selected region of the A17G S56P FF ^1H - ^{13}C HSQC correlation map, highlighting correlations from L55. (G) Strips at the G17 $^{13}\text{C}^{\alpha}$ chemical shift (47 ppm) from the A17G S56P FF 3D HSQC-NOESY-HSQC spectrum recorded at 15 °C ($k_{ex,FI2} \sim 18.8 \pm 0.9 \text{ s}^{-1}$; $p_{I2} \sim 20.4 \pm 0.5\%$), 20 °C ($k_{ex,FI2} \sim 35.1 \pm 0.9 \text{ s}^{-1}$; $p_{I2} \sim 25.1 \pm 0.4\%$), and 25 °C ($k_{ex,FI2} \sim 62.8 \pm 2 \text{ s}^{-1}$; $p_{I2} \sim 27.6 \pm 0.4\%$) showing NOE correlations between G17 $^1\text{H}^{\alpha 1/\alpha 2}$ and L55 $^1\text{H}^{\delta 1/\delta 2}$ protons. (A, B, F, G) The methyl HSQC and NOESY experiments for A17G FF and A17G S56P FF were recorded with samples that were [U- ^{15}N , ^{13}C] at all sites other than the ILV sidechains that were selectively methyl labeled (Ile δ_1 -[^{13}C], Leu, Val-[^{13}C], $^{12}\text{CD}_3$) (SI Appendix, Materials and Methods). All the experiments were carried out in 30% [^2H] glucose, 10% D_2O buffer. The NOE mixing time was set to 100 ms in all the NOESY experiments. The slight difference in fitted exchange parameters for the $F \rightleftharpoons I2$ interconversion (C, Left; G, Right) reflects the difference in the extent of ^2H enrichment in the protein samples analyzed.

S56P FF was obtained (Fig. 4C), with only small changes in peak positions relative to A17G FF (^{15}N RMSD = 0.2 ppm, excluding positions 55 to 57), and ^{15}N CEST data were recorded at 20 °C in 30% glucose buffer to stabilize I2 relative to U and I1 (Fig. 4D). The data were well fit to a two-state process of interconversion between F and I2 states (Fig. 4E) with $k_{ex,FI2} \sim 41.4 \pm 1.3 \text{ s}^{-1}$ and $p_{I2} \sim 27.2 \pm 0.5\%$ ($\chi^2_{red} \sim 1.2$), although the exact kinetic parameters depend on the degree of ^2H enrichment in the protein and the fraction of D_2O in the solvent. The relatively small differences in $\Delta\omega_{FI2}$ values between A17G and A17G S56P FF samples (Fig. 4E, RMSD = 0.4 ppm, excluding residues 55 to 57) indicate that the addition of the proline mutation is unlikely to significantly affect the I2 conformation. Notably, the large increase in p_{I2} (from $\sim 2\%$ to $\sim 25\%$, 20 °C) for A17G S56P FF now results in the observation of NOEs between H^{α} protons of G17 and the methyl δ protons of L55 at 15, 20, and 25 °C ($T_{MIX} = 100 \text{ ms}$; Fig. 4G), consistent with the fact that L55 and G17 are proximal in the I2 state and in support of the CEST derived I2 structure. It is worth emphasizing that although these NOEs are detected between F state resonances, the NOE enhancement occurs when the molecule samples the I2 state, in a manner analogous to a transferred NOE effect. The origin of these NOEs is made clear from the increase of their intensities with temperature (Fig. 4G) which occurs concomitantly with the temperature-dependent increase in p_{I2} (Fig. 4G and SI Appendix, Figs S4 and S5). In contrast, if the NOEs were to arise from magnetization transfer exclusively within the F state an opposite temperature dependence would be predicted, since in the macromolecular limit the NOE scales directly with

rotational correlation time, which becomes smaller with increase in temperature. Finally, the observed NOEs do not arise from transfer within the I1 conformer as a three-state analysis of the A17G S56P FF ^{15}N CEST data establishes that $p_{I1} < 2\%$ at 20 °C (SI Appendix, Fig. S6), too low of a population to give rise to observable NOE peaks (Fig. 4B). The excellent correlation between the amide ^{15}N $\Delta\omega_{FI2}$ values of A17G and WT FF [RMSD $\sim 1.1 \text{ ppm}$; (31, 32)] suggests that the I2 state structure derived here is a good mimic of the I2 conformer in WT FF. However, it is worth noting that the A17G mutation can lead to differences between the two structures at the level of sidechain packing.

CEST Derived Urea m -Values Provide Additional Insights into the FF Domain Folding Mechanism. With atomic resolution structures of the F, I1, and I2 states now available, we next sought to quantify the structural rearrangements that occur during their interconversion. For example, does the FF domain undergo extensive unfolding when it interconverts between compact F and I2 states to facilitate breakage of nonnative interactions and formation of native contacts (I2 to F) or vice versa (F to I2)? The urea m -value is a measure of how the free energy differences between states vary with urea concentration (45), with the m -value of state J with respect to the folded state (F) defined as $m_J = -\frac{d\Delta G_{FJ}}{d[\text{urea}]}$, where [urea] is the urea concentration, $\Delta G_{FJ} = G_J - G_F$, and G_J is the Gibbs free energy of state J. States with greater solvent-exposed surface areas have more urea binding sites and consequently higher urea m -values, so that the urea m -value is a reporter for

the compactness of the interconverting states, including the transition states involved (note that J can be a transition state) (52). For example, a U state that is completely disordered and expanded would have a higher m -value than other, partially folded or completely folded, conformers. To obtain m -values for the three minor states and four transition states of the A17G FF folding trajectory, a series of ^{15}N CEST datasets (Fig. 5A) was recorded at 2.5 °C using A17G FF samples prepared with varying amounts of urea (from 0 to 1 M). The resulting CEST profiles were fit to the four-state exchange model of Fig. 1B to obtain exchange parameters at the different urea concentrations (SI Appendix, Table S4). As discussed in SI Appendix, m -values for the three minor states were obtained from the urea dependencies of their populations, while the transition state m -values were calculated from the urea dependencies of populations and rate constants (Fig. 5B and C). Recall that all m -values reported here are with respect to the F state. As expected, among the three minor states, U has the highest m -value ($6.6 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) as it is disordered, I1 the second-highest m -value ($2.8 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$), consistent with it being fairly structured but with H4 disordered, while m_{I2} ($-0.1 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) is indistinguishable from m_F ($0 \text{ kJ mol}^{-1} \text{ M}^{-1}$) by definition, consistent with an I2 state structure that is as compact as the F state. To some extent these m -values can be rationalized by visually inspecting the amide ^{15}N CEST profiles recorded at different urea concentrations (Fig. 5A). In the ^{15}N CEST profile of S56 the size of the I2 state dip at ϖ_{I2} (near the green line) does not change with urea concentration consistent with little change in p_{I2} and, hence, $m_{I2} \sim 0 \text{ kJ mol}^{-1}$

M^{-1} . On the other hand, the dip that results from the (rapid) I1-U interconversion, located between ϖ_{I1} (blue line) and ϖ_U (red line) becomes more prominent and moves toward ϖ_U as the urea concentration increases, as the population of the U state increases more rapidly with urea than the population of I1, a result consistent with $m_U > m_{I1}$. Transition state m -values, m_{TSAB} , can also be interpreted in terms of the compactness of the respective transition state (TS_{AB}), in this case connecting states A and B. For example, $m_{TS_{UI1}}$ ($4.0 \pm 0.2 \text{ kJ mol}^{-1} \text{ M}^{-1}$) lies in between m_U ($6.6 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) and m_{I1} ($2.8 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$), establishing that TS_{UI1} is more compact than U but less compact than I1. Considering the two folding trajectories of the four-state model of Fig. 1B used to analyze the data, $m_U > m_{TS_{UI1}} > m_{I1} > m_{TS_{I1F}} > m_F$ and $m_U > m_{TS_{UI1}} > m_{I1} > m_{TS_{I12}} > m_{I2}$ (Fig. 5D), it is clear that the FF domain progressively becomes more compact as the folding reaction proceeds from U to I1 to F and from U to I1 to I2. However, $m_{TS_{I2F}}$ ($0.3 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) is only slightly larger than $m_F \sim m_{I2}$ ($\sim 0 \text{ kJ mol}^{-1} \text{ M}^{-1}$), establishing that interconversion from the compact I2 state to F, involving changes in secondary structure and breaking of nonnative interactions, also proceeds via a compact transition state TS_{I2F} that is only slightly larger than F and I2 but significantly more compact than both TS_{I1F} ($m_{TS_{I1F}} = 2.1 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$) and the I1 folding intermediate ($m_{I1} = 2.8 \pm 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$). The profile of urea m -values along the A17G FF folding trajectory reported here agrees, generally, with that published previously (4); however, there are some differences. In the present work, measurements were carried out at 2.5 °C, as opposed to 25 °C, and it is known that urea m -values decrease with

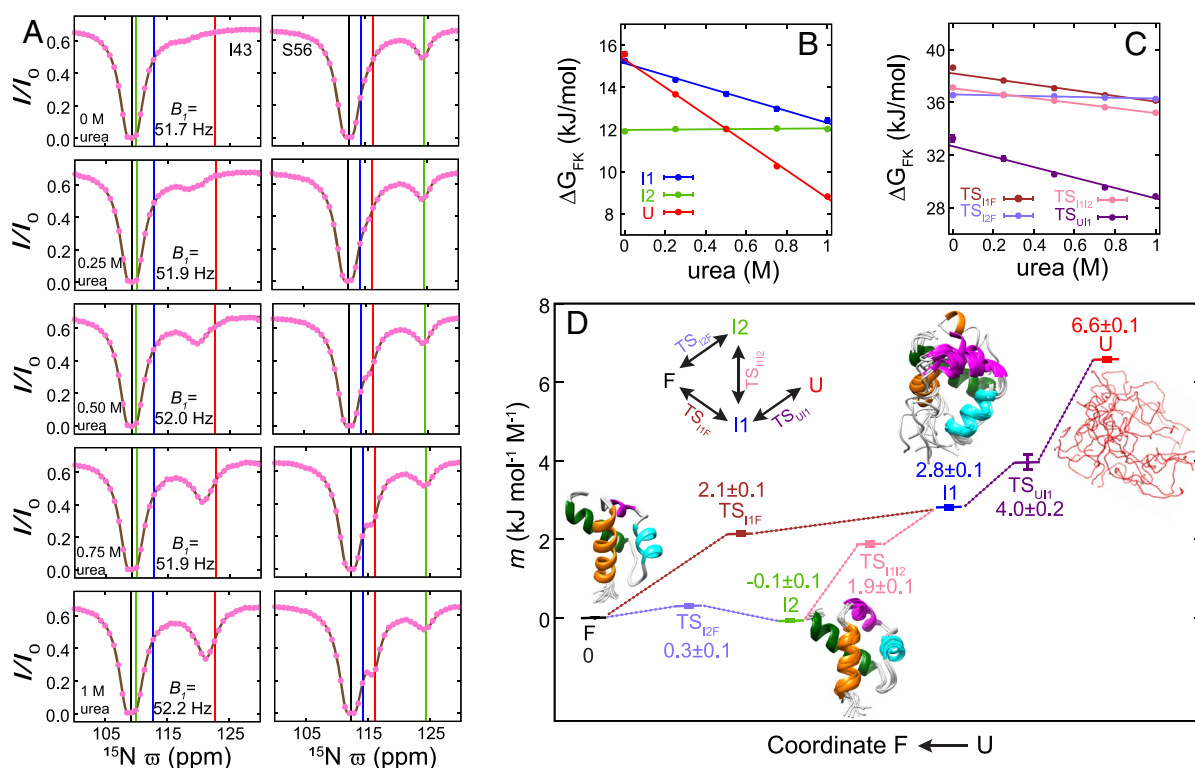


Fig. 5. m -values describe how the compaction of A17G FF evolves along the folding trajectory. (A) Representative amide ^{15}N CEST profiles from I43 and S56 in A17G FF (16.4 T; $T_{\text{Ex}} = 400 \text{ ms}$; 2.5 °C) recorded with varying amounts of urea. Experimental data are shown with pink circles while the continuous brown line in each panel is calculated from the best-fit parameters. The vertical black, blue, green, and red lines correspond to the fitted m_F , m_{I1} , m_{I2} , and m_U values, respectively. (B) Variation of ΔG_{F1} , ΔG_{F2} , and ΔG_{FU} as a function of urea concentration. The m -value for state K (m_K) is obtained from the slope of the urea concentration dependence of $\Delta G_{FK} = -RT \ln(p_K/p_F)$, with p_K and p_F obtained from the analysis of the amide ^{15}N CEST profiles recorded at different urea concentrations, as illustrated in (A). (C) Variation of $\Delta G_{F_{TS11F}}$, $\Delta G_{F_{TS12F}}$, $\Delta G_{F_{TS112}}$, and $\Delta G_{F_{TS_{UI1}}}$ as a function of urea concentration, where $\Delta G_{F_{TS_{kj}}}$ is the free energy difference between TS_{kj} and the F state. The m -value for transition state TS_{KL} ($m_{TS_{KL}}$) is obtained from the slope of the urea concentration dependence of $\Delta G_{F_{TS_{KL}}} = -RT \ln(p_K/p_F) - RT \ln(k_{KL}/C)$, where p_K , p_F and k_{KL} are derived from the analysis of ^{15}N CEST profiles recorded as a function of urea concentration. The constant C is set to be 10^7 s^{-1} and does not affect the value of $m_{TS_{KL}}$. Best-fit four-state exchange parameters obtained at various urea concentrations are listed in SI Appendix, Table S4. (D) Urea m -values obtained from (B) and (C) at various points along the folding landscape of A17G FF. Ten structures (W11-K71) of the F state (WT FF), the I1 state (WT FF), and the I2 state (A17G FF) are shown. In the case of the I1 state, residues corresponding to the disordered H4 helix have been added for the purpose of illustration.

increasing temperature (53). In addition, the previous analysis of FF domain folding data did not consider the I2 state (i.e., only included F, I1, and U), yet it is now clear that I1 can either convert to F directly or via I2. It is likely that neglect of this second I1 to F pathway through I2 leads to some of variations in m -values in the two studies.

Discussion

Herein we present a four-state model describing the folding pathway of the FF domain, including atomic resolution structural models of the two intermediate states, I1 and I2, that are formed during this process. Starting from the U state there is rapid collapse on the ~ 100 μ s timescale, forming I1 that contains H1, H2, and an elongated H3, with stabilization of these elements occurring through a network of nonnative (and native) contacts. State I1 ($m_{I1} = 2.8 \pm 0.1$ kJ mol⁻¹ M⁻¹) is more compact than state U ($m_U = 6.6 \pm 0.1$ kJ mol⁻¹ M⁻¹), but more loosely packed compared to state F. Maximizing hydrophobic interactions presumably is the driving force behind the rapid collapse from U to I1, resulting in the formation of nonnative structural elements in I1. The I1 intermediate is then able to fold to F either in a single step that involves both the breaking of nonnative interactions and the formation of H4 or in a two-step manner via I2 that first involves formation of H4 followed by the breaking of nonnative interactions when I2 converts to F. From the fitted kinetic rates it is clear that the barrier for the interconversion between native and nonnative structural features ($F \rightleftharpoons I1$, $F \rightleftharpoons I2$) is higher than the barrier for the order/disorder transition ($I1 \rightleftharpoons U$, $I2 \rightleftharpoons I1$). Slow rates associated with “fixing” nonnative interactions to form the folded state have also been observed in a comparative study of the WT and L24A FF domains (24, 35). The I1 state of the L24A FF variant was shown to be stabilized by an increased number of nonnative contacts relative to the corresponding WT intermediate, accounting for the approximate ten-fold slower I1 to F folding rate for the mutant domain (24). The fact that formation of the F state from I2 can proceed without significant changes in compaction of TS_{I2F} is notable, considering the significant structural rearrangements that must occur. A similar situation has been observed previously in studies of the L99A cavity mutant of T4 lysozyme. In this case, F114 that is at the junction of two helices and exposed to solvent in the major state becomes buried in the core of the protein in a transiently populated minor state (23). Just as in the case of the $I2 \rightleftharpoons F$ FF domain interconversion, computations show that F114 is able to swing into the core of protein without any largescale expansion as two neighboring helices transiently move apart by a few Å (54).

Additional insight into the accumulation of structure along the folding trajectory, focusing on transition states, may be obtained via Φ -value analysis (27). An extensive set of experiments on the WT FF domain and associated mutants (35), analyzed assuming a simple $F \rightleftharpoons U$ folding reaction, shows that only a small part of the protein that includes the end of H1, the beginning of H2 and intervening loop residues are structured in the rate-limiting transition state. The transition state described by this analysis largely corresponds to TS_{I1F} (3), as it is more stable than TS_{I2F} under the conditions used to perform the Φ -value experiments (2 and 6 M urea). Thus, only regions of I1 that are native-like are conserved in TS_{I1F} , while the start of H1 and the H3-H4 loop that form nonnative contacts and H3 that is elongated in I1 are not preserved. Helix H4 is not formed in TS_{I1F} but tertiary interactions between H4 residues and the protein core are beginning to emerge (35). Data from TS_{I2F} , TS_{I1I2} , and TS_{UI1} are not available from this Φ -value analysis. However, Φ -values for all these transition-states

can now be measured because ¹⁵N CEST data recorded using various FF mutants (for Φ -value analysis) can be analyzed (separately) to obtain four-state exchange parameters.

The I2 state structure determined in this study is as compact and as ordered as the F state. We wondered, therefore, whether I2-like structures might serve as ground states of other FF domains, and, if this is the case, whether such a structure could be functionally relevant. Indeed, in the native state structure of the FF1 domain from human p190-A Rho GAP (RhoGAP-FF1) helix H3 is elongated relative to what is observed in the canonical FF native state (55), and the equivalent of L55 in the canonical FF domain (L311) makes a contact with the equivalent of A17 (A273) in H1, as in the I2 state structure (*SI Appendix, Fig. S7*). Interestingly Y308 (L52 here) of RhoGAP-FF1 that becomes phosphorylated to inhibit an interaction with the transcription factor TFII-I (55, 56) is buried in the core of the domain and cannot be accessed by the kinase. Based on the observation that phosphorylation occurs above 310 K, a temperature at which the amide ¹H-¹⁵N correlation map becomes less well-dispersed, it had been suggested that RhoGAP-FF1 must unfold for Y308 to be phosphorylated (55, 57). However, the melting point of RhoGAPFF1 is 325 K, somewhat higher than 310 K at which kinase activity has been observed. Building on our current understanding of FF dynamics where states I1 and I2 rapidly interconvert with each other, it may well be the case that rather than phosphorylating the U state the kinase phosphorylates RhoGAP-FF1 in the I1 state where Y308 is accessible because H4 is disordered (*SI Appendix, Fig. S7*). The observation that the native state of RhoGAP-FF1 has the same conformation as the I2 folding intermediate of the FF domain from human HYPA/FBP11 suggests that compact, ordered folding intermediates can be repurposed by nature in different, but important functional roles in structurally related molecules. Another interesting take-away that emerges, in this case from the structures of both the I1 and I2 states, is that nonnative helical extensions are found in both minor conformers (helix H3 in both cases), as is also observed in the case of the L99A cavity mutant of T4 lysozyme (23) where helices F and G reposition to form a single long helical structure in an excited state. Albeit only a few examples, this may suggest that intermediates use helix elongation as a mechanism for stabilization and that, therefore, elongated helices may be a feature in other transiently populated minor states as well.

Previous NMR studies have generated structures of single intermediates along reaction pathways (3, 18, 20–26). We show here that CEST is a particularly powerful method for more extensive structural studies, involving pathways with at least two intermediates. In the strategy proposed in this study, we have used peak positions and linewidths in ¹⁵N CEST profiles to advance a folding model, in this case four-state (32), and then a combination of single mutations and/or additives to isolate one of the exchange reactions, for example, $F \rightleftharpoons I2$, so that state I2 can subsequently be studied in detail through a comprehensive set of ¹H, ¹⁵N, and ¹³C CEST experiments. In general, the isolated two-state exchange component, $F \rightleftharpoons X$, of the more complex four-site model, can be identified by comparing CEST-derived X state ¹⁵N chemical shifts with those obtained from fits of ¹⁵N CEST profiles in the four-state exchanging system. Although in this case, the I1 structure was determined previously (3), had this not been the case we could have chosen the A39G FF domain dissolved in 10% TFE buffer, as the exchange reaction reduces to a two state $F \rightleftharpoons I1$ process, with $p_{I1} \sim 9\%$ (1 °C) (32). Since the CEST experiment is not restricted to folding studies, but can be used to explore other biomolecular process, it is likely that the approach we have utilized here to characterize the folding landscape of the FF domain at

atomic resolution will be applicable to other exchanging protein systems as well.

Materials and Methods

The CEST datasets were recorded using previously published sequences (30, 42, 43, 46–49) on protein samples that were isotopically ($^{15}\text{N}/^{13}\text{C}/^2\text{H}$) enriched at the appropriate sites. FF samples with the desired isotopic enrichment patterns were prepared by overexpressing the protein in *E. coli* BL21(DE3) cells grown in suitable M9 media (58, 59). To obtain the exchange parameters and minor-state chemical shifts, CEST datasets were analyzed using the program *ChemEx* (60). CS-ROSETTA (38) was used to calculate structures of A17G FF in the I2 state from the CEST derived ω_{I2} values. All the CEST NMR experiments were performed on a 16.4 T (700 MHz) Bruker Avance III HD spectrometer equipped with a cryogenically cooled triple resonance probe. *SI Appendix* contains details regarding protein expression and purification, NMR experiments, data analysis, and brief discussions about CEST and m -value analyses.

Data, Materials, and Software Availability. The CEST derived chemical shifts of A17G FF in the I2 state (*SI Appendix, Table S3*) and the coordinates of the ten lowest energy A17G FF I2 state structures (A17GFF_I2_lowest10.pdb.gz) are included in the [supporting information](#).

ACKNOWLEDGMENTS. We thank Dr. G Bouvignies (ENS Paris) for providing the program *ChemEx* that was used to analyze the CEST data. We acknowledge

TIFR-Hyderabad NMR facility for an extensive grant of spectrometer time and Dr. K Rao for maintenance of the facility. This study made use of NMRbox: National Centre for Biomolecular NMR Data Processing and Analysis, a Biomedical Technology Research Resource (BTRR), which is supported by NIH grant P41GM111135 (NIGMS). P.V. acknowledges intramural funding from TIFR Hyderabad (DAE, Government of India, Project No. RTI 4007). Y.T. was supported through a Japan Society for the Promotion of Science Overseas Research Fellowship, an Uehara Memorial Foundation postdoctoral fellowship, and a fellowship from the Canadian Institutes of Health Research (CIHR). L.E.K. acknowledges support from the Canadian Institutes of Health research (FND-503573) and the Natural Sciences and Engineering Council of Canada (2015-04347) and D.F.H. thanks the United Kingdom Research and Innovation and Engineering and Physical Sciences Research Council for financial support (EP/X036782/1). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

Author affiliations: ^aTata Institute of Fundamental Research Hyderabad, Ranga Reddy District, Hyderabad 500046, India; ^bDepartment of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada; ^cDepartment of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada; ^dDepartment of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada; ^eCenter for Biosystems Dynamics Research, RIKEN, Kanagawa 230-0045, Japan; ^fDepartment of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, United Kingdom; and ^gThe Francis Crick Institute, London NW1 1AT, United Kingdom

1. T. R. Sosnick, D. Barrick, The folding of single domain proteins—have we reached a consensus? *Curr. Opin. Struct. Biol.* **21**, 12–24 (2011).
2. D. M. Korzhnev *et al.*, Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
3. D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, L. E. Kay, A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
4. D. M. Korzhnev, T. L. Religa, P. Lundstrom, A. R. Fersht, L. E. Kay, The folding pathway of an FF domain: Characterization of an on-pathway intermediate state under folding conditions by (^{15}N), (^{13}C) α and (^{13}C -methyl relaxation dispersion and (^1H)/(^2H)-exchange NMR spectroscopy. *J. Mol. Biol.* **372**, 497–512 (2007).
5. A. Zhuravleva, D. M. Korzhnev, Protein folding by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **100**, 52–77 (2017).
6. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy*. Princ. Pract. (Academic Press, ed. 2, 2006).
7. A. G. Palmer III, C. D. Kroenke, J. P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238 (2001).
8. V. Tugarinov, G. M. Clore, Exchange saturation transfer and associated NMR techniques for studies of protein interactions involving high-molecular-weight systems. *J. Biomol. NMR* **73**, 461–469 (2019).
9. A. Sekhar, L. E. Kay, NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12867–12874 (2013).
10. P. Vallurupalli, A. Sekhar, T. Yuwen, L. E. Kay, Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: A primer. *J. Biomol. NMR* **67**, 243–271 (2017).
11. D. Ban *et al.*, Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J. Magn. Reson.* **221**, 1–4 (2012).
12. A. Rangadurai, E. S. Szymanski, I. J. Kimsey, H. Shi, H. Al-Hashimi, Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance R1 ρ relaxation dispersion. *Prog. Nucl. Magn. Reson. Spectrosc.* **112–113**, 55–102 (2019).
13. P. Vallurupalli, L. E. Kay, Complementarity of ensemble and single-molecule measures of protein motion: A relaxation dispersion NMR study of an enzyme complex. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11910–11915 (2006).
14. J. Milojevic, V. Esposito, R. Das, G. Melacini, Understanding the molecular basis for the inhibition of the Alzheimer's β -peptide oligomerization by human serum albumin using saturation transfer difference and off-resonance relaxation NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 4282–4290 (2007).
15. P. Vallurupalli, V. P. Tiwari, S. Ghosh, A double-resonance CEST experiment to study multistate protein conformational exchange: An application to protein folding. *J. Phys. Chem. Lett.* **10**, 3051–3056 (2019).
16. A. F. Schubert *et al.*, Structure of PINK1 in complex with its substrate ubiquitin. *Nature* **552**, 51–56 (2017). 10.1038/nature24645.
17. T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. *Science* **370**, eabc2754 (2020).
18. D. F. Hansen, P. Vallurupalli, L. E. Kay, Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41**, 113–120 (2008).
19. P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch, L. E. Kay, Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18473–18477 (2007).
20. P. Vallurupalli, D. F. Hansen, L. E. Kay, Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11766–11771 (2008).
21. K. Madhurima, B. Nandi, S. Munshi, A. N. Naganathan, A. Sekhar, Functional regulation of an intrinsically disordered protein via a conformationally excited state. *Sci. Adv.* **9**, eadh4591 (2023).
22. P. Neudecker *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).
23. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
24. D. M. Korzhnev *et al.*, Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: An NMR relaxation dispersion study. *J. Am. Chem. Soc.* **133**, 10974–10982 (2011).
25. P. Kucik *et al.*, Structural characterization of the early events in the nucleation-condensation mechanism in a protein folding process. *J. Am. Chem. Soc.* **139**, 6899–6910 (2017).
26. J. B. Stiller *et al.*, Structure determination of high-energy states in a dynamic protein ensemble. *Nature* **603**, 528–535 (2022).
27. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999), p. 656.
28. S. Forsen, R. A. Hoffman, Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. *J. Chem. Phys.* **39**, 2892–2901 (1963).
29. K. M. Ward, A. H. Aletras, R. S. Balaban, A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J. Magn. Reson.* **143**, 79–87 (2000).
30. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying “invisible” excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161 (2012).
31. V. P. Tiwari, D. De, N. Thapliyal, L. E. Kay, P. Vallurupalli, Beyond slow two-state protein conformational exchange using CEST: Applications to three-state protein interconversion on the millisecond timescale. *J. Biomol. NMR* **78**, 39–60 (2024). 10.1007/s10858-023-00431-6.
32. V. P. Tiwari, Y. Toyama, D. De, L. E. Kay, P. Vallurupalli, The A39G FF domain folds on a volcano-shaped free energy surface via separate pathways. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2115113118 (2021).
33. M. Allen, A. Friedler, O. Schon, M. Bycroft, The structure of an FF domain from human HYPAFBP11. *J. Mol. Biol.* **323**, 411–416 (2002).
34. P. Jemth *et al.*, Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6450–6455 (2004).
35. P. Jemth *et al.*, The structure of the major transition state for folding of an FF domain from experiment and simulation. *J. Mol. Biol.* **350**, 363–378 (2005).
36. P. Jemth, C. M. Johnson, S. Gianni, A. R. Fersht, Demonstration by burst-phase analysis of a robust folding intermediate in the FF domain. *Protein Eng. Des. Sel.* **21**, 207–214 (2008).
37. Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
38. Y. Shen *et al.*, Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4685–4690 (2008).
39. A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9615–9620 (2007).
40. D. S. Wishart *et al.*, CS23D: A web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res.* **36**, W496–W502 (2008).
41. P. Vallurupalli, G. Bouvignies, L. E. Kay, A computational study of the effects of C-13–C-13 scalar couplings on C-13 CEST NMR spectra: Towards studies on a uniformly C-13-labeled protein. *Chembiochem* **14**, 1709–1713 (2013).
42. P. Vallurupalli, L. E. Kay, Probing slow chemical exchange at carbonyl sites in proteins by chemical exchange saturation transfer NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **52**, 4156–4159 (2013).
43. G. Bouvignies, P. Vallurupalli, L. E. Kay, Visualizing side chains of invisible protein conformers by solution NMR. *J. Mol. Biol.* **426**, 763–774 (2014).
44. G. Bouvignies, L. E. Kay, Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J. Phys. Chem. B* **116**, 14311–14317 (2012).
45. I. Bahar, R. Jernigan, K. A. Dill, *Protein Actions: Principles and Modeling* (Garland Science, Taylor & Francis Group, New York, 2017), p. 322.

46. T. Yuwen, L. E. Kay, A new class of CEST experiment based on selecting different magnetization components at the start and end of the CEST relaxation element: An application to (^1H) CEST. *J. Biomol. NMR* **70**, 93–102 (2018).
47. V. P. Tiwari, P. Vallurupalli, A CEST NMR experiment to obtain glycine (^1H) (α) chemical shifts in "invisible" minor states of proteins. *J. Biomol. NMR* **74**, 443–455 (2020).
48. A. Kumar, K. Madhurima, A. N. Naganathan, P. Vallurupalli, A. Sekhar, Probing excited state (^1H) α chemical shifts in intrinsically disordered proteins with a triple resonance-based CEST experiment: Application to a disorder-to-order switch. *Methods* **218**, 198–209 (2023).
49. G. Bouvignies, L. E. Kay, A 2D C-13-CEST experiment for studying slowly exchanging protein systems using methyl probes: An application to protein folding. *J. Biomol. NMR* **53**, 303–310 (2012).
50. M. V. Berjanskii, D. S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).
51. J. C. Hoch *et al.*, Biological magnetic resonance data bank. *Nucleic Acids Res.* **51**, D368–D376 (2023).
52. J. K. Myers, C. N. Pace, J. M. Scholtz, Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148 (1995).
53. A. Amsdr, N. D. Noudeh, L. Liu, T. V. Chalikian, On urea and temperature dependences of m -values. *J. Chem. Phys.* **150**, 215103 (2019).
54. P. Vallurupalli, N. Chakrabarti, R. Pomes, L. E. Kay, Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: An experiment-guided molecular dynamics study. *Chem. Sci.* **7**, 3602–3613 (2016).
55. R. Bonet, L. Ruiz, E. Aragon, P. Martin-Malpartida, M. J. Macias, NMR structural studies on human p190-A RhoGAPFF1 revealed that domain phosphorylation by the PDGF-receptor α requires its previous unfolding. *J. Mol. Biol.* **389**, 230–237 (2009).
56. W. Jiang *et al.*, An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression. *Mol. Cell* **17**, 23–35 (2005).
57. D. M. Korzhnev, Loss of structure-gain of function. *J. Mol. Biol.* **425**, 17–18 (2013).
58. P. Lundstrom, P. Vallurupalli, D. F. Hansen, L. E. Kay, Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nat. Protoc.* **4**, 1641–1648 (2009).
59. N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, L. E. Kay, A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J. Biomol. NMR* **13**, 369–374 (1999).
60. G. Bouvignies, Chemex (<https://github.com/gbouvignies/chemex/releases>) (2012).