The A39G FF domain folds on a volcano-shaped free energy surface via separate pathways

Ved P. Tiwari[a,1], Yuki Toyama[b,c,d,1], Debajyoti De[a], Lewis E. Kay[b,c,d,e,2], and Pramodh Vallurupalli[a,2]

[a]Tata Institute of Fundamental Research, Hyderabad 500107, India; [b]Department of Biochemistry, University of Toronto, ON M5S 1A8, Canada; [c]Department of Molecular Genetics, University of Toronto, ON M5S 1A8, Canada; [d]Department of Chemistry, University of Toronto, ON M5S 3H6, Canada; and [e]Program in Molecular Medicine, The Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada

Edited by H. Jane Dyson, Scripps Research Institute, La Jolla, CA, and accepted by the Editorial Board October 1, 2021 (received for review August 23, 2021)

Conformational dynamics play critical roles in protein folding, misfolding, function, misfunction, and aggregation. While detecting and studying the different conformational states populated by protein molecules on their free energy surfaces (FESs) remain a challenge, NMR spectroscopy has emerged as an invaluable experimental tool to explore the FES of a protein, as conformational dynamics can be probed at atomic resolution over a wide range of timescales. Here, we use chemical exchange saturation transfer (CEST) to detect "invisible" minor states on the energy landscape of the A39G mutant FF domain that exhibited "two-state" folding kinetics in traditional experiments. Although CEST has mostly been limited to studies of processes with rates between ∼5 to 300 s^{-1} involving sparse states with populations as low as ∼1%, we show that the line broadening that is often associated with minor state dips in CEST profiles can be exploited to inform on additional conformers, with lifetimes an order of magnitude shorter and populations close to 10-fold smaller than what typically is characterized. Our analysis of CEST profiles that exploits the minor state linewidths of the 71-residue A39G FF domain establishes a folding mechanism that can be described in terms of a four-state exchange process between interconverting states spanning over two orders of magnitude in timescale from ∼100 to ∼15,000 μs. A similar folding scheme is established for the wild-type domain as well. The study shows that the folding of this small domain proceeds through a pair of sparse, partially structured intermediates via two discrete pathways on a volcano-shaped FES.

protein folding intermediates | chemical exchange saturation transfer NMR | FF domain | multiple folding pathways | multistate exchange

The free energy surface (FES) of a molecule dictates the relative energies as well as the interconversion kinetics of the conformers that populate it. A quantitative description of molecular structure and dynamics is predicated, therefore, on a detailed understanding of the FES. In the case of biomolecules, such as proteins, the conformational dynamics that result in the exchange between different molecular states are critical for proper function and can lead to misfunction in some cases (1–5). In principle, computational methods can be used to determine the FESs for model systems, although more accurate force fields and faster simulation methods are needed for quantitative descriptions. While it remains intractable to experimentally map the FES for complex biomolecules, such as proteins, insights into the shape of the FES can be obtained by detecting the different conformational states that are sampled by the protein under equilibrium and nonequilibrium conditions and by measuring the rates at which these states interconvert with one another (6–10). For example, determining folding pathways by detecting intermediates provides insights into how a protein folds rapidly by sampling only a subset of all possible conformations. These experimentally detected states and rate constants can be combined with computational techniques to obtain a detailed picture of the FES (11, 12). However, this strategy remains challenging because traditional biophysical techniques cannot detect minor states that are sparsely and fleetingly populated and while modern fluorescence-based methods are indeed sensitive to the presence of minor states (13, 14), they provide little direct atomic resolution structural information. NMR spectroscopy can be used to study the dynamics at virtually all backbone and sidechain positions in a protein, and relaxation-based NMR experiments are sensitive to minor conformational states with lifetimes on the order of ∼0.01 to ∼100 ms and populations as low as ∼1% (15–18). Thus, NMR has emerged as a powerful tool to study conformational dynamics of proteins at atomic resolution (7, 15, 19–23). The transiently populated minor states that have become amenable to detailed study by NMR are termed "invisible" states, as peaks arising from these conformers are not directly observed in conventional NMR spectra. Yet, as has been established in the past several decades, they can be detected by manipulating the NMR signals associated with the visible major state that they are interconverting with. A number of different NMR experiments have emerged for studies of invisible states. These include Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion, chemical exchange and dark-state exchange saturation transfer (CEST and DEST, respectively), and $R_{1\rho}$ experiments, with each technique sensitive to a different kinetic exchange regime (15, 16, 20, 24–27). These approaches have been used to obtain kinetic and thermodynamic parameters governing the exchange process(es) as well as structural information on the

Significance

Protein molecules interconvert between multiple states, and this exchange process plays an important role in function and misfunction. Here, we use powerful NMR experiments that probe sparsely populated, transiently formed conformational states, which exchange with a highly populated, stable, folded conformer in order to elucidate the folding mechanism of a small protein domain. Notably, by analyzing the exchange profiles in detail, it becomes possible to obtain structural data for states that are about 10-fold less populated and less long lived than has previously been possible. Our results establish that the small FF domain folds via two partially structured intermediates and involves a pair of discrete pathways in which the unfolded state becomes increasingly structured along a volcano-shaped free energy surface.

Author contributions: V.P.T., Y.T., D.D., L.E.K., and P.V. performed research; V.P.T., Y.T., D.D., L.E.K., and P.V. analyzed data; and L.E.K and P.V. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. H.J.D. is a guest editor invited by the Editorial Board.

[1]V.P.T. and Y.T. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: kay@pound.med.utoronto.ca or pramodh@tifrh.res.in.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115113118/-/DCSupplemental.

Published November 10, 2021.

interconverting conformers (28–33) and, in some cases, it has been possible to obtain the structures of the invisible minor states (21, 34–37). Notably, the methodology is equally applicable to studies of nucleic acids, in which strong correlations between excited conformational states and function have been clearly established (38, 39). In the context of CEST that will be used in the present study and the related DEST technique, the methodology has been exploited to elucidate reaction mechanisms in small molecules (40) and to study the folding (41, 42), aggregation (43, 44), and other conformational exchange processes in large biomolecules, such as proteins (36, 45–47) and nucleic acids (16, 38, 48). CEST has also been used to study exchange occurring between a major state and multiple invisible minor states that interconvert with the major state directly or indirectly (36, 41, 42).

One area of focus in our laboratories has involved NMR relaxation–based studies of protein folding intermediates and compact alternate protein conformers that differ from the major state conformation (12, 37, 41, 49–51). Studies of the folding mechanisms of a variety of small proteins using different biophysical techniques have provided an understanding of the general principles that govern protein folding (1, 7, 9, 10, 52, 53). One such system is the 71-residue FF domain from human HYPA/FBP11, whose folding mechanism has been studied extensively by using a variety of techniques so that it has emerged as a model system to understand protein folding (8, 37, 54–56). Under native conditions, the protein adopts a four-helix bundle structure that consists of helices H1, H2, H3, H4, and a disordered N-terminal tail (57). Initial folding studies suggested that the wild-type FF domain (WT FF) folds via a compact, transiently populated intermediate (54, 55), and subsequent CPMG relaxation dispersion NMR experiments were used to derive the structure of this invisible state (37). The structure showed that the folding intermediate is stabilized by nonnative interactions which slow down the folding processes (37). In a previous study of folding using burst phase analysis, a folding intermediate was detected for 32 out of 33 FF mutants (56). Notably, an intermediate could not be detected for the folding of A39G FF (56), although a folding rate significantly slower (~135 s^{-1}, 10 °C) than WT FF (~2,200 s^{-1}, 10 °C) was measured (55), and little insight into the folding pathway of this variant is currently available.

Here, we use ^{15}N-based CEST experiments (58) to elucidate the folding mechanism of A39G FF by taking advantage of the information contained in the minor state linewidths in CEST spectra that provides increased sensitivity to invisible states with populations and lifetimes an order of magnitude smaller than those typically analyzed using this methodology. We show that the A39G FF domain folds through two intermediates via a branched pathway on a volcano-shaped energy landscape, in which the energies of minor, partially folded states increase prior to collapse to the folded conformation and that other FF variants, including the WT FF domain, fold via similar intermediates. Our study further highlights the folding complexities of even single-domain proteins and the unique role that CEST can play in describing increasingly complicated FESs of biomolecules.

Results

Using Sparse States to Detect Other Sparse States via CEST. As mentioned in the introduction, we have made use of the linewidths of the minor state dips to increase the information content of CEST profiles. (See the *SI Appendix, A Brief Description of CEST* for an overview of the CEST experiment.) In order to illustrate the approach schematically consider the CEST profiles in Fig. 1. Fig. 1*A* shows a calculated CEST profile of an ^{15}N spin that results from a two-site exchange process, A ↔ B, with $k_{ex,AB} = k_{AB} + k_{BA} = 25$ s^{-1} (k_{ij} is the rate constant for the reaction from state i to state j) and the fractional population of the minor state $p_B = 2\%$ ($p_B = k_{AB}/k_{ex,AB}$). As states A and B differ in structure, the ^{15}N spin probe in question will have distinct chemical shifts in each state, here taken to be 0 and 7 ppm for states A and B, respectively. In this case, the CEST profile comprises a pair of dips (or peaks) centered at the resonance frequencies of the ^{15}N spin in each state and indicated by the black and green vertical lines in Fig. 1. Notably, the peak from the excited (B) state would not be observed or would be very weak in conventional spectra. In Fig. 1*B*, a CEST profile for a three-state exchange process, A ↔ B ↔ C, $k_{ex,AB} = 25$ s^{-1}, $k_{ex,BC} = 200$ s^{-1}, $p_B = 2\%$, and $p_C = 0.2\%$ is shown. The chemical shift of state C (ϖ_C) is set to 3 ppm. Three dips are observed at the chemical shifts of the ^{15}N nucleus in question in each of the three states A, B, and C (blue vertical line). When $k_{ex,BC}$ is increased to 4,950 s^{-1}, corresponding to the

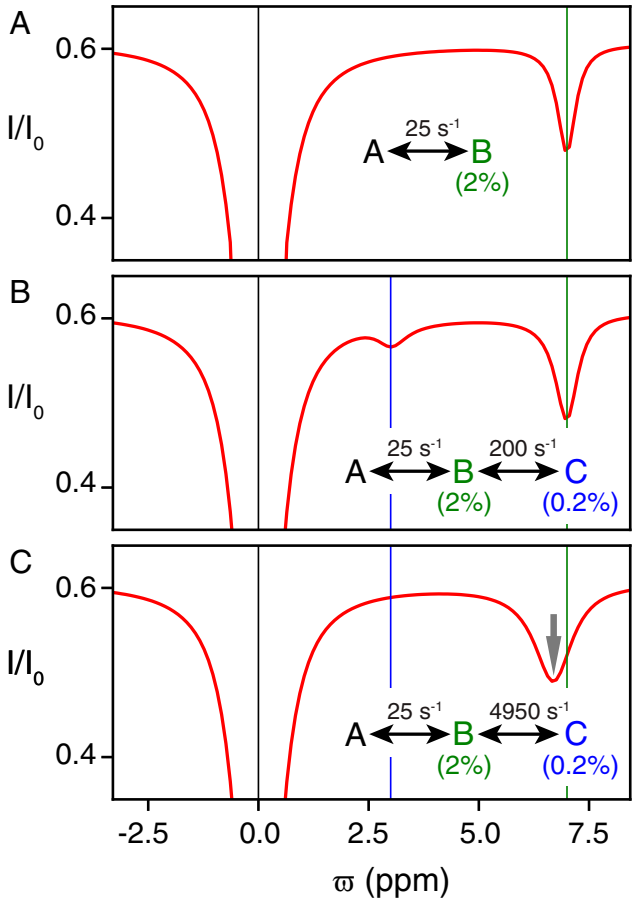

Fig. 1. ^{15}N CEST profiles are extremely sensitive to the underlying exchange process(es) that interconvert states. ^{15}N CEST profiles have been calculated for two- (*A*) and three-state (*B* and *C*) exchange using the parameters listed with $B_1 = 20$ Hz, an exchange time $T_{EX} = 0.5$ s, a static magnetic field of 23.4 T, and chemical shifts of an ^{15}N spin in states A to C of $\varpi_A = 0$ ppm, $\varpi_B = 7$ ppm, and $\varpi_C = 3$ ppm. I and I_0 are the intensities of peaks simulated with and without the T_{EX} periods, respectively (58). Longitudinal and transverse relaxation rates of 1 and 10 s^{-1} were assumed for all sites. Black, green, and blue vertical lines denote ϖ_A, ϖ_B, and ϖ_C, respectively. Because of rapid exchange between states B and C, the minor ("B") state dip indicated by the gray arrow in *C* is shifted away from ϖ_B (green line) toward ϖ_C (blue line). The ↔ symbol is used to denote equilibrium in all figures, and the number above (or below) it with units of s^{-1} (second^{-1}) is the exchange rate (k_{ex}), that is, the sum of the forward and reverse reaction rate constants. See *SI Appendix, A Brief Description of CEST* for an overview of the CEST experiment.

onset of intermediate-fast exchange ($|k_{ex,BC}/\Delta\omega_{BC}|$ ~1.95 at 23.4 T field and $\Delta\omega_{BC} = \omega_C - \omega_B$, where ω_i is the angular resonance frequency [radians/second] of the ^{15}N spin in state i), the dip from state C broadens out and disappears, while the dip derived from B broadens and moves toward the resonance frequency of the ^{15}N spin in state C (Fig. 1*C*), as expected (59–63). Thus, minor state dip positions and widths in ^{15}N CEST profiles are sensitive to the presence of other even more sparsely populated states (C in this example), suggesting that dips from state B, that are themselves only visible in CEST profiles, can be used to "spy" on other even more sparsely populated states that are invisible to CEST. Such exchange among minor states has been challenging to study previously (20, 64). As the position and linewidth of the B state dip that is affected by exchange with C will now depend on the static magnetic field, an analysis of CEST data recorded at different spectrometer field strengths is useful in studying multistate exchange (59–63, 65). Note that in all the experiments described here only magnetization from the "visible" major state (A in this case) is detected.

A39G FF Does Not Fold in a "Two-State" Manner. As described earlier, WT FF and several of its mutants have all been shown to fold via intermediates, while the slow folding A39G FF variant appears to fold in a "two-state" manner based on conventional stopped-flow experiments (56). Interestingly, the minor unfolded (U) state dip of A39G FF is broadened in ^{15}N CEST profiles (58, 66), much like the state B peak in Fig. 1*C*, suggesting that the U state is exchanging with other minor states, including possible folding intermediates. To establish whether this is the case, we recorded four ^{15}N CEST datasets with a number of different weak B_1 fields to search for invisible conformers: B_1 = 15.9 and 29.1 Hz at 600 MHz and B_1 = 8.3 and 27.7 Hz at 1 GHz (*SI Appendix*, Table S1). The amide ^{1}H^N–^{15}N correlation map (1 °C and 1 GHz; Fig. 2*A*) of A39G FF is well resolved so that conformational exchange can be studied at almost all backbone amide ^{15}N sites in the domain. ^{15}N CEST profiles recorded with B_1 = 27.7 Hz at 1 GHz (Fig. 2*B*) clearly show a pair of dips, with the main dip derived from the visible, major conformational state of the domain (red numbers in Fig. 2*A*). Notably, the chemical shifts of the minor state correspond

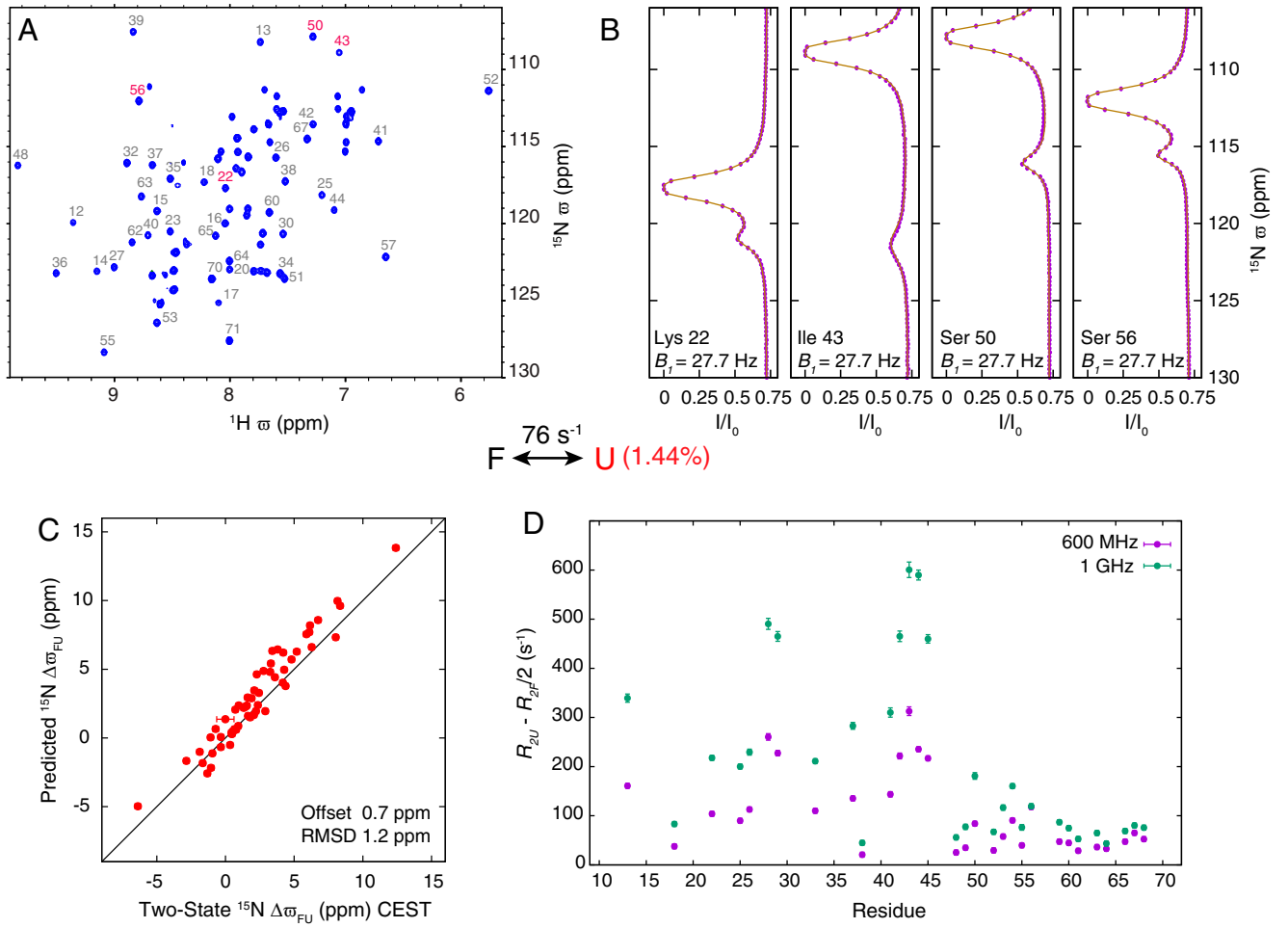

Fig. 2. The sparse, unfolded state of A39G FF exchanges with other even more rare conformers. (*A*) A39G FF ^{15}N–^{1}H Heteronuclear Single Quantum Coherence (HSQC) spectrum (1 GHz and 1 °C) with resonances labeled according to the residues from which they derive. (*B*) ^{15}N CEST profiles recorded at 1 GHz and 1 °C with B_1 = 27.7 Hz clearly show a minor state dip in addition to the major state dip. Residues for which CEST profiles are shown in *B* are labeled in red in *A*. Experimental data points are shown as magenta circles, and the brown line is calculated from the global two-state best fit parameters, with R_{2U} allowed to vary from site to site (*SI Appendix*, Table S2). (*C*) Comparison between ^{15}N $\Delta\varpi_{FU}$ values from global two-state fits of CEST profiles with chemical shift differences obtained using the unfolded state chemical shifts of A39G FF predicted by the program POTENCI (67) and the measured shifts from the folded state. Offset refers to the average difference ($Q_y - Q_x$) between the quantities plotted on the y (Q_y) and x (Q_x) axis. (*D*) R_{exU} ($R_{2U} - R_{2F}/2$) values for the (minor) unfolded state obtained from fits of CEST dips in profiles recorded at 600 MHz (magenta) and 1 GHz (green) for residues with $|\Delta\varpi_{FU}| > 2$ ppm show that many dips from the minor state are significantly broadened. A k_{ex} value of 76 s^{-1} (sum of forward and reverse rates) is obtained from the best fit of the CEST data to a two-state exchange model, as indicated above panels *C* and *D*.

to those of an unfolded protein (referred to in what follows as state U), as a strong correlation is observed between the difference of the U and folded (F) state chemical shifts $\Delta\varpi_{FU} = \varpi_U - \varpi_F$ values (parts per million), obtained from two-state fits of the CEST profiles and shift differences calculated using ^{15}N random coil values predicted with the program POTENCI (Fig. 2*C*) (67). Although we expect the U state intrinsic R_2 values (R_{2U}) to be approximately half of the F state intrinsic R_2 values (R_{2F}) (68), an exchange model in which R_{2U} was constrained to be $R_{2F}/2$ did not fit the ^{15}N CEST data well (χ^2_{red} ~5.3, *SI Appendix*, Table S2). The fits improved considerably (χ^2_{red} ~1.43, *SI Appendix*, Table S2) when R_{2U} was not constrained, with fitted rates that varied significantly from site to site. This large site-to-site variation in R_{2U} rates can be seen in the ^{15}N CEST profiles, in which it is clear that the minor state dip of Ile-43 is broader than dips from Lys-22, Ser-50, and Ser-56 (Fig. 2*B*). Notably, the exchange-induced broadening of the U state, R_{exU}, approximated by $R_{2U} - R_{2F}/2$, increases with static magnetic field (Fig. 2*D*) with R_{exU} rates as large as ~600 s^{-1} at 1 GHz, suggesting an exchange mechanism that is more complex than two state. It is interesting to note that although the predicted and CEST-derived $\Delta\varpi_{FU}$ values are in good agreement (1.2 ppm RMSD; Fig. 2*C*), the predicted U state shifts are larger on average than those measured (offset) by ~0.7 ppm (Fig. 2*C*).

The ^{15}N CEST data were subsequently analyzed using a linear three-state folding model F↔I↔U, in which folding proceeds via an intermediate I (*SI Appendix*, Fig. S1). Here, we included three additional ^{15}N CEST datasets recorded with higher B_1 values at 1 GHz (64.9, 159.6, and 224.5 Hz; *SI Appendix*, Table S1). The linear three-state model with constraints $R_{2U} = R_{2F}/2$ and $R_{2I} = R_{2F}$ did not fit the ^{15}N CEST data well (χ^2_{red} = 2.44; *SI Appendix*, Table S2), but the fits improved considerably (χ^2_{red} = 1.27; *SI Appendix*, Fig. S1*A* and Table S2) when R_{2U} was unconstrained and $R_{2I} = R_{2F}$. As in the case of the two-state analysis, the unfolded state chemical shifts obtained from these fits are in good agreement with the predicted values (1.2 ppm RMSD; *SI Appendix*, Fig. S1*B*), yet there still remains a 0.6 ppm offset.

Previous studies using CPMG experiments recorded at 30 °C (37) established that WT FF folds via an intermediate; however, further states could not be observed via the CPMG class of experiments. The ^{15}N chemical shifts of the WT FF intermediate are in reasonable agreement with the I state chemical shifts obtained here for A39G FF (RMSD 1.8 ppm; *SI Appendix*, Fig. S1*C*), suggesting that WT and A39G FF domains fold via similar intermediates. Yet the fitted R_{2U} rates are still large (*SI Appendix*, Fig. S1*D*), with some values exceeding ~300 s^{-1} (at 1 GHz), indicating that there is likely an additional exchange mechanism at play that is not captured by the three-state model. Notably, similar poor-quality fits, as those obtained using a linear folding scheme, were also obtained (*SI Appendix*, Table S2) with a more complex triangular model and the quality of the fits improved (χ^2_{red} ~1) only when R_{2U} was not constrained. The large, fitted R_{2U} values are consistent with more complex dynamics.

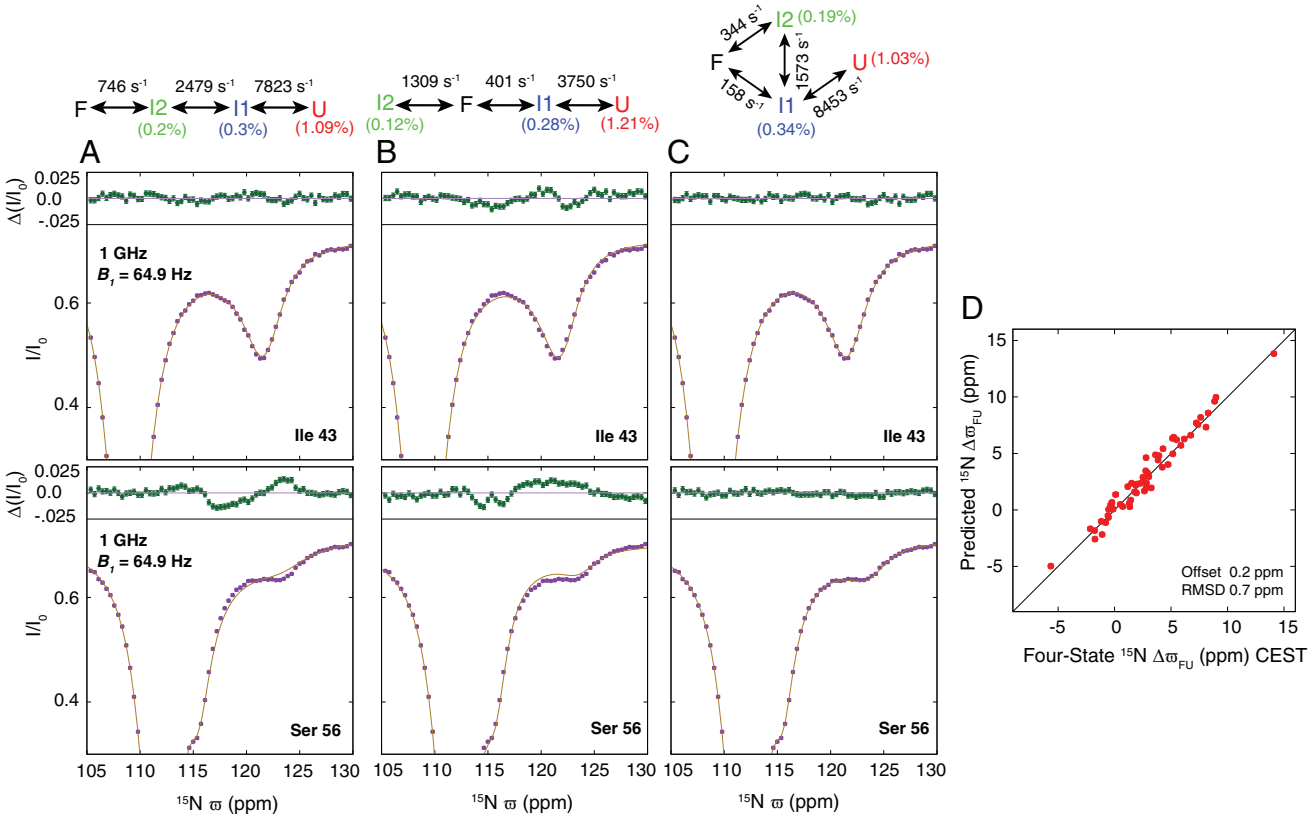

Fig. 3. A two-pathway, four-state folding model reproduces A39G FF CEST data. (*A–C*) ^{15}N CEST profiles of Ile-43 and Ser-56 (B_1 = 64.9 Hz and 1 GHz) fitted to different four-state models as shown, along with extracted k_{ex} values and populations. Experimental data are represented using magenta points, and the brown line was calculated using the appropriate best fit parameters for models shown on top of the panels. The best fit parameters were obtained from a global analysis of 600 MHz and 1 GHz ^{15}N CEST profiles from 14 residues of A39G FF (*SI Appendix*, Table S2). The CEST data cannot be satisfied by the linear models (*A* and *B*) but is well reproduced using a model in which both I1 and I2 exchange with F (*C*). Above each of the CEST profiles is plotted the difference between experimental and fitted I/I_0 values [$\Delta(I/I_0)$]. (*D*) Comparison between $\Delta\varpi_{FU}$ values obtained using the four-state model shown in *C* with those calculated from predicted chemical shifts of amides in U by POTENCI (67).

A39G FF Folds via Two Intermediates along Two Pathways. Having established that neither two- nor three-state folding models could reproduce the experimental CEST data using reasonable R_{2U} values, we next considered different linear four-state exchange schemes with a pair of intermediates, referred to as I1 and I2 in what follows, subject to the constraint that $R_{2I1} = R_{2I2} = R_{2F}$ and $R_{2U} = R_{2F}/2$. Two of these linear exchange schemes are shown in Fig. 3, and although the F ↔ I2 ↔ I1 ↔ U model fit the data better (Fig. 3*A*) than the I2 ↔ F ↔ I1 ↔ U scheme (Fig. 3*B*) (χ^2_{red} ~ 1.1 versus 1.63; *SI Appendix*, Table S2) and was superior to both two- and three-state models where $R_{2U} = R_{2F}/2$, the linear four-state fits were still not satisfactory, as illustrated in Fig. 3 *A* and *B* for profiles derived from both Ile-43 and Ser-56. However, a four-state model in which all states were allowed to exchange with each other and where $R_{2I1} = R_{2I2} = R_{2F}$ and $R_{2U} = R_{2F}/2$ fit the data well with values of $k_{ex,FU}$ and $k_{ex,I2U}$ ~0 s^{-1} (*SI Appendix*, Fig. S2*A*), resulting in the simplified four-state model of Fig. 3*C*. This model reproduced the CEST data better than previous constrained models described earlier in this section (Fig. 3*C*, χ^2_{red} ~0.9; *SI Appendix*, Table S2) and, importantly, allowing R_{2U} to float during the fits did not improve their quality (χ^2_{red} ~0.86; *SI Appendix*, Table S2), in contrast to what was observed for the two- and three-state models (*SI Appendix*, Table S2). More complex folding pathways were, therefore, not considered. In the four-state exchange model of Fig. 3*C*, the U state rapidly interconverts with I1 ($k_{ex,I1U}$ ~8,453 s^{-1}) and then folds to F either directly from I1 or via I2 with exchange rates $k_{ex,FI1} = 158 \pm 20$ s^{-1}, $k_{ex,FI2} = 344 \pm 20$ s^{-1}, $k_{ex,I1I2} = 1{,}573 \pm 50$ s^{-1}, and $k_{ex,I1U} = 8{,}453 \pm 400$ s^{-1} and fractional populations p_{I1}= 0.34 ± 0.05%, $p_{I2} = 0.19 \pm 0.01\%$, and $p_U = 1.03 \pm 0.05\%$. Thus, despite the small size of the FF domain, there appears to be more than one folding pathway.

Validation of the Four-State Folding Pathway. A simple test of the four-state model can be made by comparing the four-state $\Delta\varpi_{FU}$ values with those predicted, and indeed, the agreement is significantly improved relative to what was obtained for both the two- and three-state schemes considered (RMSD 0.7 ppm, Fig. 3*D*, versus 1.2 ppm, Fig. 2*C* and *SI Appendix*, Fig. S1*B*). Furthermore, the offset between the CEST-derived and -predicted $\Delta\varpi_{FU}$ values has essentially vanished (0.2 ppm, Fig. 3*D*, versus 0.7 ppm, Fig. 2*C* and 0.6 ppm *SI Appendix*, Fig. S1*B*). In order to establish that the I1 and I2 state chemical shifts are meaningful, we attempted to "simplify" the four-state exchange equilibrium by manipulating the populations of the different minor states through the addition of chemicals or by point mutations (Fig. 4 *A*–*E*). As a point of reference, Fig. 4*A* shows four-state fits of ^{15}N CEST data recorded at 1 GHz with $B_1 = 64.9$ Hz for Lys-22, Ile-43, Ser-50, and Ser-56 along with the ^{15}N chemical shifts for these residues in F, I1, I2, and U, indicated using black, blue, green, and red vertical lines, respectively. Note that distinct dips for the I1 state were not observed in any of the ^{15}N CEST profiles, and only Ser-56 showed an I2 state dip (Figs. 3*C* and 4*A*). The addition of 1 M urea, that is known to stabilize unfolded structure (1), increased p_U (Fig. 4*B*) so that ^{15}N CEST data (18.3 and 33.9 Hz, 700 MHz) could be well fit to a two-state F↔U exchange model ($k_{ex,FU} = 17.9 \pm 2$ s^{-1} and $p_U = 7.4 \pm 0.4\%$). The resulting $\Delta\varpi_{FU}$ values are in very good agreement with F, U shift differences obtained both from the four-state fits (RMSD 0.4 ppm; Fig. 4*F*), and from predicted unfolded state shifts (RMSD 0.6 ppm; *SI Appendix*, Fig. S3*A*). Next, we added 10% trifluro ethanol (TFE) to the buffer. TFE can stabilize or destabilize native protein states depending on the system and, additionally, may stabilize folding intermediates (69). Notably, the addition of TFE increased the population of the I1 state (Fig. 4*C*) so that ^{15}N CEST data (10.4 and 18.2 Hz, 700 MHz) could be fit to a two-state exchange model ($k_{ex,FI1} = 95.5 \pm 4$ s^{-1} and $p_{I1} = 8.6 \pm 0.1\%$), with the resulting minor state chemical shifts in agreement with those obtained for the I1 state from the four-state model (RMSD 0.9 ppm; Fig. 4*G*) and with those previously obtained for the folding intermediate of WT FF, based on the analysis of CPMG experiments (37) recorded at 30 °C (RMSD 1 ppm; *SI Appendix*, Fig. S3*B*). Thus, A39G FF folds via an intermediate similar to the I state of WT FF, supporting the four-state analysis. The underlying mechanism for the relative stabilization of I1 due to TFE is currently not understood. We also recorded ^{15}N CEST profiles ($B_1 = 10.3$ and 25.8 Hz, 700 MHz, 5 °C; Fig. 4*D*) of A17G FF (dissolved in buffer containing 20% glycerol), in which the population of the I2 state was elevated. Fits of the data to a two-site exchange model ($k_{ex,FI2} = 104.6 \pm 10$ s^{-1} and $p_{I2} = 1.0 \pm 0.1\%$) produced intermediate chemical shifts in good agreement with 1) those from the four-state model (RMSD 1 ppm; Fig. 4*H*) and 2) from analysis of CPMG experiments recorded on A17G FF at 25 °C (8) (RMSD 0.7 ppm; *SI Appendix*, Fig. S3*C*). It is worth emphasizing that none of the three systems of Fig. 4 *B*–*D* are strictly two-state since significant exchange contributions to the minor state dips were observed. Nonetheless, these systems provide a means to test the predictions of the four-state model (Fig. 4 *F*–*H*).

Fits using the four-state exchange model established that $p_U \sim 3p_{I1}$, with U and I1 in rapid exchange (Fig. 4*A*). Thus, states I1 and U should give rise to a single dip in ^{15}N CEST profiles at a position close to ϖ_U and shifted slightly toward ϖ_{I1} (60–63). This is clearly apparent for Ile-43 and Ser-50 in Fig. 4*A*, in which the minor dip is in between the blue and red vertical lines denoting ϖ_{I1} and ϖ_U, respectively. As p_U is reduced upon addition of 10% TFE, the averaged I1-U dip should appear closer to ϖ_{I1} (blue line), as observed for Ile-43 and Ser-50 (compare Fig. 4 *A* and *C*). Note that the fact that the minor dip does not superimpose with ϖ_{I1} for Ile-43 (Fig. 4*C*) suggests that $p_U \neq 0$, even with 10% TFE, and that a two-site model of exchange is an oversimplification for A39G FF + 10% TFE, as discussed earlier in this section.

Glycerol is known to stabilize proteins (1), and indeed, upon addition of 25% glycerol to the A39G FF sample (Fig. 4*E*), p_U was reduced significantly, and dips at ϖ_U were not observed in the ^{15}N CEST profiles (compare Fig. 4 *B* and *E*), while dips at ϖ_{I1} were observed (compare Fig. 4 *C* and *E*). Notably, however, the broad I1 state dip for Ile-43 is shifted toward U (Fig. 4*E*), suggesting that a small residual amount of the U state is present, even with 25% glycerol. The ϖ_{I1} and ϖ_{I2} chemical shifts obtained from fits to a three-state exchange process ($k_{ex,FI1} = 145.5 \pm 10$ s^{-1}, $k_{ex,FI2} = 330.3 \pm 15$ s^{-1}, $k_{ex,I1I2} = 1{,}237.2 \pm 50$ s^{-1}, $p_{I1} = 0.82 \pm 0.02\%$, and $p_{I2} = 0.66 \pm 0.02\%$; *SI Appendix*, Table S3) are in good agreement with those obtained from the four-state analysis (Fig. 4 *I* and *J*).

As a final test of the four-state analysis, we compared $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FU}$[two-state] and $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FI1}$[10% TFE], where $\Delta\varpi_{FU}$[1 M urea] (Fig. 4*B*), $\Delta\varpi_{FU}$[two-state] (Fig. 2*C*), and $\Delta\varpi_{FI1}$[10% TFE] (Fig. 4*C*) are obtained from two-state fits of A39G FF CEST profiles recorded on samples with and without urea, and with 10% TFE, respectively (Fig. 4*K*). For fast exchange on the NMR chemical shift timescale between U and I1 with $p_U > p_{I1}$, the averaged U-I1 dip position, ϖ_{avg}, will be shifted by an amount $\frac{p_{I1}}{p_{I1}+p_U}(\varpi_{I1} - \varpi_U)$ from the pure U state position towards I1 so that $\varpi_U - \varpi_{avg} = \frac{p_{I1}}{p_{I1}+p_U}(\varpi_U - \varpi_{I1})$. As ϖ_U[two-state] = ϖ_{avg}, the difference, $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FU}$[two-state], is, to an excellent approximation, given by $\frac{p_{I1}}{p_{I1}+p_U}(\varpi_U - \varpi_{I1})$. Similarly, $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FI1}$[10% TFE] ~ $(\varpi_U - \varpi_{I1})$, assuming that only states F and I1 are populated when 10% TFE is added. Thus, the slope of the linear correlation in Fig. 4*K* should be $\frac{p_{I1}}{p_{I1}+p_U}$, which is calculated to be 0.25 ± 0.02 based on the parameters extracted from fits of the four-state model to the CEST data. This agrees reasonably well with a slope of 0.32 ± 0.01 that is measured from

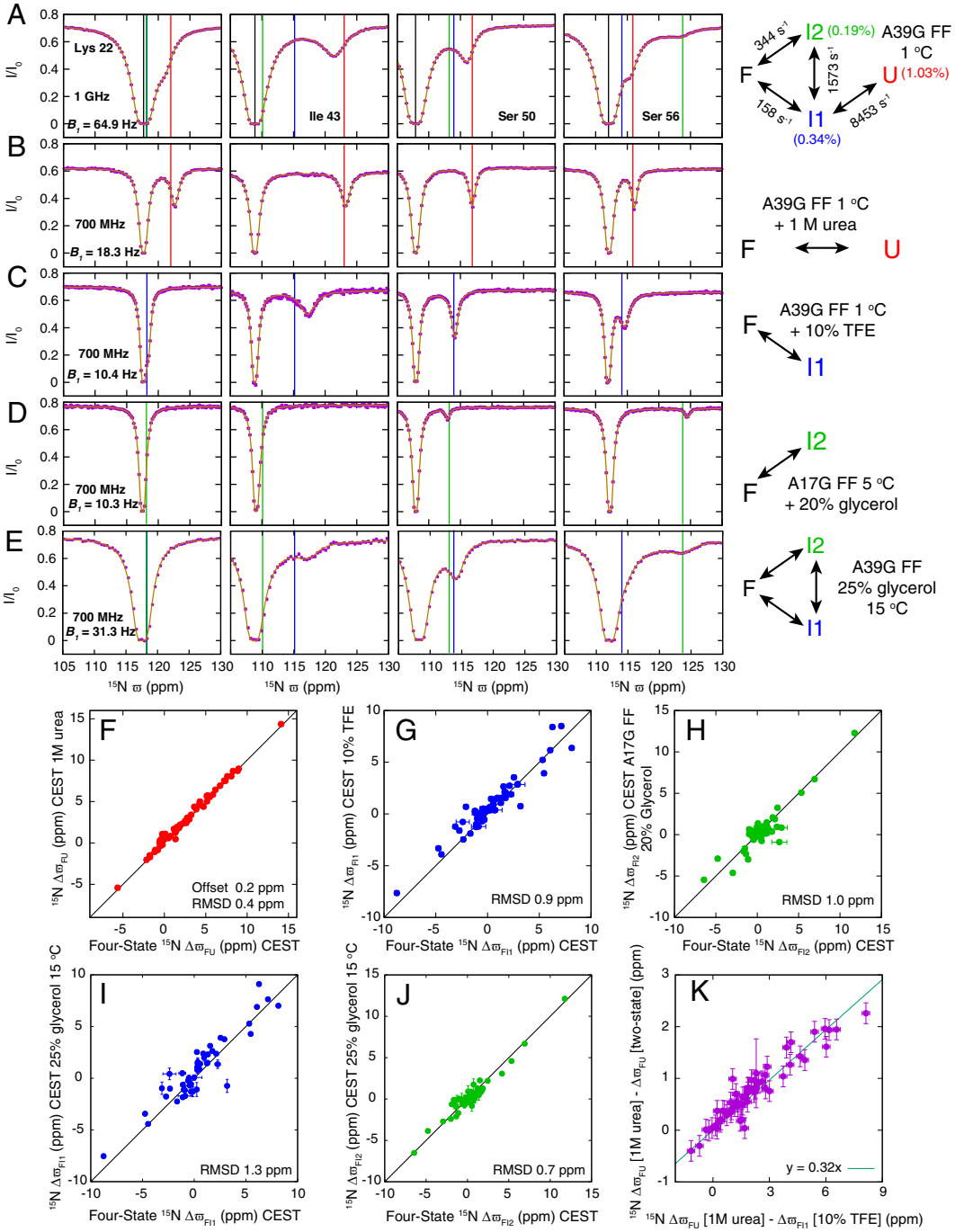

Fig. 4. Fits of the CEST data to the four-state exchange model are meaningful. (*A*) ^{15}N CEST profiles recorded for Lys-22, Ile-43, Ser-50, and Ser-56 with B_1 = 64.9 Hz, 1 GHz, and 1 °C. The four-state model is shown on the right, along with extracted exchange k_{ex} rates and populations. In *A*–*E*, the chemical shifts of the minor states obtained from the four-state analysis are shown using blue, green, and red lines for I1, I2, and U, respectively, while the position of the major native state is indicated with a black line in *A*. (*B*) A39G FF ^{15}N CEST profiles (B_1 = 18.3 Hz, 700 MHz, and 1 °C) recorded in the presence of 1 M urea and analyzed using the two-state exchange model shown on the right. Minor state dips appear close to chemical shifts predicted for U by the four-state analysis (red line). (*C*) A39G FF ^{15}N CEST profiles (B_1 = 10.4 Hz, 700 MHz, and 1 °C) recorded with 10% TFE and analyzed using the two-state model shown on the right. Minor state dips close to the I1 state chemical shift predicted by the four-state analysis are observed (blue line). (*D*) ^{15}N CEST profiles (B_1 = 10.3 Hz, 700 MHz, and 5 °C) for A17G FF + 20% glycerol analyzed using a two-state model. Minor state dips proximal to I2 state chemical shifts predicted by the four-state analysis (green line) are observed. (*E*) A39G FF ^{15}N CEST profiles recorded at 15 °C in the presence of 25% glycerol, analyzed using the three-state model, shown on the right. A dip at I1 is now visible close to the I1 state chemical shift predicted by the four-state analysis (blue line). In *A*–*E*, the experimental data points are shown in magenta, with the brown lines calculated using the appropriate global best fit parameters. (*F*–*H*) Comparison of A39G FF $\Delta\varpi_{FU}$ (*F*), $\Delta\varpi_{FI1}$ (*G*), or $\Delta\varpi_{FI2}$ (*H*) values obtained using the four-state analysis (*A*) with the corresponding $\Delta\varpi_{FU}$ (*F*), $\Delta\varpi_{FI1}$ (*G*), or $\Delta\varpi_{FI2}$ (*H*) values fitted from the two-state analysis of CEST profiles recorded using A39G FF in the presence of 1 M urea (*B*), 10% TFE (*C*), or A17G FF in the presence of 20% glycerol (*D*). (*I* and *J*) Comparison of the $\Delta\varpi_{FI1}$ (*I*) or $\Delta\varpi_{FI2}$ (*J*) values obtained using the four-state analysis (*A*) with $\Delta\varpi_{FI1}$ or $\Delta\varpi_{FI2}$ values from three-state fits of A39G FF CEST data recorded in presence of 25% glycerol (*E*). (*K*) Correlation plot of $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FU}$[two-state] versus $\Delta\varpi_{FU}$[1 M urea] − $\Delta\varpi_{FI1}$[10% TFE]; all $\Delta\varpi$ values used in (*K*) were obtained from two-state fits of CEST data (see the section titled *Validation of the Four-State Folding Pathway* for details).

the correlation plot, with the modest difference likely due to a residual small population of U in the 10% TFE sample. Notably, if $\frac{p_{I1}}{p_{I1}+p_U}$ is constrained to 0.32 in the four-state fits, the extracted $\Delta\varpi_{FI1}$, $\Delta\varpi_{FI2}$, and $\Delta\varpi_{FU}$ values remain essentially the same, and the fit quality is unaltered (χ^2_{red} ~0.91) (*SI Appendix*, Fig. S4). *SI Appendix*, *How Robust Is the Four-State Model?* also discusses some more aspects of the validity and robustness of the four-state model.

Having established that robust and meaningful measures of chemical shift differences can be obtained from the four-state analysis and that I1 and U interconvert rapidly, we can now understand why $\Delta\varpi_{FU}$ values estimated on the basis of two-state fits (Fig. 2*C*) are, on average, less than the predicted differences. Since most of the $\Delta\varpi_{FU}$ and $\Delta\varpi_{I1U}$ values are positive, with $\Delta\varpi_{FU} > \Delta\varpi_{I1U}$, the resulting shift in the dip position from ϖ_U toward ϖ_{I1} due to exchange will reduce the fitted $\Delta\varpi_{FU}$ values. In contrast, when the exchange kinetics are properly taken into account, the fitted $\Delta\varpi_{FU}$ values provide an accurate measure of the actual differences between chemical shifts in the F and U states (Fig. 3*D*). As rapid exchange between the minor states can shift the dips in CEST profiles considerably (Figs. 1*C* and 4 *A* and *B*), it may be prudent to test if the CEST- (or CPMG-, for that matter) derived minor state shifts are physically meaningful when interpreting them in terms of structure. If the shifts obtained are not physical, this implies that the exchange model used for data fitting is incomplete.

Folding Intermediates I1 and I2 Are Largely Structured. Prior to describing the structural features of I1 and I2, it is important to establish that indeed these two intermediates are unique and that they can be distinguished on the basis of their ^{15}N chemical shifts. That this is the case is made clear by the good agreement between $\Delta\varpi_{FI1}$ values obtained from the four-state fit and $\Delta\varpi_{FI1}$[10% TFE] (RMSD 0.9 ppm; *SI Appendix*, Fig. S5*A*) and by the poor agreement with $\Delta\varpi_{FI2}$ [A17G FF + 20% glycerol] (RMSD 2.8 ppm; *SI Appendix*, Fig. S5*B*). Conversely, $\Delta\varpi_{FI2}$ values from the four-state fit agree well with $\Delta\varpi_{FI2}$ obtained from the analysis of CEST profiles of A17G FF + 20% glycerol (RMSD 1.0 ppm; *SI Appendix*, Fig. S5*C*) but are in poor agreement with $\Delta\varpi_{FI1}$[10% TFE] (RMSD 2.3 ppm; *SI Appendix*, Fig. S5*D*); see *SI Appendix*, Tables S4–S7 for compilation of chemical shift values.

Having established that the ^{15}N chemical shifts are meaningful, we can then use them with confidence to obtain insights into the conformational preferences of the I1 and I2 states. Figs. 5 *A*–*C* show $|\Delta\varpi_{FU}|$ (Fig. 5*A*), $|\Delta\varpi_{FI1}|$ (Fig. 5*B*), and

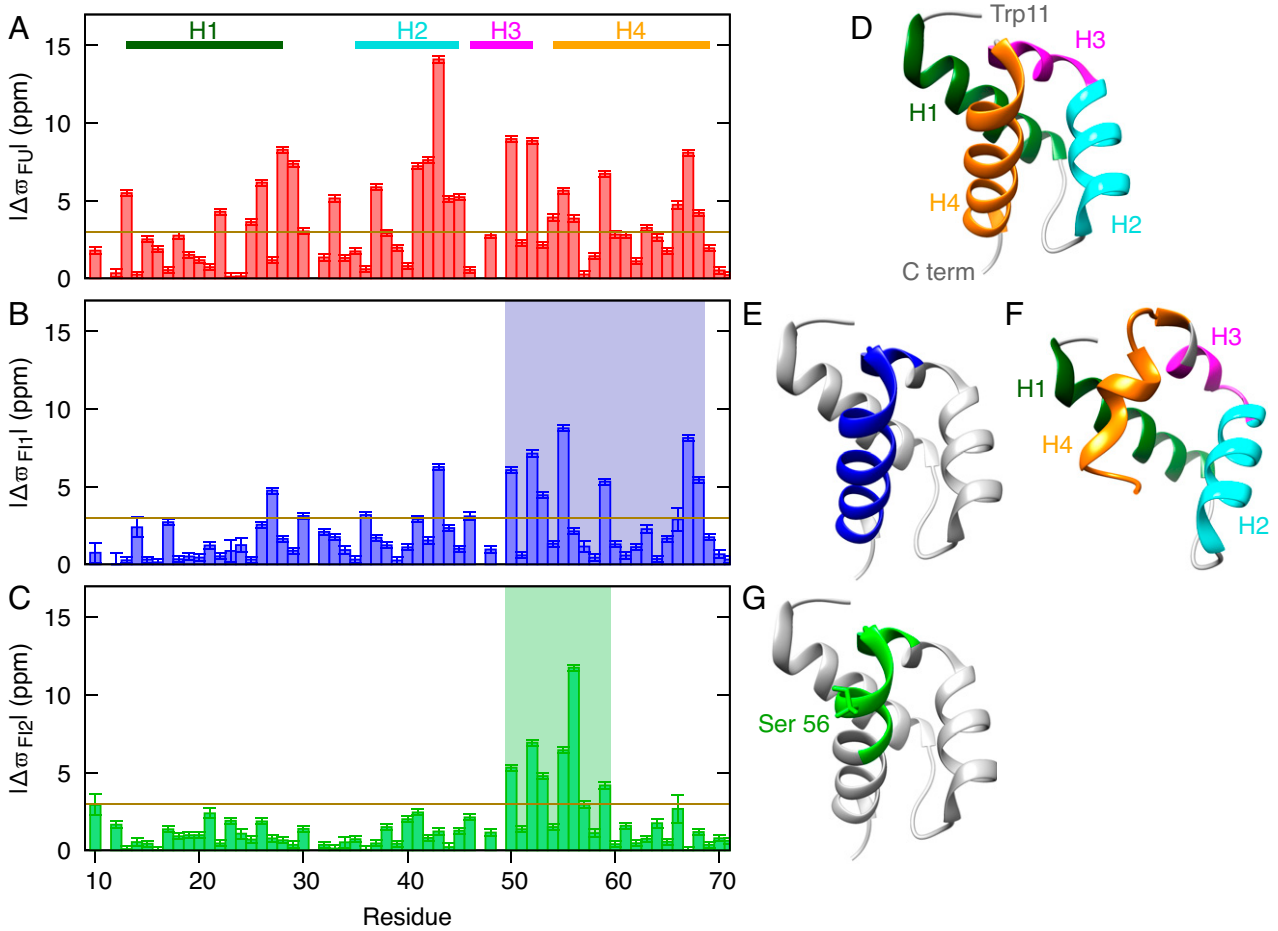

Fig. 5. Chemical shifts report on the progression of structure from U to I1 to I2. (*A*–*C*) Values of $|\Delta\varpi_{FU}|$, $|\Delta\varpi_{FI1}|$, and $|\Delta\varpi_{FI2}|$ (from the four-state analysis) are plotted along the sequence. (*D*) The native structure (57) of the FF domain (PDB: 1UZC) consists of four helices: H1 (Thr13 to Lys28; dark green), H2 (Ser-35 to Asn-45; cyan), H3 (Asp-46 to Leu-52; magenta), and H4 (Lys-54 to Thr-69; orange). (*E*) Residues with large $|\Delta\varpi_{FI1}|$ values (>3 ppm) are largely localized (*B*) to the sequence between Ser-50 and Gln-68 that is colored in blue on the structure of the WT FF domain. (*F*) The structure of the folding intermediate of FF WT (PDB: 2KZG) (37), colored as in *D*. The last five disordered residues (Val-67 to Lys-71) are not shown. (*G*) Highlighted on the structure of WT FF in green are residues from the Ser-50 to Lys-59 region, where the large $|\Delta\varpi_{FI2}|$ values (>3 ppm) are localized, as in *C*. The sidechain of Ser-56 that has the largest $|\Delta\varpi_{FI2}|$ value (>10 ppm) is shown. The olive line in each panel is drawn at 3 ppm. The first 10 (Gly-1 to Thr-10) disordered residues are not shown in *D*–*G*.

$|\Delta\varpi_{FI2}|$ (Fig. 5*C*) values for A39G FF, along with the structure of WT FF in Fig. 5*D* [Protein Data Bank (PDB) ID: 1UZC (57)]. As expected, the U state chemical shifts differ significantly from those of the native state throughout the molecule (Fig. 5*A*). In the case of the I1 state, the ^{15}N chemical shifts of the first ∼50 residues are similar to those of the native state, while amide shifts of the last ∼20 residues that encompass helices H3 and H4 differ significantly from shifts in the folded conformation (Fig. 5 *B* and *E*). Given that the I1 ^{15}N chemical shifts of A39G FF are well correlated with the I shifts of WT FF determined by a three-state analysis of CPMG data (37) (*SI Appendix*, Fig. S3*B*), the structure of the FF WT I state derived previously (37) (Fig. 5*F*) is likely to be an excellent model for the I1 conformer of A39G FF. In WT FF, both helices H1 and H2 are present in I, while helix H3 is longer and helix H4 shorter than in the native state (compare Fig. 5 *D* and *F*). An A39G FF I1 conformation with these structural features would give rise to these observed $\Delta\varpi_{FI1}$ values.

The ^{15}N chemical shifts of I2 are more native like than those for I1 (compare Fig. 5 *B* and *C*), with large $\Delta\varpi_{FI2}$ values observed only for residues in the stretch between Ser50 to Lys59 (Fig. 5 *C* and *G*) that includes the end of helix H3 and beginning of helix H4. Thus, the conformation of I2 is closer to that of the native state than I1, with the C-terminal half of helix H4 that was unstable and partially disordered in the I state of WT FF adopting a helical, native-like conformation in I2 (note the small $\Delta\varpi_{FI2}$ values for residues Gln-60 to Lys-71; Fig. 5*C*). Detailed structural studies of I2 are currently underway using a variety of CEST experiments to probe the exchange dynamics at backbone and sidechain positions in the domain (58, 66, 70–73).

The Four-State Model Explains Why A39G FF Appears To Fold without Intermediates in Stopped-Flow Experiments. The kinetic matrix for the four-state exchange model (Fig. 3*C*) can be propagated (*SI Appendix*) to predict the folding trajectory of A39G FF. Fig. 6*A* shows the populations of states U, I1, I2, and F, starting from U (p_U = 1 at t = 0 s), as a function of time obtained by solving kinetic rate equations using the rate constants from the four-state fits of our CEST measurements. Fig. 6*A* shows a gradual increase in p_{I1} (blue), followed by a subsequent buildup of I2 (green) as the U state (red) population decreases. During this folding process, both p_{I1} and p_{I2} increase to values higher than at equilibrium and then decrease as p_F (black) grows to its equilibrium value. The kinetic matrix describing the evolution of an N state system has N-1 nonzero

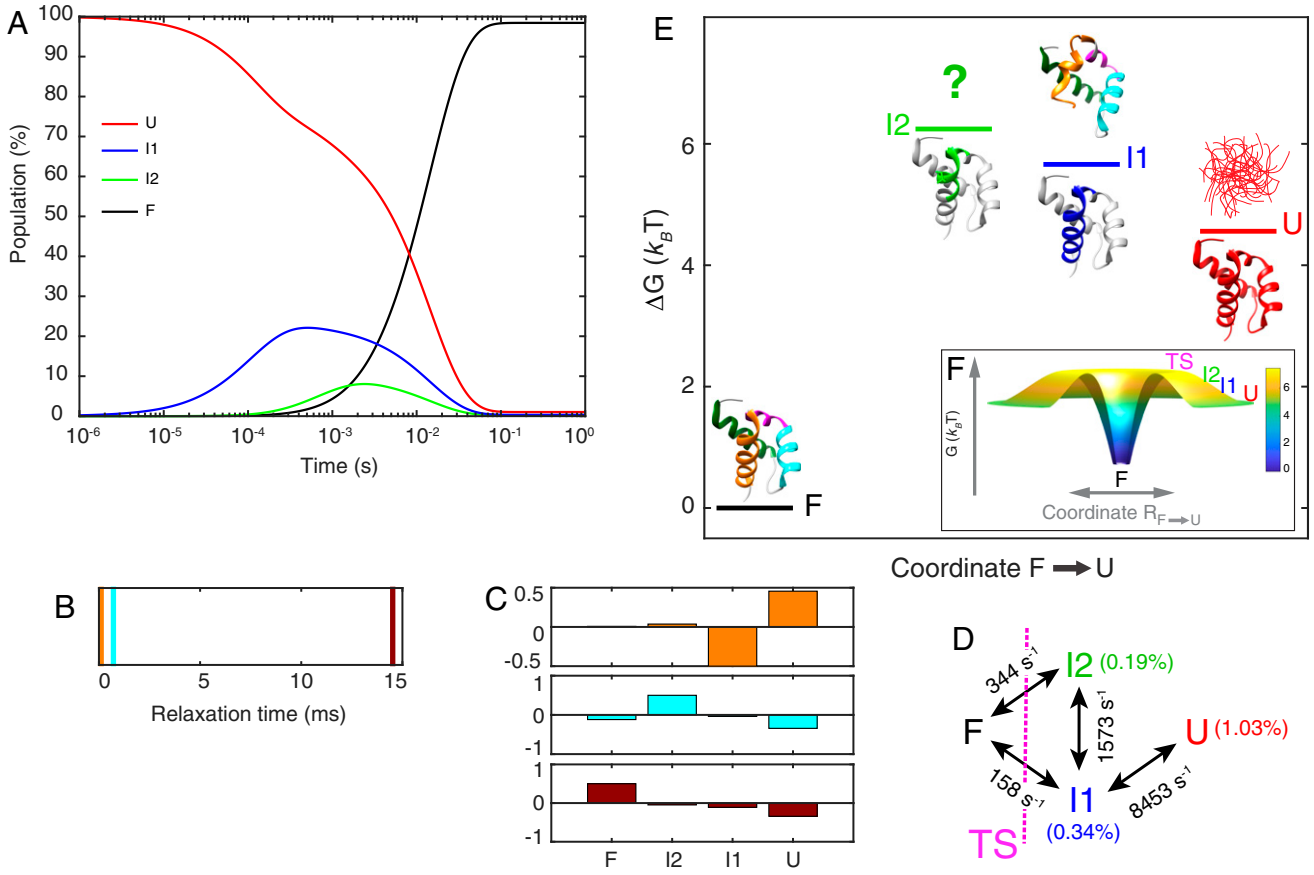

Fig. 6. A volcano-shaped FES for the A39G FF domain. (*A*) Dynamics of the folding reaction with all the molecules in state U at t = 0 s, computed by numerically propagating the kinetic equations (*SI Appendix*), describing the four-state folding model using measured rate constants from ^{15}N CEST. (*B*) Relaxation times obtained from diagonalization of the kinetic matrix (*SI Appendix*) show one slow timescale process at ∼14.52 ms (maroon) and two faster processes occurring with time constants of ∼0.71 ms (cyan) and ∼0.11 ms (orange). (*C*) Eigenvectors corresponding to the three relaxation rates in *B* highlight the principal dynamic modes as A39G FF folds. The heights of the bars correspond to the relative contribution of a given state to the eigenvector. (*D*) The folding transition state (TS) separates F from the other three states, as indicated on the four-state folding model. (*E*) Energy level diagram for A39G FF at 1 °C. The structures above the horizontal energy levels for F and I1 are those determined previously (PDB ID: 1UZC and 2KZG for F and I1, respectively), while the structure for I2 is currently not known (indicated with "?"). Below each of the U, I1 and I2 energy levels is the native structure colored according to Fig. 5 *A*–*C* to highlight the regions where chemical shifts differ significantly from those of the F state. (*F*) A schematic illustration of the volcano-shaped FES of A39G FF derived from the CEST experiments showing the increase in free energy from U to I1 to I2, as the structures become more native like, before the energy dramatically decreases because of the formation of tertiary interactions that stabilize the dominant F state.

eigenvalues that are related to the relaxation rates of "transition modes" connecting the different states in the system that in turn are given by the eigenvectors of the matrix (1, 74). The final (N^{th}) eigenvalue is zero, and the corresponding eigenvector lists the equilibrium populations. The relaxation times of the kinetic rate matrix (Fig. 6*B*) indicate a slow process (14.52 ms, maroon) and two significantly faster events (0.71 ms, cyan and 0.11 ms, orange). The eigenvectors corresponding to these relaxation times show that the fastest process (Fig. 6*C Top*, orange) is predominantly one involving the I1-U interconversion; the next fastest can be mainly described in terms of an exchange between U and I2 (Fig. 6*C*, *Middle*, cyan), while the slowest process (Fig. 6*C Bottom*, maroon) corresponds to interconversion between F and the other three states, although mainly between F and U. Thus, a picture emerges whereby F interconverts slowly with I1, I2, and U that, in turn, rapidly interconvert with each other. The folding transition state, denoted by the pink dashed line in Fig. 6*D*, separates F from the other three states. The large separation of timescales, which effectively sequesters F from the remaining conformers, leads to an apparent two-state folding mechanism via stopped-flow experiments (56), with the folding rate determined by the slowest nonzero eigenvalue.

Discussion

In this study, we have used ^{15}N CEST to elucidate the folding mechanism of the A39G FF domain by focusing on the linewidths of the minor state dips in the CEST profiles. Contributions to the linewidths of minor state dips are a powerful, and often overlooked feature, of CEST profiles that can be used to inform on the existence of additional sparse states that would otherwise escape detection, even by relaxation-based NMR experiments. Moreover, based on the present study, including linewidths in the analysis can, in some cases, extend the applicability of CEST to studies of states with lifetimes an order of magnitude shorter and populations close to 10-fold smaller than what is typically characterized using CEST. In this manner, we were able to show that A39G FF, for which folding intermediates were not detected using more traditional approaches, such as stopped-flow fluorescence (56), clearly folds via two distinct intermediates along at least two pathways and that the transition state separates the folded state from the sparely populated intermediates characterized here. As CEST "reconstructs" the spectra of these minor states, it is possible to obtain detailed structural information that in some cases can lead to structural models (51, 75–77). ^{15}N chemical shifts from the present study make it clear that I1 is structurally similar to the folding intermediate of the WT FF domain (37) that was detected using CPMG-based relaxation methods (37). I1 is thus largely structured, except for the C-terminal half of H4. Notably, the second intermediate, I2, is more natively structured, with the C-terminal half of H4 folded. The CEST-derived four-state exchange matrix establishes that F interconverts slowly with U, I1, and I2, while these three states rapidly interconvert with each other (Fig. 6), explaining why A39G FF appears to fold without intermediates when studied using other spectroscopic techniques.

The increase in structure from U to I1 to I2 that is based on a comparison of ^{15}N $\Delta\varpi$ values from CEST is consistent with previously measured *m*-values using CPMG-based experiments recorded on a number of different FF variants (8). Thermodynamic *m*-values measure how the free energy difference between states depends on urea concentration and are defined relative to the folded state, such that $m_{K-F} = m_K - m_F = -\frac{d\Delta G_{K-F}}{d[urea]}$ and $\Delta G_{K-F} = G_K - G_F$. Positive values of m_{K-F} indicate less compaction of state K relative to F, with m_K increasing as state K becomes progressively less folded (78, 79). The values, m_{U-F} = 0.74 (obtained from A17G FF), m_{I1-F} = 0.31 (obtained from WT FF), and m_{I2-F} = 0.16 (obtained from A17G FF), further establish that compaction increases as the folding reaction progresses and also show that the native state is more compact than the other minor states. Furthermore, the relatively high value of m_{U-F} compared to m_{I1-F} and m_{I2-F} explains why the addition of 1 M urea to A39G FF (Fig. 4*B*) results in a relative increase of the U state population compared to the I1 and I2 states. As p_U is significantly larger than either of p_{I1} and p_{I2} (Fig. 4*A*, *Right*), even in the absence of urea, the addition of urea effectively simplifies the exchange to two state. Interestingly, almost every FF variant studied folds via an intermediate (8, 54, 56). The results from the present study establish that the I2 state is relatively compact (compared to I1 and U), with a conformation similar to that of the native state (Fig. 5). The fact that U interconverts with the less compact and less native-like I1 conformer but not I2 suggests, rather intuitively, that folding proceeds along pathways whereby intermediates accumulate structure on route to the formation of the folded state.

The relative energies of minor state conformers identified from the four-state analysis of A39G FF domain folding reported here increase from U to I1 to I2 (i.e., $p_{I2} < p_{I1} < p_U$; Fig. 6*E*), as the amount of secondary structure in the domain grows and the tertiary structures becomes more compact. The transition to the native structure then occurs with a dramatic decrease in free energy. These observations have broad implications for the nature of the FF domain energy landscape. The foldon–funnel model proposed by Rollins and Dill (80) that builds on the "foldon" concept pioneered by Englander and coworkers (81, 82) assumes that a protein folds sequentially along its folding pathway(s) in units of secondary structure (foldons) that are stabilized by tertiary interactions. However, when the secondary structural elements are unstable, this model leads to a volcano-shaped FES (Fig. 6*F*) with the transition state located between F and the remaining states, resulting in an apparent two-state folding mechanism (80). The folding intermediates are expected to be loosely packed according to this model so that the entropy lost as the (unstable) structure is progressively formed is not yet fully compensated by interactions between secondary structural elements, giving rise to increases in free energies of states that are formed as the folding reaction progresses. Such a scenario is observed for A39G FF domain folding, where I1 has two native secondary structural elements (helices H1 and H2), while I2 has three (helices H1, H2 and the C-terminal half of helix H4), yet these helices interact with each other only loosely, giving rise to a volcano-shaped landscape (Fig. 6 *E* and *F*).

Although the present study has focused on the folding of A39G FF, showing at least a four-state process, our results have broader implications, informing on folding principles of FF domains in general. We have reexamined the folding behavior of the WT FF domain that was studied previously via CPMG approaches (8, 37), in which the data were well fit by a model containing only a single folding intermediate that is similar to the I1 state reported here. Notably, ^{15}N CEST profiles of WT FF recorded at 1 °C show clearly that the I2 state is populated as well (*SI Appendix*, Fig. S6). The folding of A17G FF, also studied by CPMG approaches, was analyzed previously in terms of a single intermediate (8), corresponding to I2 for A39G FF characterized in this work (Fig. 4*D*). Yet ^{15}N CEST experiments indicate that state I1 is also populated (*SI Appendix*, Fig. S7). Thus, the A39G variant is not unique, and a four-state folding mechanism is operative for the WT and A17G FF domains as well. It is likely that the low populations of these additional states preclude observation via CPMG. Indeed, even with excellent CPMG data, it remains unclear as to whether processes more complex than (global) three state can be properly analyzed. We have found that CEST profiles can be extremely

sensitive to the underlying exchange kinetics, even for low populated states, as illustrated in Fig. 1 and in *SI Appendix*, Fig. S8 for three- and four-state processes, respectively, increasing its utility for studying the conformational exchange between multiple states with disparate populations and lifetimes. We expect that the extraction of the full information content inherent in CEST datasets will significantly benefit from the development of machine learning–based approaches, in which algorithms are trained on vast amounts of synthetic data (83–85) covering a wide range of different potential exchange models so as to provide reasonable starting exchange parameters and minor state chemical shifts. The continued development of NMR relaxation experiments and the computational tools to best analyze them will open up the possibility for obtaining a detailed understanding of the underlying energy landscapes governing a plethora of different biological processes, including, for example, protein folding/misfolding, molecular recognition, and enzyme function.

Materials and Methods

All the FF variants were expressed in *Escherichia coli* and purified by ion exchange chromatography followed by size exclusion chromatography. Each NMR sample contained ~2 mM [U–^{15}N] protein dissolved in the desired buffer. NMR experiments were performed at fields ranging from 11.7 to 23.5 Tesla. Details of protein expression and purification, NMR experiments, and CEST profile analysis are provided in *SI Appendix*.

Data Availability. All relevant data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by funding from the Canadian Institutes of Health Research to L.E.K. and intramural funding from the Tata Institute of Fundamental Research (TIFR) (Department of Atomic Energy, Government of India) and the TIFR Hyderabad NMR center to P.V. Dr. K. Rao (TIFR Hyderabad) is acknowledged for maintenance of the TIFR Hyderabad spectrometer facilities, and Dr. G Bouvignies (Ecole Normale Superieure Paris) is acknowledged for providing the program *ChemEx*. Y.T. was supported by a Japan Society for the Promotion of Science Overseas Research Fellowship and a Uehara Memorial Foundation postdoctoral fellowship. L.E.K. acknowledges a Canada Research Chair in Biochemistry.

1. I. Bahar, R. L. Jernigan, K. A. Dill, *Protein Actions: Principles and Modeling* (Garland Science, Taylor & Francis Group, New York, 2017).
2. P. G. Wolynes, W. A. Eaton, A. R. Fersht, Chemical physics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17770–17771 (2012).
3. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
4. M. Karplus, Aspects of protein reaction dynamics: Deviations from simple behavior. *J. Phys. Chem. B* **104**, 11–27 (2000).
5. C. M. Dobson, Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
6. H. Gelman, M. Gruebele, Fast protein folding kinetics. *Q. Rev. Biophys.* **47**, 95–142 (2014).
7. A. Zhuravleva, D. M. Korzhnev, Protein folding by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **100**, 52–77 (2017).
8. D. M. Korzhnev, T. L. Religa, P. Lundström, 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(alpha) and (13)C-methyl relaxation dispersion and (1)H/(2)H-exchange NMR spectroscopy. *J. Mol. Biol.* **372**, 497–512 (2007).
9. C. Charlier *et al.*, Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4169–E4178 (2018).
10. W. A. Eaton, Modern kinetics and mechanism of protein folding: A retrospective. *J. Phys. Chem. B* **125**, 3452–3467 (2021).
11. H. S. Chung, S. Piana-Agostinetti, D. E. Shaw, W. A. Eaton, Structural origin of slow diffusion in protein folding. *Science* **349**, 1504–1510 (2015).
12. P. Vallurupalli, N. Chakrabarti, R. Pomès, 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).
13. X. S. Xie, Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics. *J. Chem. Phys.* **117**, 11024–11032 (2002).
14. J.-Y. Kim, H. S. Chung, Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **368**, 1253–1257 (2020).
15. 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).
16. A. Rangadurai, E. S. Szymaski, I. J. Kimsey, H. Shi, H. M. Al-Hashimi, Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance $R_{1\rho}$ relaxation dispersion. *Prog. Nucl. Magn. Reson. Spectrosc.* **112-113**, 55–102 (2019).
17. P. Lundström, M. Akke, Microsecond protein dynamics measured by 13Calpha rotating-frame spin relaxation. *ChemBioChem* **6**, 1685–1692 (2005).
18. J. G. Reddy *et al.*, Simultaneous determination of fast and slow dynamics in molecules using extreme CPMG relaxation dispersion experiments. *J. Biomol. NMR* **70**, 1–9 (2018).
19. A. Sekhar, L. E. Kay, An NMR view of protein dynamics in health and disease. *Annu. Rev. Biophys.* **48**, 297–319 (2019).
20. A. G. Palmer III, H. Koss, Chemical exchange. *Methods Enzymol.* **615**, 177–236 (2019).
21. 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).
22. D. A. Torchia, NMR studies of dynamic biomolecular conformational ensembles. *Prog. Nucl. Magn. Reson. Spectrosc.* **84-85**, 14–32 (2015).
23. 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).
24. N. J. Anthis, G. M. Clore, Visualizing transient dark states by NMR spectroscopy. *Q. Rev. Biophys.* **48**, 35–116 (2015).
25. 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).
26. 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).
27. 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).
28. 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).
29. P. Vallurupalli, D. F. Hansen, L. E. Kay, Probing structure in invisible protein states with anisotropic NMR chemical shifts. *J. Am. Chem. Soc.* **130**, 2734–2735 (2008).
30. D. F. Hansen, P. Vallurupalli, L. E. Kay, Quantifying two-bond 1HN-13CO and one-bond 1H(alpha)-13C(alpha) dipolar couplings of invisible protein states by spin-state selective relaxation dispersion NMR spectroscopy. *J. Am. Chem. Soc.* **130**, 8397–8405 (2008).
31. T. I. Igumenova, U. Brath, M. Akke, A. G. Palmer III, Characterization of chemical exchange using residual dipolar coupling. *J. Am. Chem. Soc.* **129**, 13396–13397 (2007).
32. B. Zhao, Q. Zhang, Measuring residual dipolar couplings in excited conformational states of nucleic acids by CEST NMR spectroscopy. *J. Am. Chem. Soc.* **137**, 13480–13483 (2015).
33. R. S. Ma *et al.*, Determination of pseudocontact shifts of low-populated excited states by NMR chemical exchange saturation transfer. *Phys. Chem. Chem. Phys.* **18**, 13794–13798 (2016).
34. 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).
35. M. J. Grey, C. Wang, A. G. Palmer III, Disulfide bond isomerization in basic pancreatic trypsin inhibitor: Multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J. Am. Chem. Soc.* **125**, 14324–14335 (2003).
36. T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. *Science* **370**, eabc2754 (2020).
37. 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).
38. B. Zhao, S. L. Guffy, B. Williams, Q. Zhang, An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.* **13**, 968–974 (2017).
39. I. J. Kimsey *et al.*, Dynamic basis for dG•dT misincorporation via tautomerization and ionization. *Nature* **554**, 195–201 (2018).
40. V. Ramanujam, C. Charlier, A. Bax, Observation and kinetic characterization of transient Schiff base intermediates by CEST NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **58**, 15309–15312 (2019).
41. 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).
42. J. Lim, T. Xiao, J. Fan, D. Yang, An off-pathway folding intermediate of an acyl carrier protein domain coexists with the folded and unfolded states under native conditions. *Angew. Chem. Int. Ed. Engl.* **53**, 2358–2361 (2014).
43. J. Milojevic, V. Esposito, R. Das, G. Melacini, Understanding the molecular basis for the inhibition of the Alzheimer's Abeta-peptide oligomerization by human serum albumin using saturation transfer difference and off-resonance relaxation NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 4282–4290 (2007).
44. N. L. Fawzi, J. Ying, R. Ghirlando, D. A. Torchia, G. M. Clore, Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* **480**, 268–272 (2011).
45. A. Sekhar *et al.*, Thermal fluctuations of immature SOD1 lead to separate folding and misfolding pathways. *eLife* **4**, e07296 (2015).
46. C. Gladkova *et al.*, An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. *EMBO J.* **36**, 3555–3572 (2017).

47. L. Deshmukh, V. Tugarinov, D. H. Appella, G. M. Clore, Targeting a dark excited state of HIV-1 nucleocapsid by antiretroviral thioesters revealed by NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **57**, 2687–2691 (2018).
48. A. Rangadurai, H. Shi, H. M. Al-Hashimi, Extending the sensitivity of CEST NMR spectroscopy to micro-to-millisecond dynamics in nucleic acids using high-power radio-frequency fields. *Angew. Chem. Int. Ed. Engl.* **59**, 11262–11266 (2020).
49. 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).
50. P. Neudecker *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).
51. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
52. P. Neudecker, P. Lundström, L. E. Kay, Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.* **96**, 2045–2054 (2009).
53. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999).
54. 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).
55. 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).
56. 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).
57. M. Allen, A. Friedler, O. Schon, M. Bycroft, The structure of an FF domain from human HYPA/FBP11. *J. Mol. Biol.* **323**, 411–416 (2002).
58. 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).
59. O. Millet, J. P. Loria, C. D. Kroenke, M. Pons, A. G. Palmer, The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *J. Am. Chem. Soc.* **122**, 2867–2877 (2000).
60. P. Vallurupalli, G. Bouvignies, L. E. Kay, Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J. Phys. Chem. B* **115**, 14891–14900 (2011).
61. N. R. Skrynnikov, F. W. Dahlquist, L. E. Kay, Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* **124**, 12352–12360 (2002).
62. A. B. Gopalan, P. Vallurupalli, Measuring the signs of the methyl ^{1}H chemical shift differences between major and 'invisible' minor protein conformational states using methyl ^{1}H multi-quantum spectroscopy. *J. Biomol. NMR* **70**, 187–202 (2018).
63. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy, Principles and Practice* (Academic Press, ed. 2, 2007).
64. H. Koss, M. Rance, A. G. Palmer III, General expressions for $R_{1\rho}$ relaxation for N-site chemical exchange and the special case of linear chains. *J. Magn. Reson.* **274**, 36–45 (2017).
65. G. Karunanithy, J. Reinstein, D. F. Hansen, Multiquantum chemical exchange saturation transfer NMR to quantify symmetrical exchange: Application to rotational dynamics of the guanidinium group in arginine side chains. *J. Phys. Chem. Lett.* **11**, 5649–5654 (2020).
66. 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).
67. J. T. Nielsen, F. A. A. Mulder, POTENCI: Prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *J. Biomol. NMR* **70**, 141–165 (2018).
68. N. A. Farrow, O. Zhang, J. D. Forman-Kay, L. E. Kay, Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **36**, 2390–2402 (1997).
69. Y. Luo, R. L. Baldwin, Trifluoroethanol stabilizes the pH 4 folding intermediate of sperm whale apomyoglobin. *J. Mol. Biol.* **279**, 49–57 (1998).
70. G. Bouvignies, P. Vallurupalli, L. E. Kay, Visualizing side chains of invisible protein conformers by solution NMR. *J. Mol. Biol.* **426**, 763–774 (2014).
71. T. Yuwen, A. Sekhar, L. E. Kay, Separating dipolar and chemical exchange magnetization transfer processes in ^{1}H-CEST. *Angew. Chem. Int. Ed. Engl.* **56**, 6122–6125 (2017).
72. D. Long, A. Sekhar, L. E. Kay, Triple resonance-based $^{13}C(\alpha)$ and $^{13}C(\beta)$ CEST experiments for studies of ms timescale dynamics in proteins. *J. Biomol. NMR* **60**, 203–208 (2014).
73. V. P. Tiwari, P. Vallurupalli, A CEST NMR experiment to obtain glycine $^{1}H^{\alpha}$ chemical shifts in 'invisible' minor states of proteins. *J. Biomol. NMR* **74**, 443–455 (2020).
74. F. Noé, S. Fischer, Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **18**, 154–162 (2008).
75. 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).
76. 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).
77. 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).
78. 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). Correction in: *Protein Sci.* **5**, 981 (1996).
79. E. S. Courtenay, M. W. Capp, R. M. Saecker, M. T. Record Jr., Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: Interpretation of urea and guanidinium chloride m-values and their correlation with changes in accessible surface area (ASA) using preferential interaction coefficients and the local-bulk domain model. *Proteins* **4** (suppl. 4), 72–85 (2000).
80. G. C. Rollins, K. A. Dill, General mechanism of two-state protein folding kinetics. *J. Am. Chem. Soc.* **136**, 11420–11427 (2014).
81. H. Maity, M. Maity, M. M. G. Krishna, L. Mayne, S. W. Englander, Protein folding: The stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4741–4746 (2005).
82. S. W. Englander, L. Mayne, M. M. G. Krishna, Protein folding and misfolding: Mechanism and principles. *Q. Rev. Biophys.* **40**, 287–326 (2007).
83. M. A. Beckwith, T. Erazo-Colon, B. A. Johnson, RING NMR dynamics: Software for analysis of multiple NMR relaxation experiments. *J. Biomol. NMR* **75**, 9–23 (2021).
84. F.-A. Chao, Y. Zhang, R. A. Byrd, Theoretical classification of exchange geometries from the perspective of NMR relaxation dispersion. *J. Magn. Reson.* **328**, 107003 (2021).
85. G. Karunanithy, D. F. Hansen, FID-Net: A versatile deep neural network architecture for NMR spectral reconstruction and virtual decoupling. *J. Biomol. NMR* **75**, 179–191 (2021).