



Structure of a transient protein-folding intermediate by pressure-jump NMR spectroscopy

Elahe Masoumzadeh^{a,1}, Joseph M. Courtney^{a,1,2}, Cyril Charlier^{a,3}, Jinfa Ying^a, Philip Anfinrud^a , and Adriaan Bax^{a,4} 

Contributed by Adriaan Bax; received July 20, 2025; accepted September 12, 2025; reviewed by Lewis E. Kay and Peter E. Wright

Protein folding, as commonly portrayed, involves exploration of a rough, high-dimensional landscape, ending with a final descent into a low-energy folded state. During that journey, the protein may visit shallow basins corresponding to metastable structures, potentially of biological importance. Structural characterization of transiently populated metastable states is challenging due to their low population, which limits traditional NMR, and also makes crystallization for X-ray diffraction difficult without stabilizing mutations, covalent modifications, or the addition of antibodies. Here, we report the structural characterization of the on-pathway folding intermediate of a pressure-sensitized ubiquitin mutant. The obtained non-native β -sheet registry was previously shown to be necessary in the PINK1 mitophagy pathway. We used fast pressure jumps to repeatedly initiate folding and advanced NMR measurements to probe the evolving ensemble of protein conformations. The results reported here demonstrate that the non-native β -sheet hydrogen bond registry can act as a metastable trap during protein folding. This work provides a template for future investigation of metastable conformations and protein folding with rich structural detail.

ubiquitin | retracted strand | metastable protein-folding intermediate | CS-Rosetta | molecular dynamics

Over the past half-century, it has been shown that proteins fold from an ensemble of disordered initial states to the final folded state through multiple, potentially metastable conformations (1). Because the energy wells of most natively folded proteins are very deep, the populations of unfolded and excited states at physiological temperatures and solvent conditions often are very low, complicating their study (2, 3). However, despite their low populations and short lifetimes, near-native excited states of proteins can be necessary for posttranslational modification (3) or can lead to the formation of toxic oligomers (4, 5).

Structural investigation of metastable states visited during protein folding is complicated by short (ms timescale) lifetimes and small populations of these species under native conditions. These “invisible states” traditionally have been investigated under static conditions using backbone amide hydrogen exchange or NMR experiments that indirectly observe their effects on spectra of the major state (6–9). To study excited states with energies much higher than the native state, sample conditions are typically changed abruptly from those favoring the unfolded state to those favoring the native state. For example, rapid changes of buffer conditions by dilution of denaturant, rapid mixing with requisite cofactors, or pH jumps allow experimentalists to initiate folding and conformational changes with high time resolution (10–13). Alternatively, rapid changes in physical conditions may be accomplished by a sudden jump in temperature, which typically destabilizes the protein and triggers unfolding, or by an abrupt change in hydrostatic pressure, which can trigger either protein unfolding or folding (14). In particular, rapid cycling of hydrostatic pressure can be used to repeatedly alter the energy surface, permitting extensive measurements of the excited state conformations accessed by the protein during folding (14, 15). For small proteins, the folding process often proceeds via a downhill energy pathway (16). However, along their folding trajectory, proteins can also encounter local energy minima associated with metastable states, which will differ if there exist multiple parallel folding pathways (17–20). These metastable states may be on-pathway folding intermediates, aggregates, misfolded species, or functional conformational states.

Pressure switching is arguably the most benign and easily reversible perturbation of the energy landscape, enabling protein structural probing under native buffer conditions (21). In order to significantly populate the unfolded state of a protein at high pressure, its volume must be substantially smaller than that of the native state (22, 23). For such proteins, pressure-jump fluorescence measurements have yielded valuable insights into the kinetics of these structural switches during the folding process, including the presence of trapped intermediate states (14). Here, we demonstrate that combining reversible, millisecond pressure jumps with NMR spectroscopy enables determination of the

Significance

Extensive computational and experimental studies have aimed to deepen our understanding of the pathways by which proteins fold. However, the short-lived nature of folding intermediates has severely limited experimental access to high-resolution information on their structure and their evolution during protein folding. Using pressure-jump NMR spectroscopy, we find that during folding, ubiquitin transiently adopts a non-native β -sheet registry, previously identified for a phosphorylated state of the protein that is key to its role in the PINK1 mitophagy pathway.

Author affiliations: ^aLaboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD 20892-0520

Author contributions: J.M.C. and A.B. designed research; E.M., J.M.C., C.C., and J.Y. performed research; P.A. contributed new reagents/analytic tools; E.M. and J.M.C. analyzed data; and E.M., J.M.C., and A.B. wrote the paper.

Reviewers: L.E.K., University of Toronto; and P.E.W., Scripps Research Institute Department of Integrative Structural and Computational Biology.

Competing interest statement: Research in the authors' laboratory was funded by the intramural program of the National Institute of Diabetes and Digestive and Kidney Diseases.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹E.M. and J.M.C. contributed equally to this work.

²Present address: Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT 06030-3205.

³Present address: Toulouse Biotechnology Institute, University of Toulouse, Toulouse 31077, France.

⁴To whom correspondence may be addressed. Email: bax@nih.gov.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2519493122/-/DCSupplemental>.

Published October 8, 2025.

three-dimensional structure of such intermediates at atomic resolution under near-native conditions. Our analysis uses conventional input restraints, including chemical shifts, ^1H - ^1H nuclear Overhauser (NOE) cross-relaxation, and residual dipolar couplings.

Ubiquitin is a small, 76-residue eukaryotic protein involved in diverse intracellular signaling pathways. Its phosphorylation, which occurs while in a transient conformational state, is essential for the initiation of mitophagy (3). To gain insight into ubiquitin's folding pathway, we developed a pressure-jump apparatus that can repeatedly switch the hydrostatic pressure in an NMR sample cell between 1 bar and up to 2.8 kbar within a few milliseconds (21). Using this hardware, we monitor in real time the folding of ubiquitin V17A/V26A mutant (VA2-ubiquitin), where the mutations were introduced to increase the difference in volume between the unfolded and folded states by *ca* 60 mL/mol (24), thereby lowering the midpoint of pressure-induced denaturation from 5.4 to 1.4 kbar (21). The folding pathway of wild-type ubiquitin has been the subject of much experimental and computational work, with disagreement on whether it is a two-state folder or passes through metastable intermediates (25–31). While many experiments failed to detect it, fluorescence measurements on F45W ubiquitin pointed to the presence of a metastable folding intermediate (28, 29, 32). Our own ensemble-averaged ^{15}N chemical shift measurements during folding revealed the parallel presence of both single-barrier and double-barrier folding pathways, with their relative efficiency strongly dependent on temperature (21, 33).

Results

A two-dimensional ^1H - ^{15}N pressure-jump heteronuclear single quantum correlation (PJ-HSQC) spectrum of VA2-ubiquitin, recorded with ^{15}N evolution starting 50 ms after the pressure drop, reveals numerous weak resonances, in addition to those of the natively folded species (Fig. 1*A*). These weak resonances disappear when the start of ^{15}N evolution is delayed to 500 ms after the pressure drop (*SI Appendix, Fig. S1A*) and correspond to a metastable structure that converts to the native state at a rate of *ca* 15 s^{-1} at 298 K (21). Because the V17A/V26A mutations involve contacts within the N-terminal half of ubiquitin, which has been implicated as the “folding nucleus” of ubiquitin (34, 35), it seemed possible that the metastable intermediate was a consequence of these mutations. To address this question, we also recorded a pressure-jump HSQC spectrum for the L50A mutant of ubiquitin (*SI Appendix, Fig. S1B*), which can be unfolded by hydrostatic pressures comparable to those of VA2-ubiquitin (36). The L50A mutation site is located outside the region where the largest chemical shift differences are observed between the intermediate and native states (Fig. 2*A*) and well outside the “folding nucleus” (*SI Appendix, Fig. S2*). As shown in *SI Appendix, Fig. S1B*, the L50A-ubiquitin mutant also shows weak resonances that disappear over time. Assignment of these additional intermediate state resonances for both mutants revealed that they had very similar chemical shifts (Fig. 1*B*), indicating that the same metastable folding intermediate is present for both L50A and VA2-ubiquitin. All further analysis presented below pertains to the VA2 mutant for which a more extensive set of structural restraints was obtained, but conclusions regarding the folding pathway equally apply to L50A and, by inference, to native ubiquitin.

Chemical Shifts of the Intermediate State. To determine the 3D structure of the transient intermediate state (I-state), we measured its backbone ^1H , ^{15}N , and ^{13}C chemical shifts during repeated

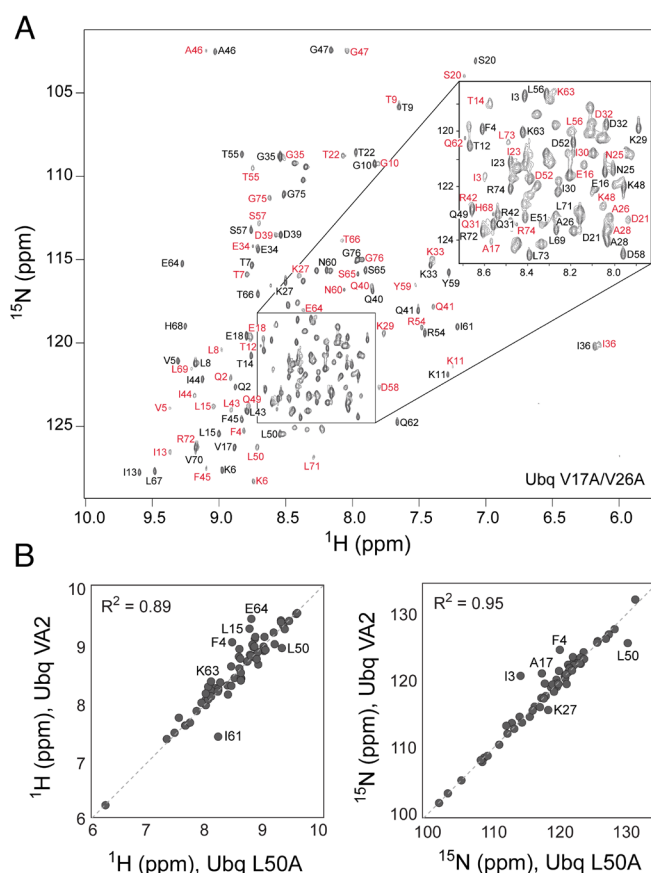


Fig. 1. ^1H and ^{15}N chemical shift assignments of V17A/V26A ubiquitin, recorded at 298 K using a PJ-HSQC pulse sequence. Each indirect acquisition point is preceded by an 8-s high-pressure (2.4 kbar) equilibration delay, with the pressure dropped to 1 bar immediately after polarization transfer from ^1H to ^{15}N . ^{15}N chemical shift evolution is initiated after a 50-ms folding delay, with all three species—folded, unfolded, and intermediate—present during ^1H acquisition. Consequently, peaks corresponding to all three states are observed in the HSQC spectrum. (A) The 900 MHz ^1H - ^{15}N PJ-HSQC NMR spectrum showing folded (black), intermediate (red), and unfolded peaks (unlabeled). (B) Comparison of ^1H and ^{15}N chemical shifts of the L50A I-state (recorded at 278 K, see *SI Appendix, Fig. S1*) and the VA2 I-state (298 K).

pressure-jump cycles that were synchronized with standard heteronuclear correlation NMR experiments described previously (21, 37). Comparison of these shifts with those in the native state revealed the largest differences for strands $\beta 1$ and $\beta 5$, as well as the “S65 loop” immediately preceding $\beta 5$ (Fig. 2*A*). Analysis of these chemical shifts by TALOS-N software (38) pointed to a limited number of large backbone torsion angle differences relative to the native structure (Fig. 2*B*). These substantial differences mostly pertained to the S65 loop, but chemical shifts of residues L73 and R74, which are disordered in the native state, also indicated rigidification of the backbone up to residue G75. TALOS-N analysis (38) showed that secondary-structure differences were highly localized, leaving the majority of residues largely unaffected (Fig. 2*A* and *B*). While the secondary structure for other residues near the N and C termini of the protein remained unchanged, their comparatively large chemical shift differences between the natively folded and intermediate species (Fig. 2*A*) were consistent with the exquisite sensitivity of chemical shifts to differences in electrostatics and hydrogen bonding geometry.

CS-Rosetta Model of the Intermediate State. We next used the CS-Rosetta program to develop a full 3D structural model compatible with the chemical shifts of the I-state (39).

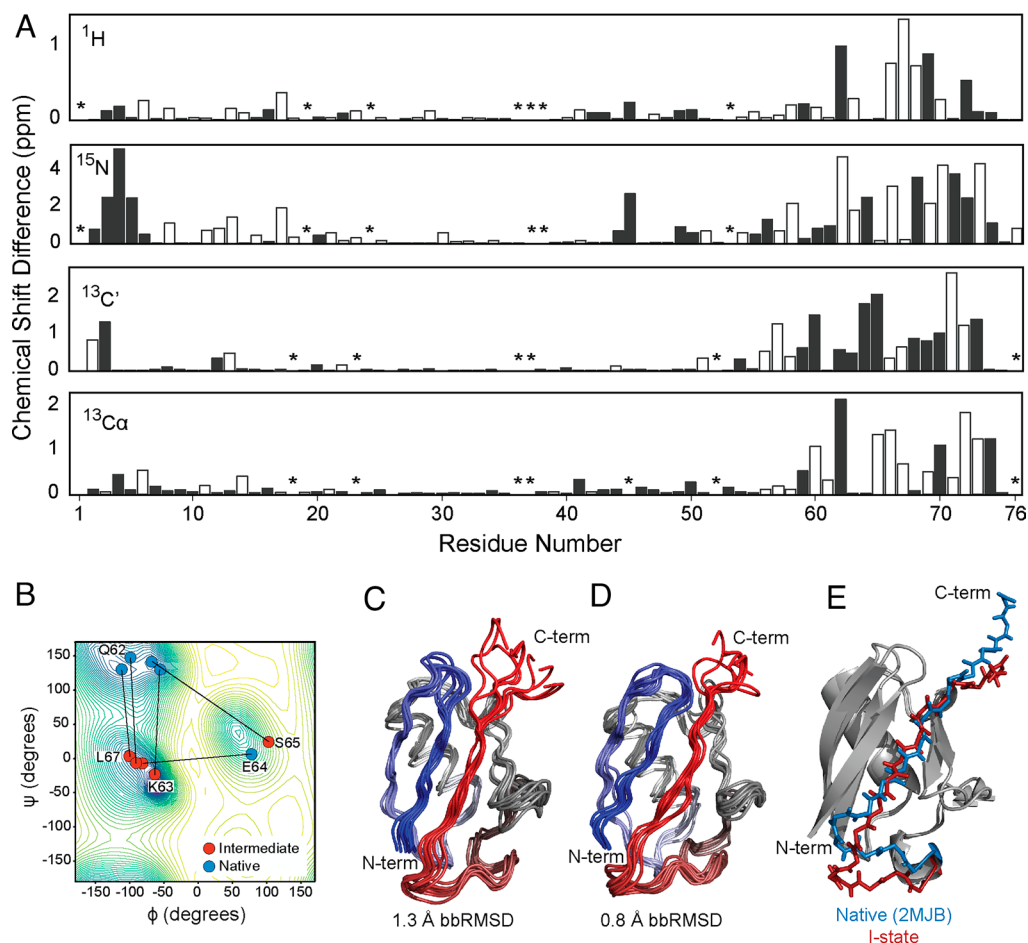


Fig. 2. Characterization of folding intermediate structure by chemical shift analysis. (A) Difference in chemical shifts of the I-state and the native state for ^1H , ^{15}N , $^{13}\text{C}'$, and ^{13}Ca (solid bars: positive values; open bars: negative values; asterisks: missing assignments, presumably due to overlap with native species) (BMRB code: 53268). (B) Comparison of native state backbone dihedral angles (blue) to TALOS-N values for the I-state (red). For clarity, only residues for which $|\Delta\phi| + |\Delta\psi| \geq 90^\circ$ are shown. (C) Ensemble of 10 lowest energy conformers obtained by CS-Rosetta using only I-state backbone chemical shifts. Coloring of the chain progresses from blue (N terminus) to red (C-terminus). The backbone atomic coordinate RMSD over this ensemble is 1.3 Å. (D) CS-Rosetta ensemble when NOE and RDC restraints were included in the input, reducing the backbone RMSD to 0.8 Å. The coloring reflects progression from the N terminus (blue) to the C-terminus (red). (E) Overlay of the lowest-energy CS-Rosetta model of the I-state (red) with the native structure (blue; PDB: 2MJJB).

Superposition of the resulting 10 lowest-energy models of the folding intermediate revealed native-like topology but, remarkably, all models exhibited a shifted registry by two residues of strand $\beta 5$ relative to the native structure (Fig. 2E). However, with a rmsd (RMSD) of 1.3 Å relative to their averaged atomic backbone coordinates, these CS-Rosetta models still displayed substantial heterogeneity (Fig. 2C).

I-State Refinement by ^1H - ^1H NOE. To validate the change in $\beta 5$ registry and improve its coordinate precision, we carried out a pressure-jump version of the common three-dimensional NOE experiment. A high-pressure (2.5 kbar) interscan equilibration delay of 8 s resulted in $\sim 84\%$ of the proteins to unfold at 298 K prior to the start of each repeat of the pulse sequence (Fig. 3A). Then, following t_1 evolution that encoded the resonance frequencies of this unfolded state, the sample pressure was dropped to 1 bar, thereby initiating the folding process. After the 150-ms NOE mixing period, at which time point folding to the native state was nearly complete, the magnetization was read out by means of a standard gradient-enhanced HSQC pulse sequence (SI Appendix, Fig. S3A) (40). The result was a 3D spectrum with resolved cross-peaks that correlated unfolded ^1H chemical shifts to ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts of the folded state, with intensities modulated by the sum of all NOE interactions that were present

during the mixing period, comprising unfolded, intermediate, and folded states.

The maximum I-state population increases with temperature, but its lifetime decreases (SI Appendix, Fig. S4). As a compromise, the NOE measurement was performed at 298 K where the maximum population reaches $\sim 25\%$ and the lifetime of the I-state is *ca* 60 ms (21). Because the protein exists in three different states during the NOE period, the resulting cross-peak intensities reflect a time-average of the cross-relaxation rates in those states. The 150-ms NOE period covered both the build-up and decay of the I-state population. Under these conditions, about 50% of the protein passes through the metastable I-state, whereas the other half follows a single-barrier pathway (21). Therefore, NOE cross peaks arising from the I-state are expected to be weak. Indeed, the vast majority of peaks observed in the resulting spectrum correspond to contacts in the native conformation (Fig. 3B). NOE interactions arising from the unfolded state are vanishingly weak because its population is short-lived and its rapid internal motions are of large amplitude, leading to slow cross-relaxation. However, the 3D pressure-jump NOESY spectrum also shows numerous cross peaks that are entirely inconsistent with the folded structure, absent in the pressure-denatured unfolded state, and only compatible with a shift in registry of strand $\beta 5$ (Fig. 3 C and D).

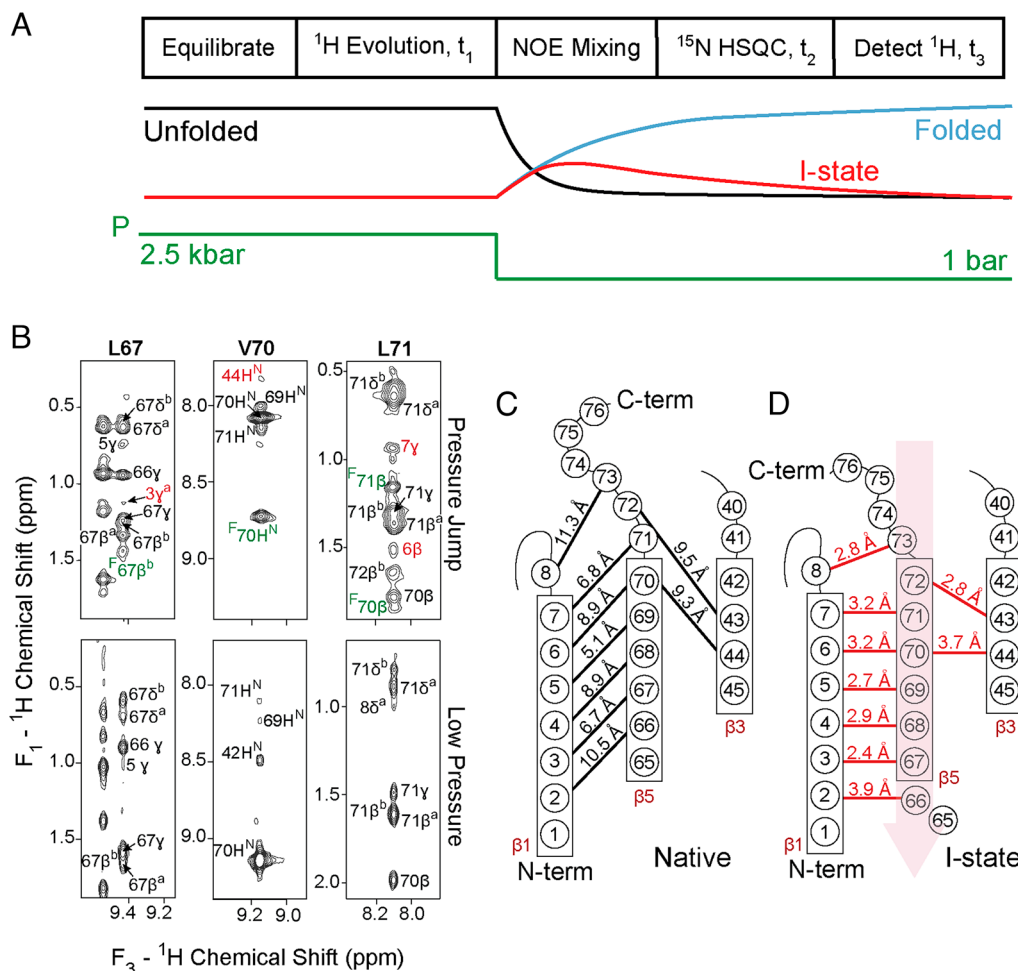


Fig. 3. Measurement of NOEs in the I-state. (A) Schematic of the pressure-jump NOESY experiment. (B) Narrow strips taken through two 800-MHz 3D NOESY spectra, showing cross peaks for the transient I-state (Top row; pressure jump) and the native state (Bottom row; static ambient pressure). The amide proton serving as the source of magnetization is selected by its ^{15}N frequency during t_2 , as marked by the column labels. Chemical shifts were detected at 1 bar in the folded state (both experiments). Black labels correspond to short-range contacts in the natively folded species. Red labels in the upper row of strips indicate long-range NOE contacts that are present only in the I-state. Green labels result from the ~16% of protein that had not yet unfolded at the end of the equilibration phase. (C and D) Schematic representation of the hydrogen bond registry of strands $\beta 1$, $\beta 5$, and $\beta 3$ for (C) the native state and (D) the I-state. Distances corresponding to the shortest observed interstrand I-state NOEs involving a backbone amide proton are marked in red and were taken from the retracted strand X-ray structure (50XH). Distances marked in black correspond to the native state NMR structure (2MJB).

While most observed NOE interactions correspond to the native state, peaks marked with red labels in Fig. 3B are only consistent with a retracted strand conformation in which the registry between $\beta 5$ and $\beta 3$ shifts by two residues (Fig. 2E). For example, contacts are observed between $\beta 5$ and the flanking strands $\beta 1$ and $\beta 3$ (L67H $^{\text{N}}$ -I3H $^{\text{Y}}$ $^{\text{a}}$, V70H $^{\text{N}}$ -I44H $^{\text{N}}$, and L71H $^{\text{N}}$ -T7H $^{\text{Y}}$) as well as contacts in the C-terminus (R72H $^{\text{N}}$ -L43H $^{\text{a}}$ and L73H $^{\text{N}}$ -L8H $^{\text{Y}}$; SI Appendix, Fig. S5) indicating rigidification compared to the native state, and all consistent with a shift in registry for $\beta 5$ (Fig. 3 C and D). NOE contacts are also seen between the end of strand $\beta 2$ and the S65 loop (e.g., K63H $^{\text{N}}$ -P19H $^{\text{b}}$, E64H $^{\text{N}}$ -P19H $^{\text{b}}$), indicating that the S65 loop contacts the loop containing residue P19 (SI Appendix, Fig. S5).

I-State Validation By Residual Dipolar Couplings. Residual dipolar couplings (RDCs) in weakly aligned proteins are precise reporters of bond vector orientations and highly effective for structure validation (41). Remarkably, a dilute, *ca* 10 mg/mL nematic suspension of Pf1 filamentous phage, commonly used for obtaining protein alignment, was previously shown to be sufficiently robust to withstand 2.5 kbar of hydrostatic pressure (42). Even the repeated 8% solvent volume compression and

expansion associated with the pressure jumps did not destroy the nematic liquid crystalline phase of Pf1, provided that pressure jumps were lengthened to have a duration of at least 4 ms. For measurement of the RDCs under Pf1-aligned conditions, an 8-s high-pressure period is used to unfold the protein. At the end of this period, ^1H magnetization is transferred to ^{15}N , immediately followed by a pressure drop to 1 bar which initiates folding (Fig. 4A). Once a sufficient population of I-state has built up, the ^{15}N chemical shift is encoded and scalar and dipolar couplings are allowed to evolve for either 5.35 ms [$\sim 1/(2 J_{\text{NH}})$] or 10.7 ms ($\sim 1/J_{\text{NH}}$) in interleaved reference and attenuated 3D NMR spectra, respectively (SI Appendix, Fig. S3B). The residual dipolar couplings are then calculated from the ratio of peak intensities in the reference and attenuated spectra (SI Appendix, Fig. S6) (43).

Because the chemical shifts and RDCs of the unfolded, native, and intermediate states are encoded simultaneously, the RDCs observed for the native and unfolded states act as an internal check. With an RMSD of 1.4 Hz between observed and predicted values, the RDCs measured for the native state fit well to the native structure (Protein Data Bank (PDB) entry: 2MJB) (44) (Fig. 4B). The unfolded state, which is dynamically highly disordered, yielded values for the RDCs that were close to zero. In contrast,

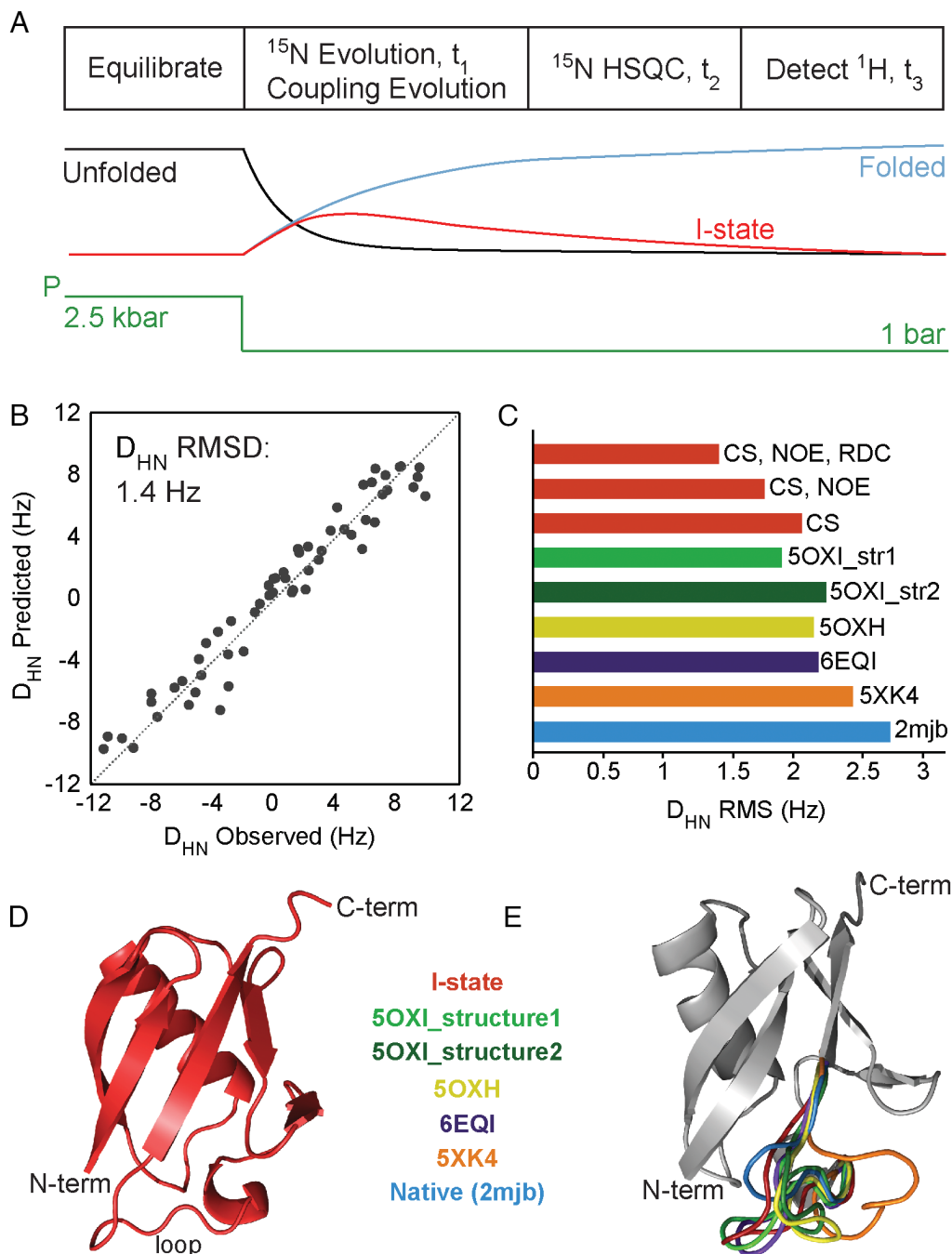


Fig. 4. Measurement of VA2 ubiquitin RDCs at 900 MHz. (A) Schematic timing diagram used to measure RDCs. The populations of the unfolded (black), intermediate (red), and natively folded (blue) states of the protein are shown together with the applied pressure (green) for a single scan of the experiment. The full spectrum consisted of 6,788 repeats of this depressurization, measure, repressurization cycle (for details, see *SI Appendix, Fig. S3B*). (B) Comparison of the RDCs observed in the P-jump experiment for the native state to values predicted for PDB entry 2MJB. (C) RMSD between observed I-state RDCs and values best fitted for various structural models. CS is the CS-Rosetta model; CS, NOE additionally includes NOEs; CS, NOE, RDC additionally includes RDC restraints. (D) Structure of the lowest energy I-state conformer, obtained with CS-Rosetta from CS, NOE, and RDC data. (E) Models, rotated by 45° about the vertical axis relative to (D), highlighting the S65-loop conformations for different structures.

RDCs of the I-state were comparable in magnitude to those measured for the natively folded structure but fit rather poorly to the native-state conformation (RMSD of 2.8 Hz). Whereas the RDCs for residues 2 to 58 of the I-state remained in good agreement with the native structure (RMSD 1.8 Hz; *SI Appendix, Fig. S7A*), the RDCs for residues 59 to 73 agreed poorly (*SI Appendix, Fig. S7B*, RMSD 4.3 Hz). Having confirmed the accuracy of the measured RDC values, we then performed another CS-Rosetta calculation incorporating these RDCs together with I-state NOEs

as structural constraints. The resulting low-energy structure retained the general features of the chemical shift-only structure (Fig. 4D) but exhibited considerably tighter convergence (0.8 Å versus 1.3 Å backbone RMSD, Fig. 2D). We cross-validated this structure by performing eight separate CS-Rosetta calculations with each omitting one of the eight amide RDCs in the S65-loop (residues 59 to 69) and found good agreement between each omitted RDC and its corresponding CS-Rosetta model prediction (RMSD 1.8 Hz; see *SI Appendix, Fig. S8*).

Comparison With Prior Retracted-Strand Structures. Studies of S65-phosphorylated ubiquitin (pUbq) and a L67S ubiquitin mutant both showed an equilibrium between natively folded ubiquitin and conformers in which $\beta 5$ was retracted (3, 45, 46). The population of retracted $\beta 5$ conformers in pUbq becomes dominant above neutral pH (46), which allowed its structure determination by NMR spectroscopy (PDB entry 5XK4). While overall similar to the X-ray structure of a T66V/L67N ubiquitin mutant (PDB entry 5OXH) (3), rather different conformations for the S65 loop were seen for the pUbq NMR structure and this mutant X-ray structure. Importantly, Komander and coworkers showed that the transient presence of retracted $\beta 5$ conformers was essential for enabling phosphorylation of S65, but no structural detail for this transient, very lowly populated state in wild-type ubiquitin could be obtained. The I-state structures derived from our pressure-jump NMR measurements therefore represent atomic resolution models for wild-type ubiquitin in the $\beta 5$ -retracted state.

I-state RDCs yielded better agreement with the various retracted strand structures than with the native ubiquitin structure (Fig. 4C). However, the S65 loop differs substantially among these retracted-strand structures (Fig. 4E). In pUbq (PDB 5OXH) (3, 45) the loop folds toward the end of $\beta 3$, while in our folding-intermediate structure, it makes NOE contacts with P19. The X-ray structure of the L67S mutant (PDB 5XOI) contains two distinct structures, differing only in the orientation of the S65 loop (labeled as structure1 and structure2 in Fig. 4E). Both structure1 and the PINK1-bound ubiquitin X-ray structure (6EQI) have loop orientations similar to our I-state, with contacts between the S65 loop and the P19 loop. The other retracted-strand ubiquitin X-ray structures (5OXH and structure2 of 5OXI) have loop orientations that are inconsistent with the observed NOE contacts between the S65 and P19 loops. Similarly, none of the solution NMR structures of enzymatically phosphorylated ubiquitin (5XK4) showed contacts between the S65- and P19-loops, but instead revealed a two-residue extension of ubiquitin's second helix up to residue I61 (46).

In our I-state structure, the S65-loop adopts a well-ordered conformation positioned toward the P19-loop (*SI Appendix, Fig. S5*). This structure is supported by NOEs from residues Q62, K63, and E64 to P19, which are located in the turn that precedes ubiquitin's main α -helix (*SI Appendix, Fig. S5*). Backcalculating eight RDC values from the S65-loop in the intermediate structure while omitting the corresponding RDC from the CS-Rosetta input showed significantly better agreement with structures based on NMR data relative to the pUbq crystal structures (*SI Appendix, Fig. S9*), thereby independently validating that the structural differences observed in the S65 loop are real and likely result from the electrostatic repulsion between the negatively charged phosphate group and the sidechains of residues E18 and D21. Moreover, the random coil index server (47) indicates that, based on its backbone chemical shifts, the S65-loop exhibits a degree of backbone order that is no lower than for the whole molecule, suggesting the absence of large amplitude internal dynamics for this loop. Excluding the S65 loop (Y59-L69) and the C terminus (R74-G76), the I-state structure has a 0.7 to 1.4 Å backbone coordinate RMSD relative to crystal structures that exhibit a retracted $\beta 5$ strand (Fig. 4C), indicating close agreement between the nonvariable structural regions.

Is the I-State Observed in Molecular Dynamics Trajectories?

Ubiquitin has been widely used as a model system for molecular dynamics (MD) simulations of protein folding (48–50). Piana et al. performed several milliseconds-scale all-atom simulations starting from both folded and unfolded states, observing multiple

folding and unfolding events (35). Notably, due to the time-reversal symmetry of their integrator, unfolding events can also be viewed as folding events. To reduce the high computational burden of searching the entire trajectory for conformers that contain the retracted $\beta 5$ -strand conformation, we simply compared the hydrogen bonds involving strand $\beta 5$ to those in the native fold and I-state. This approach reduced the search to the comparison of two numbers: The average H-bond distances across the backbone residues of $\beta 1$, $\beta 3$, and $\beta 5$. Trajectories of RMS H-bond donor–acceptor interatomic distances for each state show three, relatively short periods where frames display H-bonding more similar to that seen in our experimentally determined I-state than to the native state (boxed regions in Fig. 5A and B). Frames from these I-state-like periods were further analyzed by applying dimension reduction using the UMAP algorithm, followed by clustering with DBSCAN (51). UMAP is a powerful method for projecting high-dimensional data, such as MD trajectories of protein backbone coordinates, into a lower-dimensional space. It constructs a high-dimensional graph representation of the data and optimizes a corresponding low-dimensional diagram to preserve both global and local structures. UMAP is often preferred for its balance between structural preservation and computational efficiency. Applying UMAP allowed for better separation of distinct intermediate states and clearer identification of more populated conformations (see Methods). Twelve clusters were identified, each representing conformations similar to the experimentally determined I-state, with most structural variability localized to the 3_{10} helix (L56-Y59) and the S65 loop regions (Fig. 5C and *SI Appendix, Fig. S10*).

Discussion

Our study highlights the complementarity between pressure-jump NMR and other methods such as relaxation-dispersion and chemical exchange saturation transfer (CEST) for studying conformational exchange processes. These latter methods provide atomic-resolution access to the presence of lowly populated excited states at equilibrium that are transiently sampled from the ground state but are inherently limited to states that are revisited multiple times. By contrast, pressure-jump methods perturb the system out of equilibrium and enable structural characterization of intermediates that form only once along the folding trajectory, such as the I-state captured here. Together, the two approaches provide highly synergistic views of transient protein conformations.

CEST experiments have shown that ubiquitin's retracted-strand conformation is low enough in energy to be accessed under native conditions and thereby enable phosphorylation by PINK1 (3). Its low energy and structural similarity to the native fold suggests it is likely accessible as a folding intermediate. Indeed, about half of all molecules follow this pathway during folding (21). It is possible and perhaps even likely that such parallel folding pathways exist for many proteins that sample multiple metastable “excited-state” conformations under equilibrium conditions (9, 52).

It was gratifying to see that one quarter of the folding/unfolding transitions observed in very long all-atom molecular dynamics trajectories of ubiquitin involved a retracted $\beta 5$ intermediate (Fig. 5), which compares to roughly one half during our experimental folding study (21). Moreover, the low population of this intermediate state under equilibrium folding/unfolding conditions appears consistent with its very low (*ca* 0.7%) population observed experimentally under physiological conditions (3). Beachamp et al. analyzed very long molecular dynamics trajectories of 14 proteins by introducing novel Markov state models (53). They showed the presence of native-like but register-shifted β -sheet

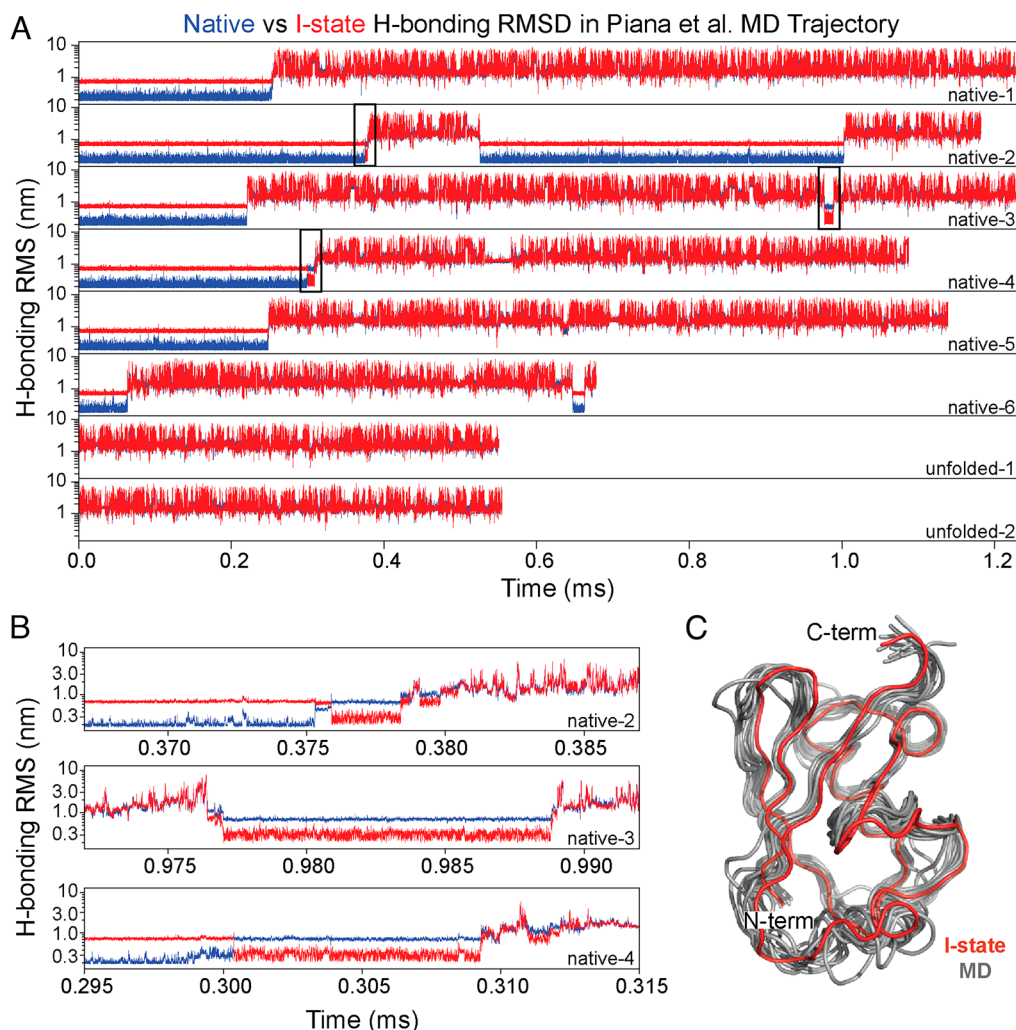


Fig. 5. Transient presence of retracted-strand I-state conformers in the Piana et al. MD trajectory (35). (A) Rms distances (log scale) for selected H-bond atom pairs involving the pairing of strands $\beta 5$ with $\beta 1$ and $\beta 3$ in the folded state (blue) and the I-state (red). For each time point in the trajectory, the Rms average $H^{N}-O$ donor-acceptor distance for folded-state H-bonding (blue) and I-state H-bonding (red) is shown. Thus, a jump in the blue distances accompanied by a drop in the red values corresponds to a transition from the native to the $\beta 5$ -retracted I-state. (B) Detail of three periods identified as being more similar to the $\beta 5$ -retracted state than to the native state, marked by black boxes in A. (C) Overlay of a representative structure from each cluster of intermediate-like conformations identified in the MD trajectory (SI Appendix, Fig. S10 for expanded plots).

states for three proteins: GTT, NTL9, and protein G. As seen for ubiquitin, populations of these out-of-register states were very low, in the range of 0.1 to 1.5%, and these authors suggested that such register shifts may be a general feature in the folding process of beta topologies. Indeed, rather than a hindrance to protein folding, a mutational study of protein G showed that the design of an out-of-register intermediate could accelerate the folding process by more than an order of magnitude (54).

Our pressure-jump NMR study unveils the structure of a metastable conformation of a protein with a lifetime much too short for full structural characterization by traditional techniques. These results highlight the potential of pressure-jump NMR for increasing our understanding of the protein folding process and for exploring the conformational space of such transiently populated states. While the very high stability of wild-type ubiquitin, which has an activation energy for unfolding of *ca* 7 kcal/mol (28, 55), necessitated the introduction of pressure-sensitizing mutations, the total void volume in many other wild-type proteins suffices to enable pressure-induced unfolding. Alternatively, void-generating mutations such as used in our study can be readily introduced to lower

the hydrostatic pressure required for unfolding (22). As demonstrated here, conventional structural parameters such as chemical shifts, NOE contacts, and residual dipolar couplings then become readily accessible by pressure-jump NMR spectroscopy, thereby providing a powerful method to gain atomic level structural detail on short-lived protein-folding intermediate states. Analogous studies are likely to yield detailed insights into pressure-sensitive excited states and folding intermediates of nucleic acids as well (56).

Materials and Methods

Protein Expression and Purification. Two codon-optimized plasmids, encoding residues 1 to 76 of human ubiquitin with V17A/V26A and L50A mutations (ATUM), were transformed into *E. coli* BL21(DE3) competent cells (New England Biolabs). For protein expression, all bacterial growth media contained 100 μ g/L of ampicillin, with M9 minimal media suitable for generating $U-^{13}C/^{15}N$ protein, or $U-^2H/^{15}N$ protein, following a previously described protocol (36). Purification also followed this previously described protocol, including a final reverse-phase HPLC step using a Vydac 214TP C4 column, which was necessary to eliminate trace protease contamination for which pressure-unfolded ubiquitin is a vulnerable substrate (21).

Pressure-Jump Apparatus. A description of the home-built pressure-jump apparatus has been reported previously (21). Briefly, a 2.8-mm inner diameter zirconia sample cell (57), rated for static pressures of up to 3 kbar (Daedalus, Inc.), was connected by a high-pressure hydraulic fluid (mineral spirit) transfer line via a manifold to two pneumatically controlled hydraulic valves. These valves connect to either an adjustable high-pressure reservoir (≤ 3 kbar) or an atmospheric pressure vessel. Spectrometer-generated logic signals control the valves, enabling to alternately pressurize and depressurize the protein sample. Pressure transition times were adjusted by flow restrictors to ca 2 ms for most measurements but lengthened to 4 ms for the RDC measurements.

NMR Spectroscopy. NMR experiments were recorded on Bruker 600-, 800-, and 900-MHz NMR spectrometers equipped with cryogenic probeheads that contained pulsed field gradient accessories. NMR samples contained 0.3 mM VA2 (or L50A) ubiquitin in 200 μ L solvent, 98% H₂O/2% D₂O, with 25 mM potassium phosphate buffer pH 6.4. Sample concentrations were limited to 0.3 mM concentrations to prevent off-pathway oligomerization during the folding process. For pressure-jump RDC measurements at 900 MHz, samples additionally contained 10 mg/mL PF1 and 200 mM NaCl. The sample temperature was regulated at 298 K during the high-pressure periods in each experiment, but it temporarily decreased by ca 2 K following each pressure drop due to adiabatic decompression. All pressure-jump 3D NMR spectra, including those for NOE and RDC measurements, were recorded using nonuniformly sampling protocols, and spectra were reconstructed using the SMILE software (58), integrated into the NMRPipe data processing package (59), and analyzed with CCPN (60) and Sparky (61) software. Distance restraints for the I-state were derived from NOE cross-peaks that were observed exclusively in the I-state spectra. Since the I-state is short lived, even short distances do not give rise to strong NOEs, and therefore the presence of an NOE was simply converted to an upper distance limit of 3.5 Å for the CS-Rosetta calculations.

CS-Rosetta Calculations. Multiple rounds of CS-Rosetta calculations were carried out with increasing levels of experimental restraints (62). First, using only the chemical shifts, a preliminary structure was obtained in which the chain

converged to a well-defined fold exhibiting the retracted $\beta 5$ strand. Second, a round that also included sparse NOE restraints was carried out. 34 NOE cross-peaks from the 3D pressure-jump NOESY experiment that were absent in the static, low-pressure NOESY spectrum were manually assigned and included in the CS-Rosetta input (63). Third, a final round of CS-Rosetta calculations also included I-state RDCs. Such RDCs were only available for 44 resonances where the I-state ¹⁵N chemical shifts differed sufficiently from both the native and unfolded states to avoid peak overlap. The 10 lowest-energy models are available in the PDB under entry 9PL1.

UMAP Embedding. UMAP embedding was generated using an Euclidean distance metric, a neighborhood size of 10, a minimum interpoint distance of 0.1, and a minimum cluster size of 100. For visualization, frames were projected onto two dimensions with UMAP and colored based on their cluster assignments (SI Appendix, Fig. S10A) (38).

Data, Materials, and Software Availability. Coordinates of the ubiquitin-folding intermediate structure have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) under accession code 9PL1 (64). NMR chemical shifts have been deposited in the BioMagResBank under accession code: 53268 (65). Pulse sequence code and parameters are available at <https://doi.org/10.5281/zenodo.16125212> (63). All study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. We thank Venkatraman Ramanujam for preparing several protein samples, James Baber and Yang Shen for assistance with experiments and analysis, and William A. Eaton, Robert Best, G. Marius Clore, Kresten Lindorff-Larsen, and Dennis A. Torchia for helpful discussions. This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) within the NIH. The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

1. C. M. Dobson, A. Sali, M. Karplus, Protein folding: A perspective from theory and experiment. *Angew. Chem.-Int. Ed.* **37**, 868–893 (1998).
2. D. M. Korzhnev, L. E. Kay, Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Acc. Chem. Res.* **41**, 442–451 (2008).
3. C. Gladkova *et al.*, An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. *EMBO J.* **36**, 3555–3572 (2017).
4. Y. O. Kamatari, S. Yokoyama, H. Tachibana, K. Akasaka, Pressure-jump NMR study of dissociation and association of amyloid protofibrils. *J. Mol. Biol.* **349**, 916–921 (2005).
5. I. Ritsch, H. J. Dyson, P. E. Wright, Initiation of transthyretin aggregation at neutral pH by fluid agitation. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2425230122 (2025).
6. 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).
7. S. W. Englander, L. Mayne, Protein folding studied using hydrogen-exchange labeling and 2-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243–265 (1992).
8. H. J. Dyson, P. E. Wright, How does your protein fold? Elucidating the apomyoglobin folding pathway. *Acc. Chem. Res.* **50**, 105–111 (2017).
9. 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).
10. C. A. Royer, Probing protein folding and conformational transitions with fluorescence. *Chem. Rev.* **106**, 1769–1784 (2006).
11. H. J. Dyson, P. E. Wright, H. A. Scheraga, The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13057–13061 (2006).
12. T. P. Causgrove, R. B. Dyer, Nonequilibrium protein folding dynamics: Laser-induced pH-jump studies of the helix-coil transition. *Chem. Phys.* **323**, 2–10 (2006).
13. K. Schlepckow, J. Wimmer, A. Bachmann, T. Kiefhaber, H. Schwalbe, Conserved folding pathways of alpha-lactalbumin and lysozyme revealed by kinetic CD, fluorescence, NMR, and interrupted refolding experiments. *J. Mol. Biol.* **378**, 686–698 (2008).
14. A. J. Wirth, Y. X. Liu, M. B. Prigozhin, K. Schulten, M. Gruebele, Comparing fast pressure jump and temperature jump protein folding experiments and simulations. *J. Am. Chem. Soc.* **137**, 7152–7159 (2015).
15. C. Dumont, T. Emilsson, M. Gruebele, Reaching the protein folding speed limit with large, sub-microsecond pressure jumps. *Nat. Methods* **6**, 515–U570 (2009).
16. S. W. Englander, L. Mayne, The case for defined protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8253–8258 (2017).
17. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein-folding - A synthesis. *Proteins-Struct. Funct. Genet.* **21**, 167–195 (1995).
18. K. A. Dill, H. S. Chan, From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
19. W. A. Eaton, P. G. Wolynes, Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9759–E9760 (2017).
20. M. M. G. Krishna, H. Maity, J. N. Rumbley, S. W. Englander, Branching in the sequential folding pathway of cytochrome c. *Protein Sci.* **16**, 1946–1956 (2007).
21. 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**, 201803642 (2018).
22. J. Roche *et al.*, Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6945–6950 (2012).
23. M. B. Prigozhin *et al.*, Misplaced helix slows down ultrafast pressure-jump protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8087–8092 (2013).
24. G. R. Hedwig, Partial molar volumes of the amino-acid side-chains of proteins in aqueous solution: Some comments on their estimation using partial molar volumes of amino acids and small peptides. *Biopolymers* **32**, 537–540 (1992).
25. M. S. Briggs, H. Roder, Early hydrogen-bonding events in the folding reaction of Ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2017–2021 (1992).
26. B. Brutscher, R. Bruschweiler, R. R. Ernst, Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by H-1, C-13, and N-15 nuclear magnetic resonance spectroscopy. *Biochemistry* **36**, 13043–13053 (1997).
27. S. Garcia-Manjares, L. Dougan, C. L. Badilla, J. Bruijic, J. M. Fernandez, Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10534–10539 (2009).
28. S. Khorasanizadeh, I. D. Peters, T. R. Butt, H. Roder, Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054–7063 (1993).
29. S. Khorasanizadeh, I. D. Peters, H. Roder, Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **3**, 193–205 (1996).
30. B. A. Krantz, T. R. Sosnick, Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696–11701 (2000).
31. A. Vallee-Belisle, S. W. Michnick, Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol.* **374**, 791–805 (2007).
32. A. Vallee-Belisle, S. W. Michnick, Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat. Struct. Mol. Biol.* **19**, 731 (2012).
33. C. Charlier, J. M. Courtney, T. R. Alderson, P. Anfinsen, A. Bax, Monitoring N-15 chemical shifts during protein folding by pressure-jump NMR. *J. Am. Chem. Soc.* **140**, 8096–8099 (2018).
34. T. R. Sosnick, R. S. Dothager, B. A. Krantz, Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17377–17382 (2004).
35. S. Piana, K. Lindorff-Larsen, D. E. Shaw, Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5915–5920 (2013).
36. E. Masoumzadeh, J. F. Ying, J. L. Baber, P. Anfinsen, A. Bax, Proline peptide bond isomerization in Ubiquitin under folding and denaturing conditions by pressure-jump NMR. *J. Mol. Biol.* **436**, 168587 (2024).
37. C. Charlier, J. M. Courtney, P. Anfinsen, A. Bax, Interrupted pressure-jump NMR experiments reveal resonances of on-pathway folding intermediate. *J. Phys. Chem. B* **122**, 11792–11799 (2018).

38. 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).
39. 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).
40. L. E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient-enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).
41. N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
42. Y. N. Fu, A. J. Wand, Partial alignment and measurement of residual dipolar couplings of proteins under high hydrostatic pressure. *J. Biomol. NMR* **56**, 353–357 (2013).
43. N. C. Fitzkee, A. Bax, Facile measurement of H-1-N-15 residual dipolar couplings in larger perdeuterated proteins. *J. Biomol. NMR* **48**, 65–70 (2010).
44. A. S. Maltsev, A. Grishaev, J. Roche, M. Zaslloff, A. Bax, Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. *J. Am. Chem. Soc.* **136**, 3752–3755 (2014).
45. A. F. Schubert *et al.*, Structure of PINK1 in complex with its substrate ubiquitin. *Nature* **552**, 51 (2017).
46. X. Dong *et al.*, Ubiquitin S65 phosphorylation engenders a pH-sensitive conformational switch. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6770–6775 (2017).
47. 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**, 496–502 (2008).
48. N. J. Marianayagam, S. E. Jackson, The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys. Chem.* **111**, 159–171 (2004).
49. G. Reddy, D. Thirumalai, Collapse precedes folding in denaturant-dependent assembly of Ubiquitin. *J. Phys. Chem. B* **121**, 995–1009 (2017).
50. K. Lindorff-Larsen *et al.*, Systematic validation of protein force fields against experimental data. *PLoS One* **7**, e32131 (2012).
51. M. Ester, H. P. Kriegel, J. Sander, X. W. Xu "A density-based algorithm for discovering clusters in large spatial databases with noise" in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, E. Simoudis, J. Han, U. M. Fayyad, Eds. (AAAI Press, 1996), pp. 226–231.
52. D. M. Korzhnev *et al.*, Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586–590 (2004).
53. K. A. Beauchamp, R. McGibbon, Y. S. Lin, V. S. Pande, Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17807–17813 (2012).
54. S. Nauli, B. Kuhlman, D. Baker, Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8**, 602–605 (2001).
55. C. L. Chyan *et al.*, Reversible mechanical unfolding of single ubiquitin molecules. *Biophys. J.* **87**, 3995–4006 (2004).
56. B. Harish *et al.*, Hidden intermediates in Mango III RNA aptamer folding revealed by pressure perturbation. *Biophys. J.* **121**, 421–429 (2022).
57. R. W. Peterson, A. J. Wand, Self-contained high-pressure cell, apparatus, and procedure for the preparation of encapsulated proteins dissolved in low viscosity fluids for nuclear magnetic resonance spectroscopy. *Rev. Sci. Instrum.* **76**, 094101 (2005).
58. J. Ying, F. Delaglio, D. A. Torchia, A. Bax, Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **68**, 101–118 (2017).
59. F. Delaglio *et al.*, NMRpipe - A multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
60. S. P. Skinner *et al.*, Ccpnmr analysisassign: A flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66**, 111–124 (2016).
61. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-sparkly: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics (Oxford, England)* **31**, 1325–1327 (2015).
62. Y. Shen, R. Vernon, D. Baker, A. Bax, De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**, 63–78 (2009).
63. E. Masoumzadeh *et al.*, Structure of a transient protein folding intermediate by pressure-jump NMR. Zenodo. <https://doi.org/10.5281/zenodo.16125212>. Deposited 19 July 2025.
64. E. Masoumzadeh, J. Courtney, A. Bax, Structure of metastable folding intermediate of V17A/V26A ubiquitin derived from pressure-jump NMR data. Protein Data Bank. <https://doi.org/10.2210/pdb9pl1/pdb>. Deposited 15 July 2025.
65. E. Masoumzadeh, J. Courtney, A. Bax, Ubiquitin V17A/V26A retracted intermediate. Biological Magnetic Resonance Data Bank. https://bmr.io/data_library/summary/?bmrld=53268. Deposited 10 July 2025.