PNAS

Supporting Information for

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

Elahe Masoumzadeh, Joseph M. Courtney, Cyril Charlier, Jinfa Ying, Philip Anfinrud, and Adriaan Bax

This PDF file includes:

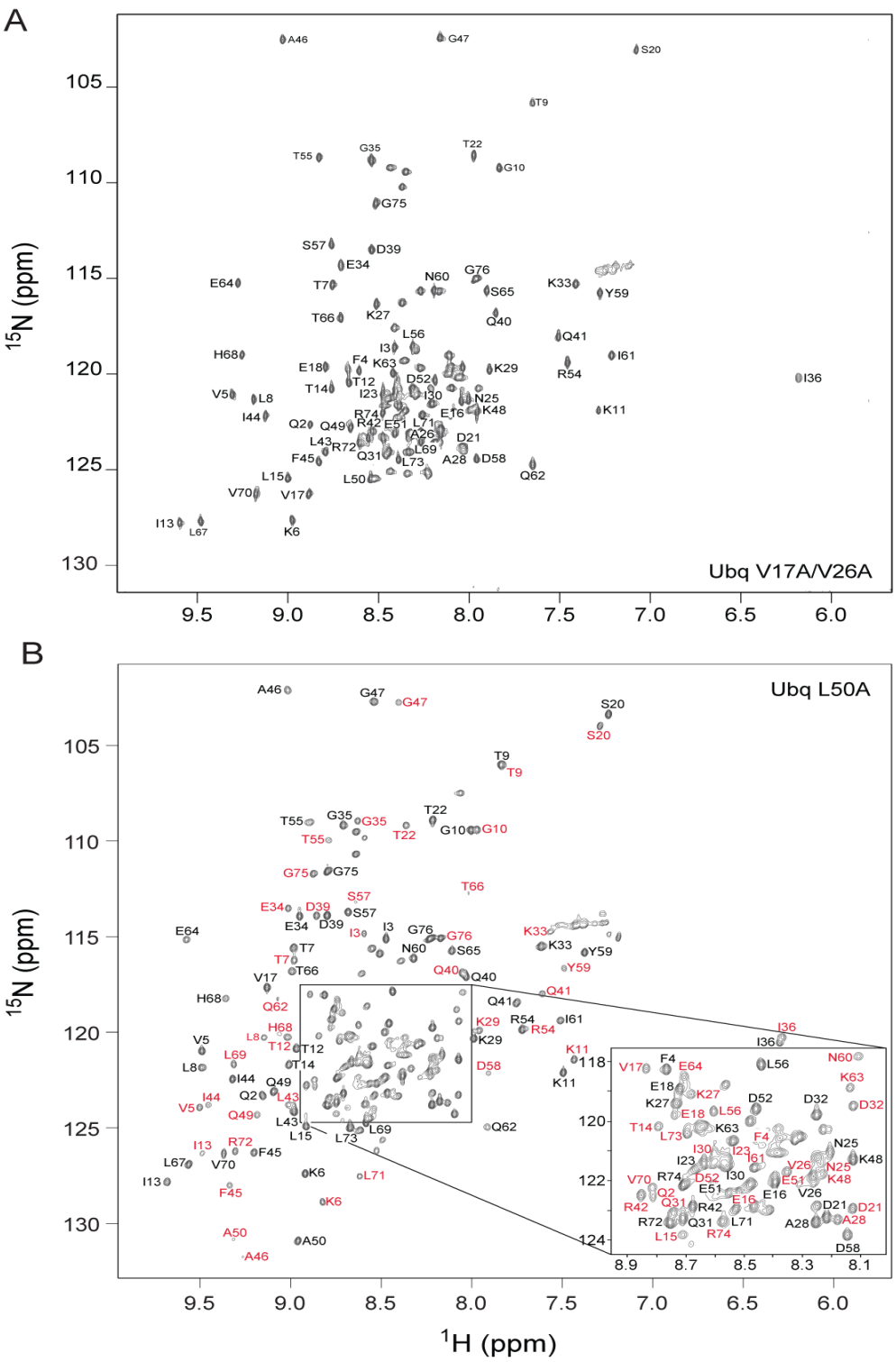

Fig. S1. 900-MHz PJ-HSQC spectra of (A) V17A/V26A ubiquitin at 298K and (B) L50A ubiquitin, at 278 K, both recorded with parameters similar to those in Fig. 1, main text. (A) VA2 ubiquitin PJ-HSQC spectrum with ^{15}N evolution started 500 ms after the drop in pressure, after I-state conformers had transitioned to the native state. (B) L50A ubiquitin PJ-HSQC spectrum with ^{15}N evolution started 40 ms after the pressure drop with I-state resonances remaining present (red labels) in addition to natively folded (black), and unfolded state peaks (unlabeled).

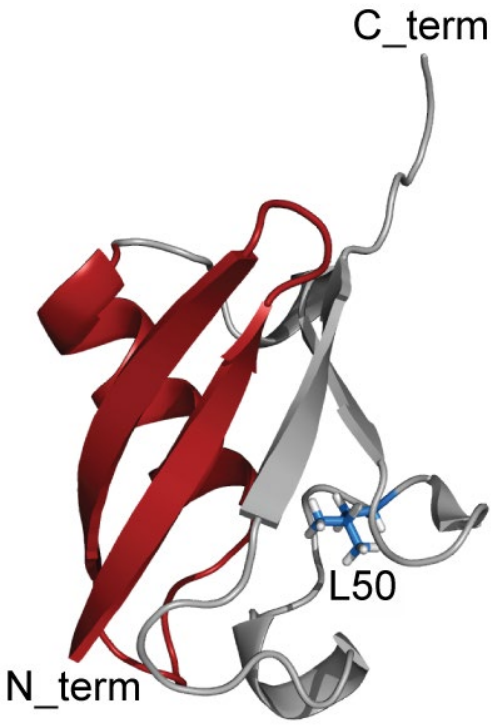

Fig. S2. Location of residue L50 (in blue stick representation) relative to the folding nucleus of ubiquitin (red) for the native state solution structure (2MJB).

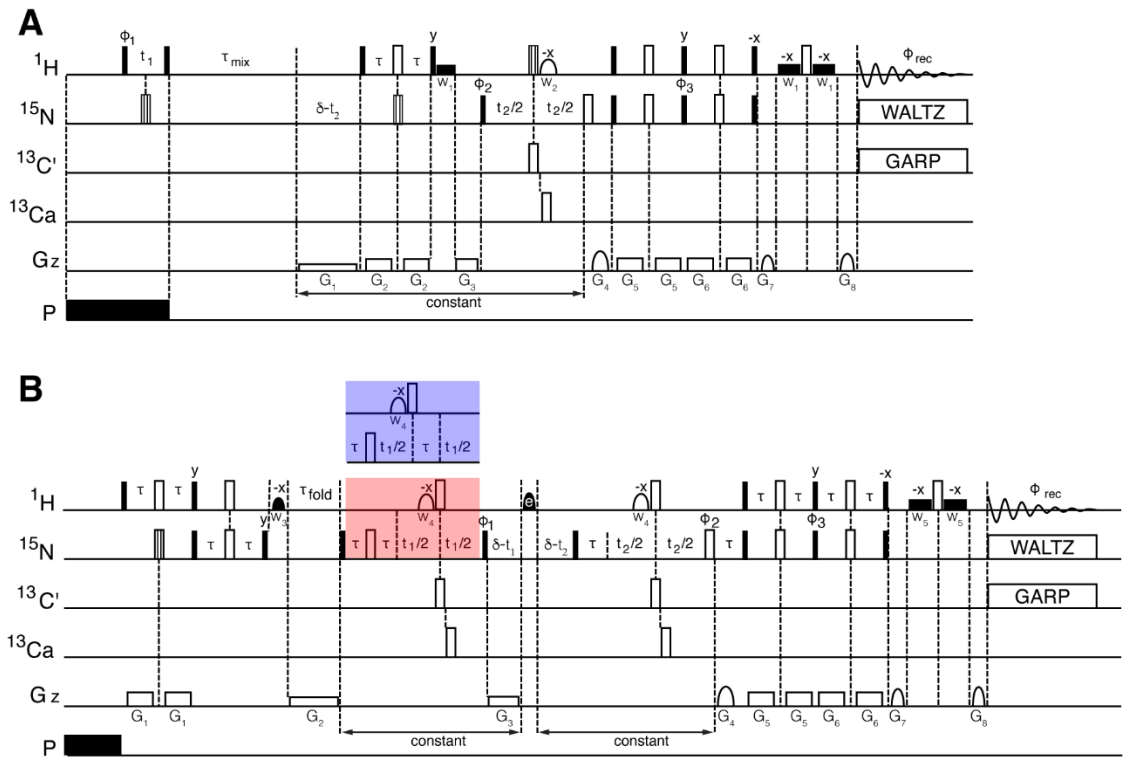

Fig. S3. Pulse sequence diagrams for (A) pressure-jump NOESY and (B) pressure-jump ARTSY experiments. In both experiments, the filled and open rectangular bars represent 90º and 180º pulses, while the vertically hatched bars denote composite 180º pulses. The filled (90º) and open (180º) rectangular boxes and bells are water flip-back pulses, labeled as w_1 (1.2 ms, rectangular 90º), w_2 (2 ms, sinc-shaped 180º), w_3 (1.5 ms, sinc-shaped 90º), w_4 (1.491 ms, sinc-shaped 180º), and w_5 (1.2 ms, rectangular 90º). The filled shape labeled 'e' denotes an amide selective Eburp2 pulse (1.6 ms with an offset of 2000 Hz from the water resonance at 600 MHz). All pulses were applied along x unless indicated otherwise. The filled boxes in the pressure channel represent durations where the sample is at a high pressure (2.5 kbar). Delay τ in both experiments is 2.7 ms. Delay δ in both experiments is 50 ms, longer than the indirect acquisition times and being decremented with t_1 or t_2 to ensure the time between the low-pressure valve opening and the beginning of ^{1}H detection was kept constant during the data acquisition. The NOE mixing period, τ_{mix}, in the NOESY experiment was 150 ms and delay τ_{fold} in the ARTSY experiment was 82.1 ms. Gradient pulses were either rectangular or sine-shaped, as depicted in the diagrams. In both experiments, WALTZ16 and GARP CPD decoupling schemes were applied to the ^{15}N and ^{13}C' nuclei, respectively, during direct detection of amide protons.

For the NOESY experiment in (A), the durations and strengths of the gradients are listed as follows: $G_{1,2,3,4,5,6,7,8}$ = 50 ms - t_2, 2.6, 1.0, 2.0, 2.6, 2.6, 0.5, and 0.5 ms, z-strengths of 2.1, 1.33, 2.52, 34.545, 0.532, 0.798, 17.5, and 31.5 G/cm. Quadrature detection in t_2 was achieved using the echo-antiecho scheme by inverting the encoding gradient G_4 together with ϕ_3 to obtain the second FID for every t_2 increment. Quadrature in the t_1 dimension was recorded using States-TPPI with ϕ_1 being incremented by 90°, and phase cycling: ϕ_1 = x, -x; ϕ_2 = x, x, -x, -x; ϕ_3 = y and ϕ_{rec} = x, -x, -x, x. The data was recorded using unweighted non-uniform sampling in the indirect (t_1 and t_2) dimensions, with 16.4% randomly selected t_3 FIDs from the full data matrix, consisting of 1024* (^{1}H, t_3, 121.7 ms) x 80* (^{15}N, t_2, 40.0 ms) x 150* (^{1}H, t_1, 24.0 ms) hypercomplex points. With an interscan delay of 7.5 s and 2 scans per FID, the total time for the data collection was approximately 35.5 hours.

For the ARTSY experiment in (B), the reference and the attenuated experiments using the ARTSY encoding schemes shaded in red and blue, respectively, were applied in an interleaved manner. Durations and strengths of the gradients were: $G_{1,2,3,4,5,6,7,8}$ = 2.6, 82.0, 50 ms - t_1, 2.0, 2.6, 2.6, 0.5,

and 0.5 ms, with z-strengths of 1.33, 1.33, 1.33, 34.545, 0.532, 0.798, 21.0, and 35.0 G/cm. Quadrature detection in t_2 was achieved using the echo-antiecho scheme by inverting the encoding gradient G_4 together with ϕ_3 to obtain the second FID for every t_2 increment. The t_1 dimension was recorded using States-TPPI with ϕ_1 being incremented by 90°, and phase cycling: ϕ_1 = x, -x; ϕ_2 = x, x, y, y, -x, -x, -y, -y; ϕ_3 = y; and ϕ_{rec} = x, -x, -x, x. The data was recorded using unweighted non-uniform sampling in the indirect (t_1 and t_2) dimensions, with 17.3% randomly selected t_3 FIDs from the full data matrix, consisting of 1024* (^{1}H, t_3, 121.7 ms) x 96* (^{15}N, t_2, 46.0 ms) x 96* (^{15}N, t_1, 46.0 ms) hypercomplex points. With an interscan delay of 8.0 s and 4 scans per FID, the total time for recording the two interleaved data sets was approximately 123 hours.

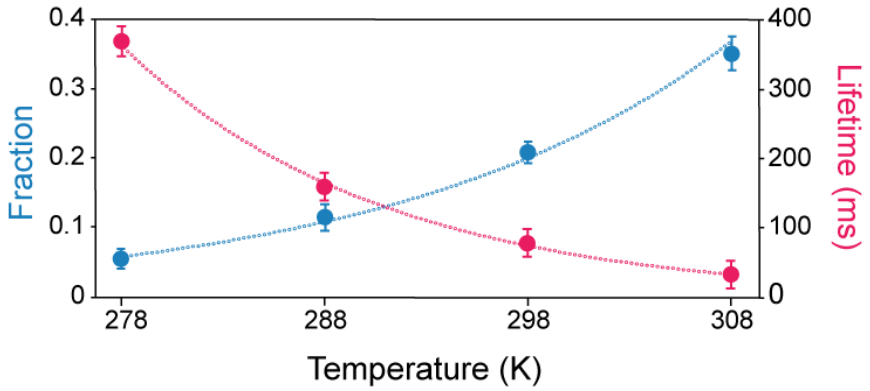

Fig. S4. Maximum I-state fraction during folding as a function of temperature (blue), showing that a greater maximum I-state population is reached at higher temperature, but with decreased lifetimes (red). The maximum I-state fraction reached was determined from the relative intensities of the upfield shifted methyl group resonance of L50 in the native and retracted strand conformations (1).

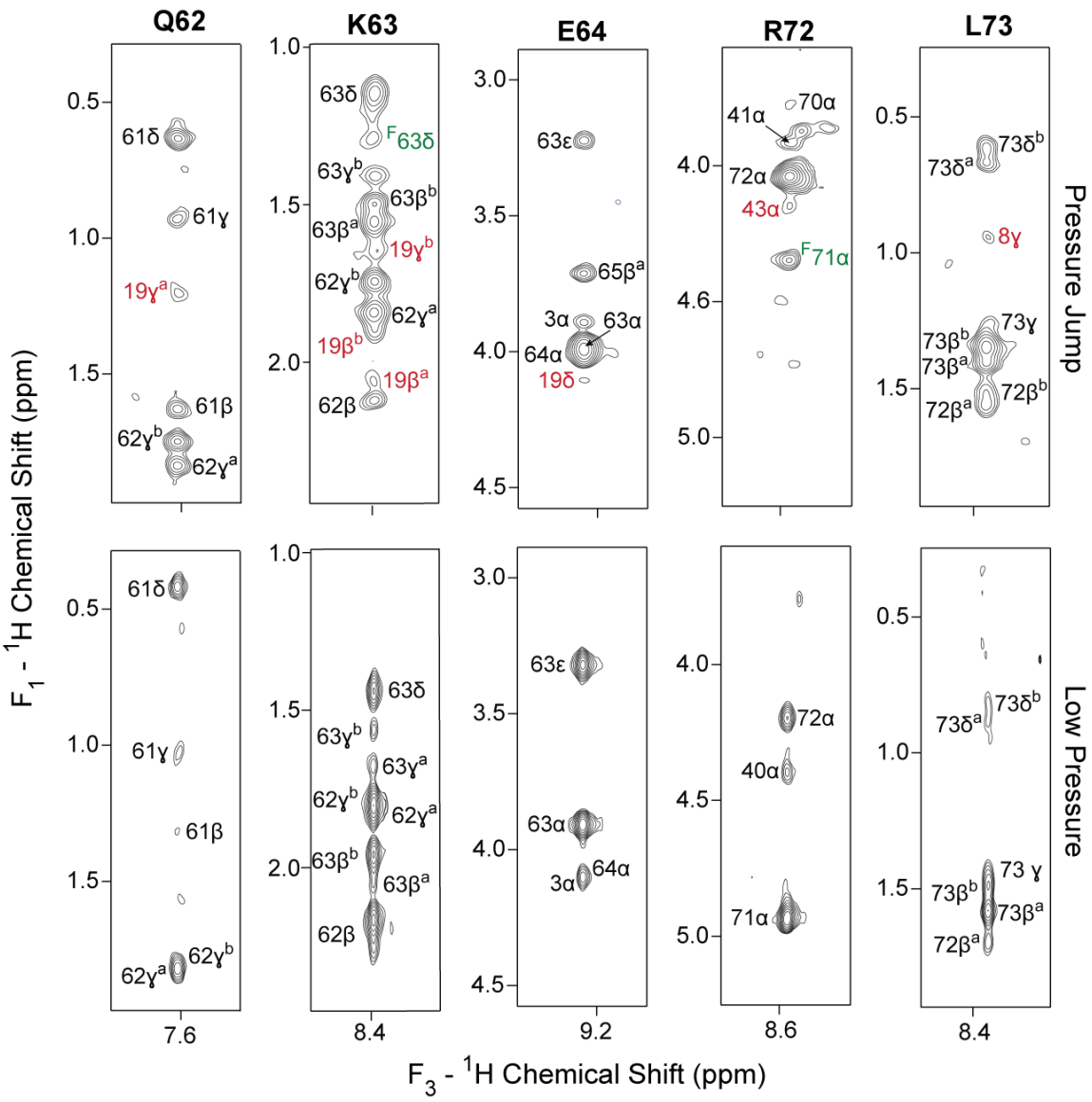

Fig. S5. Additional NOE cross peaks observed in 3D pressure-jump (top row) and corresponding, static low pressure (bottom row) NOESY spectra. The amide proton serving as the source of magnetization is indicated by the column labels. Direct dimension (t_3) chemical shifts were measured at 1 bar in the folded state (both experiments). For the pressure-jump experiment, chemical shifts in the F_1 dimension were measured at 2.5 kbar, causing the strongest peak to be an exchange peak to the unfolded, 2.5 kbar chemical shift of the same amide ^{1}H. Black and red labels indicate peaks resulting from native state and I-state contacts, respectively. The green labels with F indicate peaks associated with the residual natively folded state, resulting from incomplete unfolding of ubiquitin during the 8-second interscan delay at 2.5 kbar. Red labels correspond to resonance positions of contacts identified in the intermediate state but not observed in natively folded state (bottom row). For the pressure jump experiment, chemical shifts in the t_1 dimension evolved at 2.5 kbar, causing the strongest peak to be an exchange peak to the unfolded, 2.5 kbar chemical shift of the same amide ^{1}H. Other strips showing I state NOEs are available at https://doi.org/10.5281/zenodo.16125212

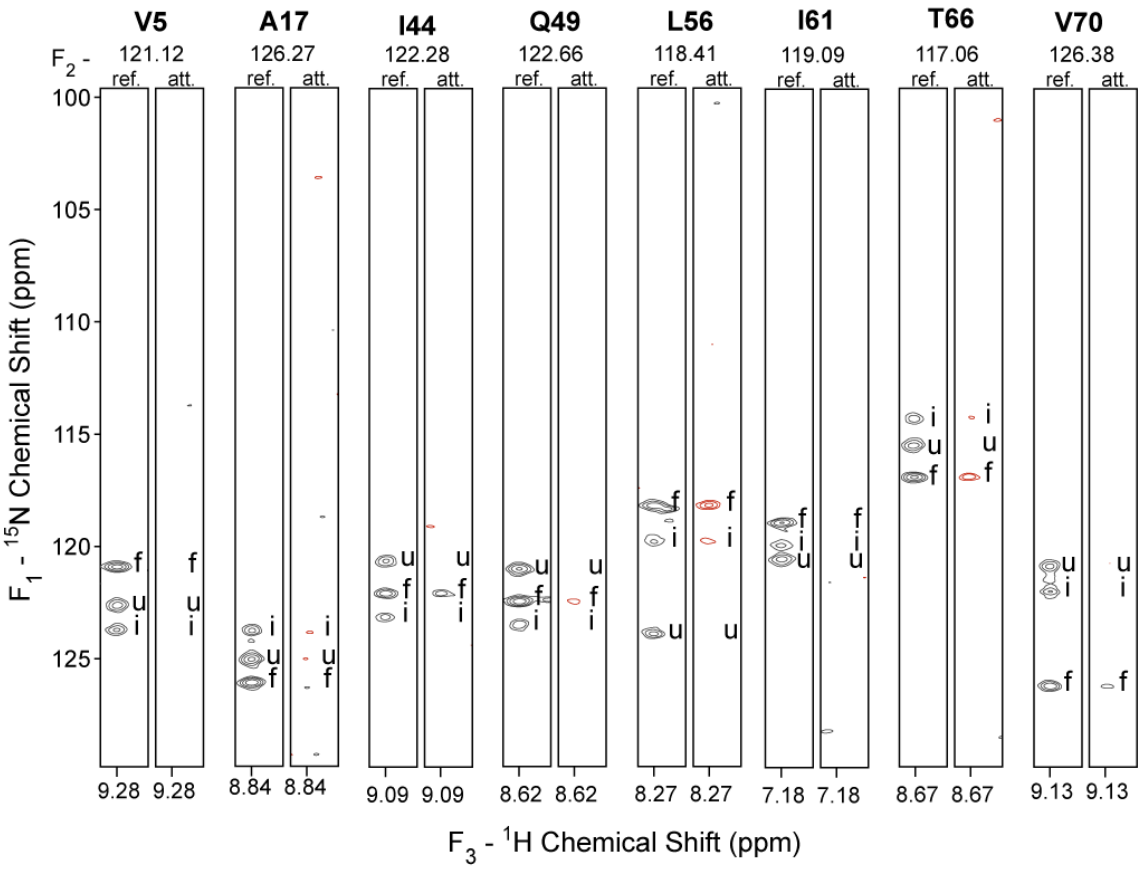

Fig. S6. Examples of pairs of strips through a 3D pressure-jump ARTSY spectrum, used for measuring RDCs of the I- state. The amide proton serving as the source of magnetization is indicated by column labels for each pair. The left strip in each pair is taken from the reference spectrum (J+D rephasing duration of 5.35 ms); the right strip from the attenuated spectrum (J+D rephasing duration of 10.7 ms). Direct dimension (F_3) chemical shifts were measured at 1 bar in the folded state (both spectra). The F_2 dimension corresponds to ^{15}N evolution, initiated 60 ms after the pressure-jump, and contains resonances for the unfolded, intermediate and folded species, marked u, i, and f, respectively, modulated by the coupling strength, $^1J_{NH} + {}^1D_{NH}$. Black and red contours correspond to positive and negative intensities, respectively. A full set of strips is available at https://doi.org/10.5281/zenodo.16125212

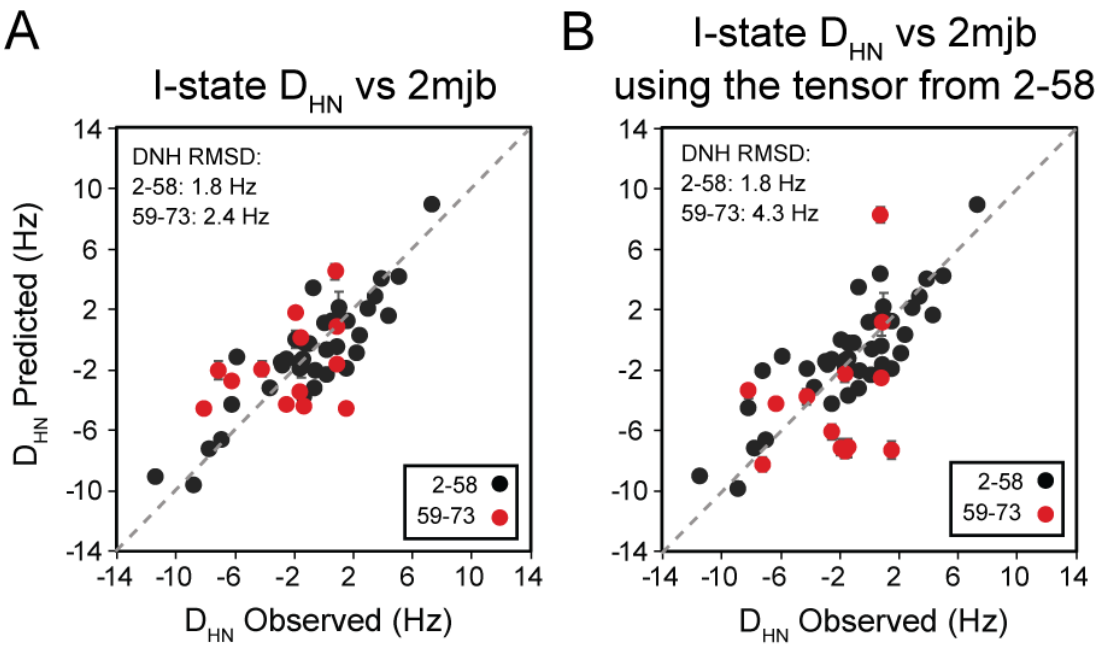

Fig. S7. Comparison of the 900 MHz RDCs observed in the P-jump ARTSY experiment for the I-state to values predicted for the native WT structure (2MJB). (A) RDCs of residues 2–58 (black dots) and 59-73 (red dots) were fitted separately and (B) the alignment tensor of residues 2–58 was used to back-calculate the RDC values for residues 59–73.

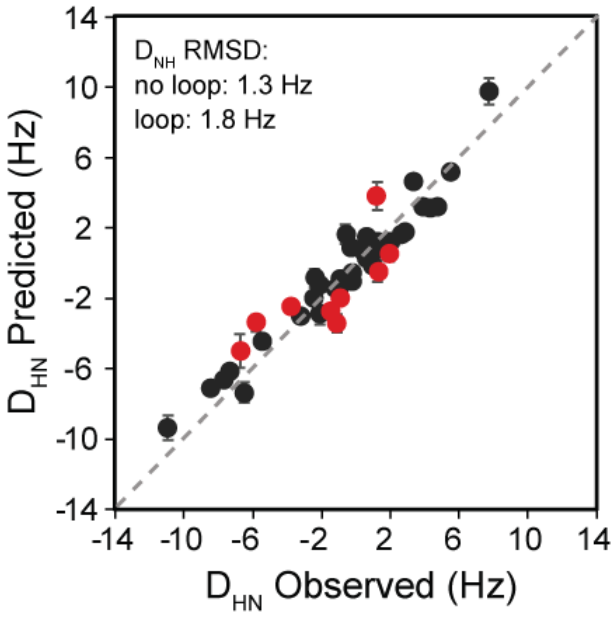

Fig. S8. Jackknifed cross-validation of the I-state structure using chemical shifts, NOE, and RDC data for the S65 loop. Eight separate CS-Rosetta calculations were carried out, each time cyclically leaving out one of the eight amide RDCs in the S65 loop. With an RMSD of 1.8 Hz, these 'left-out' RDCs (residues 59-69, red), predicted by the corresponding CS-Rosetta models, showed good agreement with the models. Black dots represent the RDC fit of all residues excluding the S65-loop to the experimental values. The RMSD overall RDCs equals 1.4 Hz.

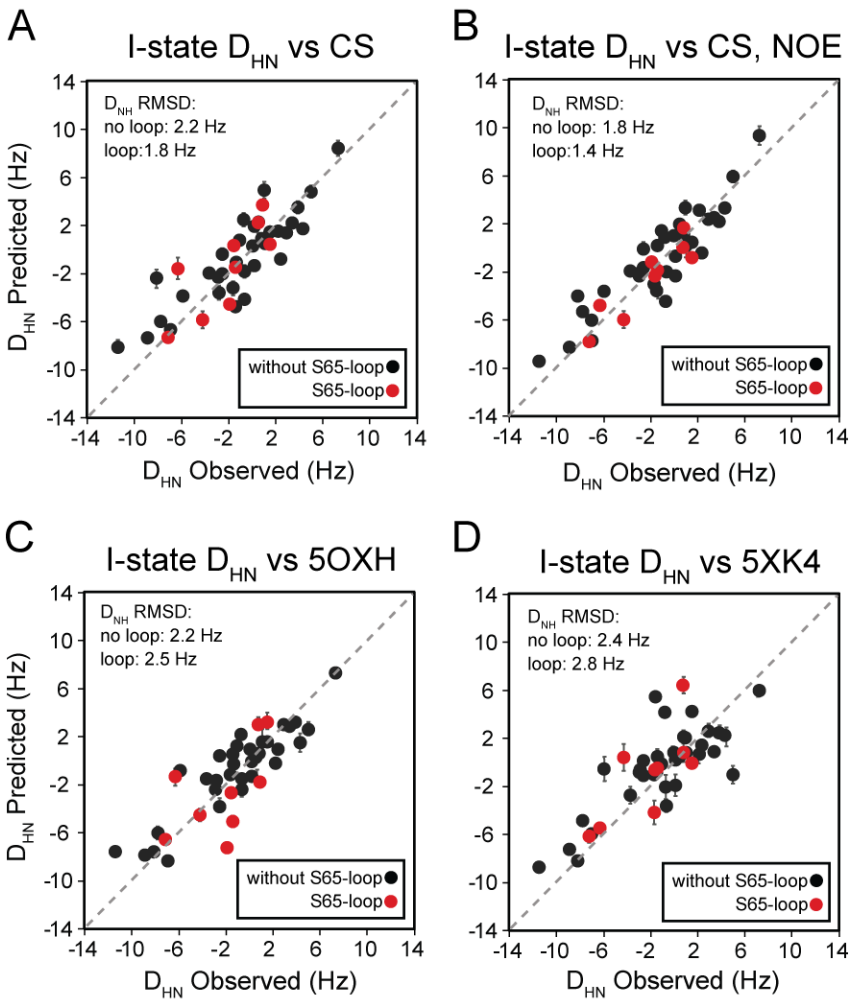

Fig. S9. Comparison of the RDCs observed for the I-state to values predicted for different I-state models generated by: (A) I-state chemical shifts, (B) chemical shifts and NOE data, (C) the crystal structure with S65 phosphorylated (5OXH), and (D) the NMR structure with S65 phosphorylated (5XK4). The RDC values for the S65 loop residues (59–69, red dots) were predicted separately using the alignment tensor obtained from fitting the remaining residues excluding the loop (2–58 and 70–73, black dots).

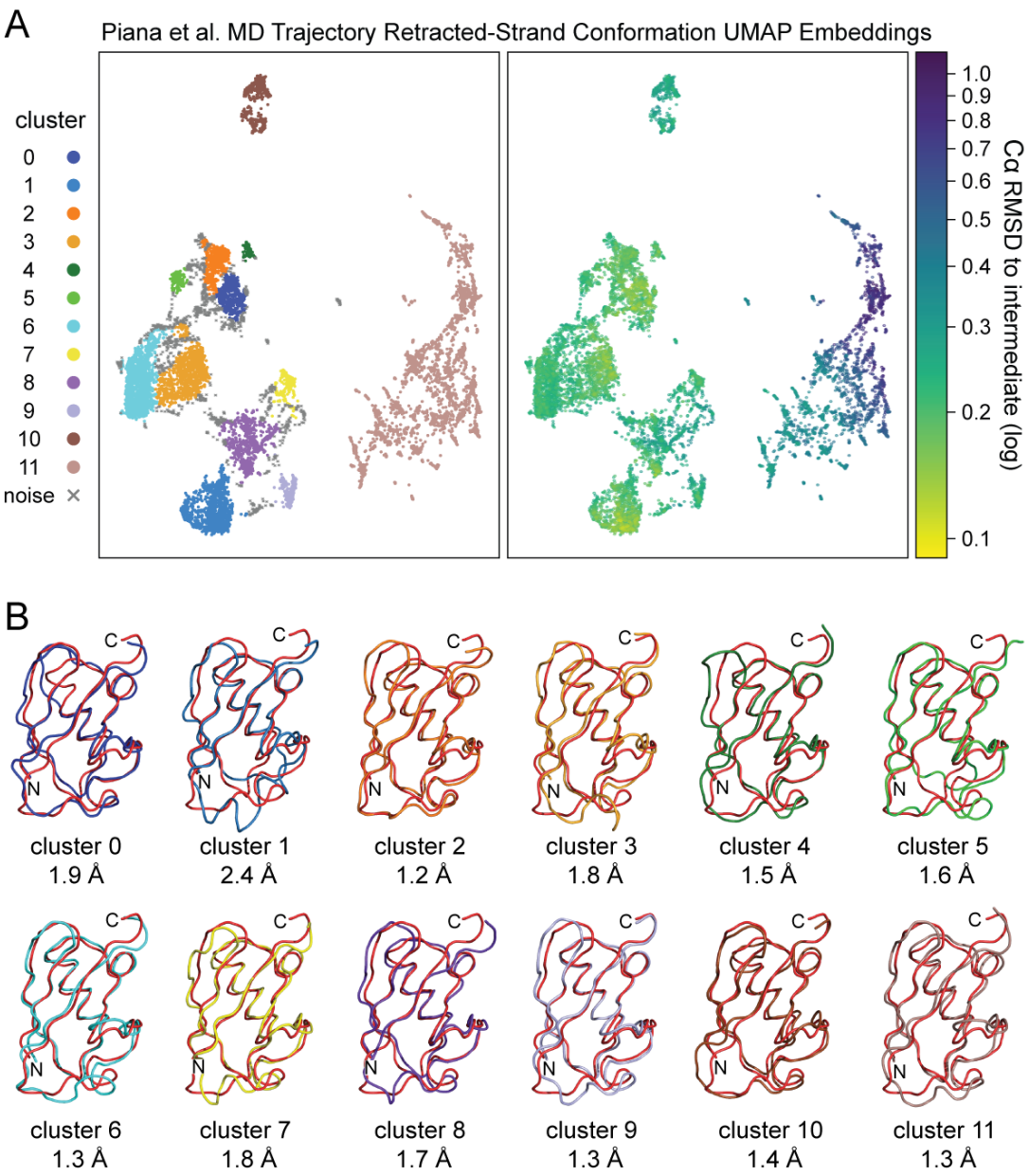

Fig. S10. Cluster analysis of Piana *et al.* trajectory frames (2) where H-bonding was more similar to the I-state than to the folded state. Frames were projected onto two dimensions with the UMAP algorithm. (A) Each dot indicates a single frame. Projection of structures to 2D were clustered using DBSCAN and dots are colored by cluster ID (left panel). The same projected frames, colored by backbone C^{α} RMSD to the I-state, are shown in the right panel, with the vertical axis displaying the C^{α} RMSD to the I-state on a log scale in units of nm. (B) Representative structures of each cluster (color-coded based on panel A, left) overlaid with the I-state structure (colored red). C^{α} backbone RMSD is indicated below each overlay.

Table S1. Experimental $^1D_{NH}$ RDC values of VA2 ubiquitin for the native fold, I-state, and unfolded state, at 1 bar in the pressure-jump ARTSY spectrum.

	D (Hz)				D (Hz)		
residue #	native fold	I-state	unfolded	residue #	native fold	I-state	unfolded
Q2	1.1	5.1	1.9	Q41	-4.5	-11.4	-0.2
I3	-6.6	-5.9	-0.9	R42		-0.6	0.8
F4	-3.7	2.2	-0.1	L43	5.7		0.4
V5	-0.4	-1.4	-1.5	I44	9.2	2.5	-2.3
K6	4.5	4.4	-1.3	F45	9.7	0.2	0.3
T7	7.2		0.8	G47	4.9		-0.8
L8	1.0	-6.9	-1.7	K48	-0.2	3.5	-1.8
T9	-6.1		0.6	Q49		-2.8	-0.5
K11	3.6	1.0	-2.1	L50	9.4	3.0	1.7
T12	9.0	-1.3	-0.2	E51	0.6	7.4	-0.9
I13	0.0	-2.5	0.4	R54	1.5		0.8
T14	-0.5	0.6	1.3	T55	5.6	3.9	-1.4
L15	-5.7	-0.6	-3.6	L56	-9.2	-7.7	-0.7
A17	-2.1	-2.9	-0.1	S57	-10.9	-8.8	2.7
E18	2.1		-1.3	D58	-5.2	0.2	1.7
S20	-2.9	-1.0	-0.1	Y59		-6.3	
D21	2.1	-3.7	-0.7	N60	-1.3	-7.1	1.7
T22	-11.2		1.8	I61	1.9	-1.9	1.4
I23	4.0	0.1	1.1	Q62	1.4	-1.4	-0.4
K27	8.1	1.1	1.9	E64	-10.0	1.6	-0.9
K29	2.8			S65	-3.1		1.6
I30		1.6	2.7	T66	-8.1	-4.2	-0.4
Q31	7.1	-0.7		L67	-5.0	-2.7	-1.7
D32	6.5		-0.1	H68		0.9	
K33	3.0		-0.4	L69	8.2	-1.5	-1.2
E34	6.5		-4.6	V70	5.9	0.9	0.3
G35	-0.2		0.4	L71	-0.4	-1.6	0.1
I36	0.5		-2.4	R72	-3.1	-2.5	-0.7
D39	-7.7		0.4	L73	-4.8	-8.1	-1.8
Q40	-8.1		-0.1	G75	-0.9	-2.0	3.0

SI References

1. 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. USA* **115**, 201803642 (2018).
2. 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).