JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Nonnative Interactions in the FF Domain Folding Pathway from an Atomic Resolution Structure of a Sparsely Populated Intermediate: An NMR Relaxation Dispersion Study

Dmitry M. Korzhnev,[†,‡] Robert M. Vernon,[§] Tomasz L. Religa,[†] Alexandar L. Hansen,[†] David Baker,[§] Alan R. Fersht,[∥] and Lewis E. Kay[†,*]

[†]Departments of Molecular Genetics, Biochemistry and Chemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada

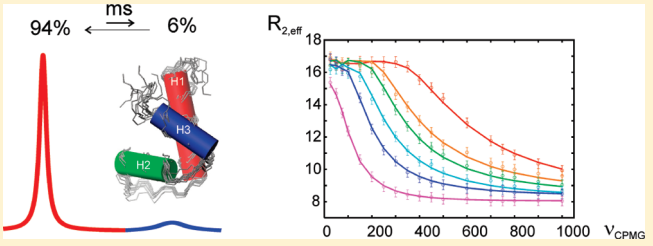
[‡]Department of Molecular, Microbial and Structural Biology, University of Connecticut Health Center, Farmington, Connecticut 06030, United States

[§]Department of Biochemistry, University of Washington, Box 357350, 1705 NE Pacific Street, Seattle, Washington 98195-7350, United States

[∥]Center for Protein Engineering, Medical Research Council, Cambridge, United Kingdom

Supporting Information

ABSTRACT: Several all-helical single-domain proteins have been shown to fold rapidly (microsecond time scale) to a compact intermediate state and subsequently rearrange more slowly to the native conformation. An understanding of this process has been hindered by difficulties in experimental studies of intermediates in cases where they are both low-populated and only transiently formed. One such example is provided by the on-pathway folding intermediate of the small four-helix bundle FF domain from HYPA/FBP11 that is populated at several percent with a millisecond lifetime at room temperature. Here we have studied the L24A mutant that has been shown previously to form nonnative interactions in the folding transition state. A suite of Carr−Purcell−Meiboom−Gill relaxation dispersion NMR experiments have been used to measure backbone chemical shifts and amide bond vector orientations of the invisible folding intermediate that form the input restraints in calculations of atomic resolution models of its structure. Despite the fact that the intermediate structure has many features that are similar to that of the native state, a set of nonnative contacts is observed that is even more extensive than noted previously for the wild-type (WT) folding intermediate. Such nonnative interactions, which must be broken prior to adoption of the native conformation, explain why the transition from the intermediate state to the native conformer (millisecond time scale) is significantly slower than from the unfolded ensemble to the intermediate and why the L24A mutant folds more slowly than the WT.



INTRODUCTION

The assumption that protein folding proceeds by the progressive formation of nativelike interactions has played an important role both in the development of theoretical models and in the interpretation of experimental studies of how protein molecules assume their three-dimensional structures.[1,2] For example, native-centric Go-like models have been used in computer simulations of protein folding,[2] while Φ-values that measure the effects of amino-acid substitutions on folding kinetics and thermodynamics are often analyzed in terms of the extent of formation of nativelike structure in the folding transition state.[1] Significant numbers of recent studies have, however, established that the mechanism of protein folding can be more complex than the simple accumulation of native contacts and that nonnative interactions often play an important role in controlling the kinetics and thermodynamics of the folding pathway.[3−9] Nonnative interactions may either assist the folding process by stabilizing transition states, so that barrier heights are lowered and hence folding rates are increased,[9−11] or trap the molecules in conformations that lead to a reduction in protein folding rates.[12] In cases where intermediates can be stabilized, these interactions have been studied in detail.[5,6] However, in many cases folding intermediates are populated at low levels, are only transiently formed, and cannot be stabilized, complicating their investigation by traditional structural biology methods.

Such "invisible" intermediate (excited) states can, nevertheless, be studied in detail by Carr−Purcell−Meiboom−Gill (CPMG) relaxation dispersion NMR spectroscopy in cases where they have fractional populations of 0.5% or greater and exchange with the "visible" folded (ground) state with rates between several hundred and several thousand per second.[13,14]

Received: April 21, 2011
Published: June 06, 2011

The exchange process leads to line-broadening of cross-peaks in NMR spectra of the ground state. In relaxation dispersion experiments, the resultant increase in transverse relaxation rates is monitored as a function of the strength of an applied radio frequency field to extract the kinetics and thermodynamics of the exchange reaction along with chemical shifts of the excited state(s).[15,16] In addition, bond vector orientations in the excited state can also be obtained, through measurement of residual dipolar couplings (RDCs) by a variant of the CPMG experiment in which molecules are weakly aligned in a static magnetic field.[17] In cases where detailed structural information about the excited state is desired, the measured chemical shift and RDC values are used as experimental restraints in computational protocols such as CS-Rosetta,[18] Cheshire,[19] and CS23D.[20] This approach provides an avenue, therefore, for the production of atomic resolution models of sparsely populated intermediates, facilitating the detailed analysis of interactions—both native and nonnative—in these states.

Recently we have used a chemical shift/RDC approach in concert with CS-Rosetta[18] to calculate the three-dimensional structure of an on-pathway folding intermediate of the wild-type (WT) FF domain from HYPA/FBP11.[21] This intermediate is populated at a level of 2−3% that of the native state with a lifetime of approximately 0.5 ms (25−30 °C).[22] While many features of the native (N) structure were preserved, a striking finding was the observation of a significant number of nonnative contacts that presumably stabilize the intermediate (I) structure. Previous studies have shown that the conversion from the intermediate to the native state is rate-limiting, close to 2 orders of magnitude slower than the transition from the unfolded (U) state to I.[21,23] The results from the structure of the WT intermediate suggest, but do not prove, that breaking non-native contacts prior to formation of the native state may be responsible for the slow step in folding. In an effort to further understand how such nonnative interactions in the I state might affect FF domain folding rates, to obtain additional evidence that such contacts are responsible for slowing folding, and to continue to elucidate the general principles of protein folding at an atomic level, we have studied a mutant of the FF domain, L24A. Jemth et al.[24] have measured a Φ-value greater than unity (1.3) for the L24A mutation, consistent with nonnative interactions in the folding transition state. This group has also reported a folding rate of 300 s^{-1} from stopped-flow fluorescence measurements, which is a factor of 7 smaller than the rate measured for the WT FF domain (2200 s^{-1}).[24] Here we report the structure of the invisible I state of the L24A FF domain based on backbone chemical shifts and $^{1}H^{N}-^{15}N$ RDC values measured via CPMG relaxation dispersion experiments. Notably nonnative contacts beyond those observed in the WT FF domain structure are present that trap the intermediate, further slowing the folding reaction. Taken together, the structures of the folding intermediates of WT and L24A FF domains provide a basis for understanding the general principles of protein folding, at a level of resolution that heretofore was not possible.

RESULTS

Measurement of NMR Relaxation Dispersion Data. A series of NMR experiments have been performed to measure relaxation dispersion profiles of backbone ^{15}N,[17,25] $^{1}H^{N}$,[26] $^{13}C^{\alpha}$,[16] $^{1}H^{\alpha}$,[27] and $^{13}C^{O}$[16,28] nuclei of L24A FF domain (Figure S1, Supporting Information), very similar to those described in our previous study of the WT FF domain folding intermediate.[21] The data for each type of nucleus were collected from protein samples with different isotope labeling schemes that are optimized for each experiment. All data sets were well fit with a model of two-site chemical exchange (reduced $\chi^2 < \sim 1$, see Supporting Information), as for the WT folding reaction,[21,22] rather than a more complex three-state model such as U ↔ I ↔ N that is sometimes necessary.[15,22] That a two-site model can be well fit to the data is consistent with the fact that the rate of interconversion between I and U is fast on the NMR chemical shift time scale (microsecond exchange[23]). Further, hydrogen−deuterium exchange experiments recorded on both WT[22] and L24A FF domains (Figure S2, Supporting Information) establish that the population of the U state, p_U, is less than 20% that of I, so that the exchange monitored by relaxation dispersion is well described by the reaction $I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N$ (see Discussion). Absolute values of chemical shift differences between states, $|\Delta\varpi_{IN}|$, were extracted from fits of the data, along with the population of the intermediate state, p_I, and the exchange rate constant, $k_{ex} = k_{NI} + k_{IN}$. Although signs of $\Delta\varpi_{IN}$ values are not available from relaxation dispersion experiments per se, they can be obtained by a comparison of peak positions in heteronuclear single/multiple quantum coherence (HSQC/HMQC) data sets recorded at multiple static magnetic fields[29] or by measurement of off-resonance, rotating-frame transverse relaxation rates.[30] Once the signs of $\Delta\varpi_{IN}$ were available, chemical shift values of backbone nuclei in the intermediate, $\varpi_I = \varpi_N + \Delta\varpi_{IN}$, were obtained and used in structure calculations (see below). Values of (p_I, k_{ex}) for L24A were compared with those extracted from the corresponding fits of dispersion profiles of the WT FF domain, providing insight into how the L24A mutation affects folding kinetics and thermodynamics.

Additional NMR relaxation dispersion experiments that are ^{15}N spin-state-selective[17] were performed on an L24A sample weakly aligned in a poly(ethylene glycol) (PEG, $C_{12}E_5$)/hexanol mixture.[31] Along with the parameters described above, these measurements allow the extraction of differences in $^{1}H^{N}-^{15}N$ RDC values between intermediate and native states, $|\Delta D_{IN}|$, from which $D_I = D_N + \Delta D_{IN}$ is obtained.[17] The relative orientations of amide bond vectors that are calculated from D_I values complement the nearly complete set of backbone chemical shifts that have been obtained. Together ϖ_I and D_I values are used in subsequent structure calculations, following a procedure that has been described previously.[21]

Effect of the L24A Mutation on Folding Kinetics and Thermodynamics. In order to evaluate how the L24A mutation affects FF domain folding kinetics and thermodynamics, we have performed ^{15}N relaxation dispersion measurements on U-^{15}N,^{2}H-labeled WT and L24A FF domain samples at 25 °C, under identical experimental conditions. Measured p_I values of 1.54% ± 0.01% and 5.96% ± 0.04% and exchange rate constants $k_{ex} = 1397 \pm 13\ s^{-1}$ and $321 \pm 2\ s^{-1}$ are obtained for FF WT and L24A domains, respectively, corresponding to a change in relative free energy of the N and I states of $\Delta\Delta G_{IN} = \Delta G_{IN}(L24A) - \Delta G_{IN}(WT) = -0.83$ kcal/mol ($\Delta G_{AB} = G_A - G_B$). Correspondingly, the change in ΔG between N and the transition state (†) for the N−I interconversion is $\Delta\Delta G_{\dagger N} = 0.07$ kcal/mol. From the reported value of $\Delta\Delta G_{UN} = -1.0$ kcal/mol,[24] it follows that the L24A mutation results in only a very moderate change in stability of the folding intermediate, $\Delta\Delta G_{UI} = \Delta\Delta G_{UN} - \Delta\Delta G_{IN} = -0.17$ kcal/mol. The small value of $\Delta\Delta G_{\dagger N}$ suggests

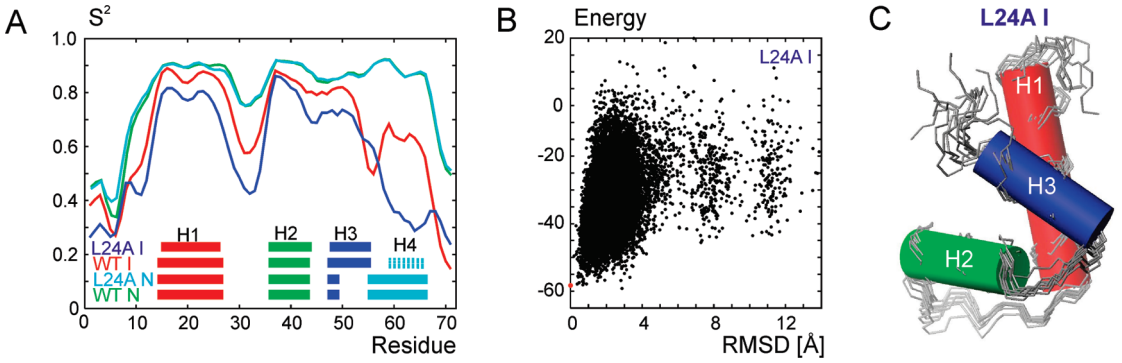


Figure 1. Secondary structure and dynamics of the L24A FF domain folding intermediate. (A) Comparison of secondary structure and backbone dynamics of the folding intermediates (I) and native states (N) of WT and L24A FF domains. Order parameters squared for the backbone amide groups, S^2, predicted by the RCI approach[33] are plotted vs residue numbers for the I states of WT (red) and L24A (blue) FF domains, with the corresponding plots for the N states shown in green and cyan. Secondary structure elements predicted by TALOS+[34] are indicated on the bottom of the plot (helix boundaries are 14−27, 36−43, 48−49, and 54−67 for the N state of WT and L24A; 14−27, 36−43, 48−55, and 59−66 for the WT I state; and 15−26, 36−44, and 49−55 for the L24A I state). (B) Energy of CS-Rosetta-generated models for the I state of L24A FF domain (residues 11−59), including Rosetta, chemical shift, and RDC terms vs rmsd to a representative low-energy structure (red dot). (C) Ensemble of nine representative structures of the I state of the L24A FF domain, calculated by the CS-Rosetta protocol described in the text and previously[21] (PDB code 219v). To outline the overall topology of the intermediate, helices 1−3 are depicted by rods.

that the relative stabilities of N and † change in concert so that k_{NI} is little affected by mutation, as observed ($k_{NI} = 21.5 \pm 0.2\ s^{-1}$ and $19.1 \pm 0.1\ s^{-1}$ for WT and L24A FF domains). By contrast, a much larger change in the barrier for the I to N folding step is observed, $\Delta\Delta G_{+I} = 0.90$ kcal/mol, that significantly decreases k_{IN} for the L24A FF domain ($1375 \pm 13\ s^{-1}$ versus $302 \pm 2\ s^{-1}$). Below we provide an explanation for the decreased folding rate based on an analysis of the structures of the WT folding intermediate determined previously[21] and the I state of the L24A mutant FF domain determined here. Finally, it is worth emphasizing that the effect of the L24A mutation on the measured I to N folding rate, k_{IN}, obtained from relaxation dispersion experiments, is different than for the U to N folding rate, as measured by stopped-flow fluorescence.[24] Notably, k_{IN} and k_{UN} decrease by 4.6- and 7.3-fold, respectively, relative to rates measured for the WT domain. This is not surprising since the L24A mutation would not be expected to change the stabilities of states I and U to the same extent, and as is shown below, mutation introduces significant conformational changes to the intermediate state. It should also be noted that the NMR experiments were recorded on highly deuterated samples, while the stopped-flow measurements were performed on fully protonated samples in solutions with a minimum of 1 M urea to destabilize the I state.[24] The differences in experimental conditions associated with both types of experiments would make it hard to compare even identical kinetic and thermodynamic parameters; as indicated above, the values reported here are based on measurements with perdeuterated WT and L24A FF domain samples, obtained under identical conditions via relaxation dispersion measurements so that valid comparisons can be made.

Secondary Structure and Dynamics of the L24A FF Folding Intermediate. Initially the L24A I state was characterized on the basis of the extracted backbone chemical shifts that provide valuable information about secondary structure[32] and conformational dynamics.[33] ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ chemical shifts of the intermediate state (92% of the possible chemical shifts were obtained) were used as input into the TALOS+ program,[34] generating a map of the I state secondary structure. Figure 1A shows a comparison of secondary structures of the native and intermediate states of the WT[21] and L24A FF domains predicted on the basis of the backbone chemical shifts. As expected, the L24A mutation does not affect the secondary structure of the native state; for both WT and L24A, four helices are predicted with identical boundaries (Figure 1A). The boundaries of the helices are consistent with those observed in the experimentally derived NMR structure of the WT protein.[35] The intermediate states of both WT and L24A FF domains include helices H1−H3, with the first two having very similar boundaries as the native state while the third helix, H3, is decisively nonnative. Helix H3 of the intermediate spans residues 48−55 (WT) and 49−55 (L24A) that form the 3_{10}-helix H3, the H3−H4 loop, and the beginning of α-helix H4 of the native protein. The C-terminal 15 residues of the intermediate state of L24A are disordered, so that H4 is essentially absent. In contrast, an "abridged" version of H4 is formed in I of the WT protein, spanning residues 59−66, although this helix is the least well formed of the four.[21]

N−H bond vector dynamics of the folding intermediate were predicted by the RCI approach,[33] in which the square of amide order parameters, S^2, was calculated on the basis of backbone chemical shifts. The predicted order parameters suggest that both of the folding intermediates of WT and L24A FF domains consist of a well-structured core of approximately 45 residues flanked by flexible N- and C-termini (Figure 1A). In both cases the core includes helices H1−H3 with RCI S^2 values ranging between 0.7 and 0.9, although the loop between helices H1 and H2 is flexible. The intermediate state of the L24A mutant is, however, more dynamic than that of the WT protein. In particular, S^2 values of about 0.6 were predicted for the C-terminal helix H4 of the WT folding intermediate, while for the L24A mutant the order parameters drop abruptly after the end of the nonnative helix H3.

More Extensive Nonnative Interactions in L24A Intermediate Than in Wild Type. Structure calculations of the folding intermediate of the L24A FF domain were carried out by the protocol described previously for the WT FF domain[21] in which backbone chemical shifts (ϖ_I) and $^{1}H^{N}-^{15}N$ RDCs (D_I) obtained from $\Delta\varpi_{IN}$ and ΔD_{IN} values measured in relaxation dispersion experiments were used as input restraints for the

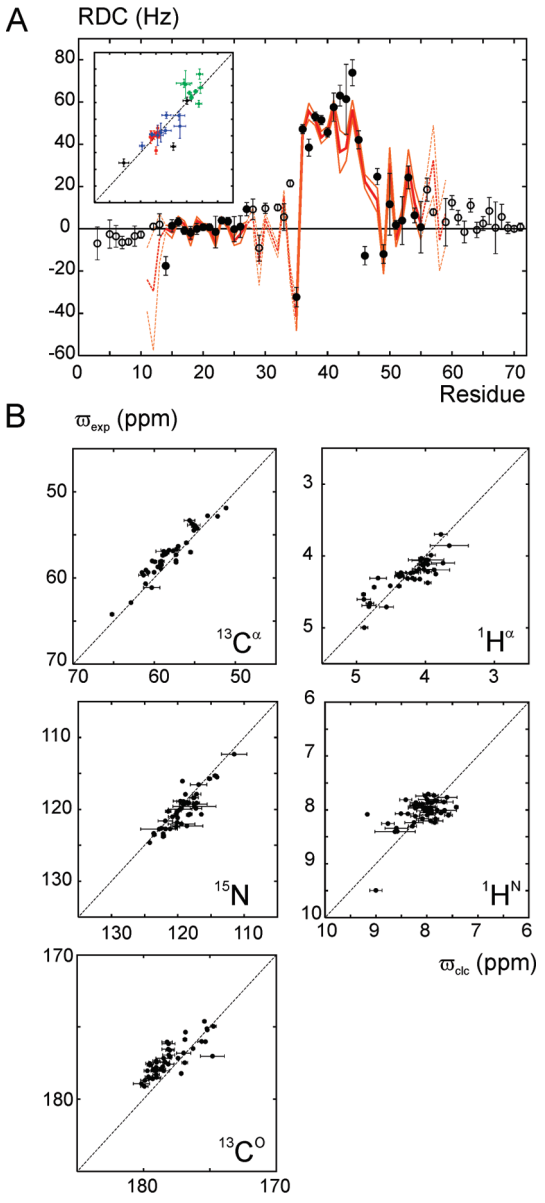


Figure 2. Correspondence between experimental and back-calculated RDCs and chemical shifts for the ensemble of representative structures of the I state of the L24A FF domain. (A) $^{1}H^{N}-^{15}N$ RDC values averaged over the nine lowest energy structures in Figure 1C (thick red line) and the standard deviation in predicted RDC values (thin orange lines). RDC data for residues with RCI S^2 > 0.6 used in scoring of assembled models are shown by solid circles/lines. (Inset) Experimental RDC values for the residues in helices H1 (red), H2 (green), and H3 (blue) and in other regions (black) plotted vs back-calculated values averaged over the ensemble. (B) Correlation plots of experimental ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ chemical shifts of the I state of the L24A FF domain, ϖ_{exp}, vs chemical shifts predicted by the SPARTA program,[36] ϖ_{clc}, averaged over the ensemble of the nine best models.

CS-Rosetta program.[18] As described previously, we have used a modified scoring function during the final structure selection stage that includes residuals between predicted and measured RDC values in addition to the Rosetta energy that has been rescored to include experimental chemical shifts.[21] RDC values are used only for final scoring and only if they derive from "rigid" residues with RCI S^2 values > 0.6. Calculations were performed for a core region comprising residues 11−59 with the flexible N- and C- terminal regions of the protein excluded. A total of 2×10^4 models were generated and scored,[21] with the 10 lowest energy structures selected for subsequent analysis. The lowest energy structure packs the N-terminal residues in such a way that if the initial unstructured 10 residues (residues 1−10 of the full-length sequence) were present, they would most certainly interfere with the interface between helices H1 and H3 and clash with helix H2 (Figure S3, Supporting Information). Hence this model was not used in further analyses. The remaining nine structures pack the N- and C-termini in a manner that ensures compatibility with the flexible extensions that are part of the "full-length" structure.

A plot of Rosetta energy supplemented with measures of the agreement with input chemical shifts and RDC values versus rmsd to a low-energy reference structure shows the characteristic funnel shape that is indicative of a converged structure calculation[18] (Figure 1B). Here the reference is chosen to be the second lowest energy structure. The nine structures chosen are well converged (Figure 1C) with a mean pairwise backbone rmsd of 1.0 ± 0.4 Å calculated over the regions with RCI S^2 > 0.6. (By contrast, the corresponding rmsd for the discarded structure to the set of nine models is 3.0 ± 0.2 Å.) Similarly, if the 50 lowest energy structures are chosen for analysis, approximately 10% of the structures have an N-terminal orientation that leads to clashes; removing these structures gives a set of 44 models for which a backbone pairwise rmsd of 1.55 Å is obtained. An ensemble of the lowest energy cluster of structures is shown in Figure 1C, with helices H1−H3 indicated.

Figure 2 presents correlation plots showing the level of agreement between either experimental $^{1}H-^{15}N$ RDCs (panel A) or chemical shifts (panel B) and the corresponding values calculated as averages over the set of nine converged, low-energy structures. As expected, the best models are in agreement with input experimental data, providing a good indication that local structure and overall topology of the intermediate are reproduced correctly. It is noteworthy that, as for the WT I state,[21] the L24A folding intermediate aligns more strongly in the PEG-($C_{12}E_5$)/hexanol medium[31] than the native structure due to an interaction of the intermediate conformer with the alignment medium, with RDC values ranging between −32:74 and −9:11 for the I and N states, respectively (see Supporting Information). We see no evidence for a similar interaction involving the native conformation, since high-quality spectra are obtained and RDC values are in keeping with expectations based on the concentration of alignment media used. We currently do not understand why the I state interacts with alignment media while the N state does not, but the difference may reflect the fact that the intermediate structure is significantly more plastic, potentially exposing side chains that could facilitate contacts with the medium.

As a control, we have also performed standard CS-Rosetta calculations of the structure of the native state of the L24A FF domain using only backbone chemical shifts, ϖ_N. Such a calculation is an important test of the methodology because it provides a direct assessment of how well the chemical shift restraints by themselves can "fold" the FF domain. Clearly, significant deviations in this case between the CS-Rosetta-derived structure and the corresponding model obtained by either NMR- or X-ray-based methods would provide an indication that structural studies of the invisible intermediate state from similar input data are likely to be problematic. Figure 3 compare the structures of the N states of WT FF domain obtained from a "standard"

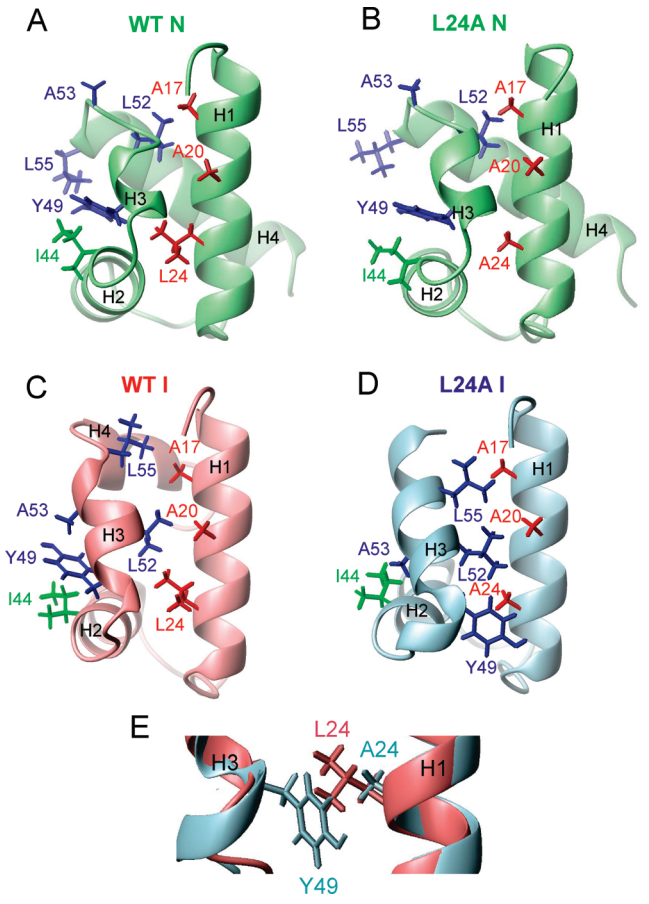


Figure 3. Nonnative contacts in the FF domain folding intermediate. Three-dimensional structures and side-chain packing of the native (A, B) and intermediate (C, D) states of WT (A, C) and L24A (B, D) FF domains are compared. Shown are the experimental NOE-based NMR structure of the native WT FF domain (PDB code 1UZC)[35] (A) and representative CS-Rosetta models of the FF WT I state[21] (C) and L24A N and I states (B, D). Outlined are side chains of residues from helices H1 (red), H2 (green), and H3 (blue) that form nonnative contacts in the intermediate state. (E) Interface of helices H1 and H3 in the I state of WT and L24A FF domains, showing that the conformation of Y49 observed in the L24A intermediate would not be possible in the WT I state due to steric clashes with the L24 side chain.

NOE-driven computation[35] (panel A) and L24A calculated by CS-Rosetta (panel B). The level of agreement is high, with a backbone rmsd of 1.0 Å over residues 11−70 (see also Figure S4, Supporting Information), providing confidence in using the Rosetta methodology for structural studies of the L24A folding intermediate.

Notably, the L24A mutation leads to a number of significant differences in the structure of the folding intermediate relative to that of the WT. This is illustrated in Figure 3 where low energy models of the folding intermediates of WT (panel C) and L24A (panel D) domains are compared. Overall, the structures share many common features. In both cases the spatial arrangement of helices H1 and H2 and the loop between them is similar and notably nativelike. For the I states of both WT and L24A, a nonnative α-helix H3 is formed in place of the 3_{10}-helix H3, the H3−H4 loop and the beginning of α-helix H4 in the native structure (compare Figure 3A,B with Figure 3C,D). In addition, in both intermediates helix H3 forms a nonnative interface with helix H1 involving residues A17, A20, L52, and L55. An important distinction, however, lies in the increased number of nonnative interactions in the L24A mutant due to the change in orientation of the aromatic side chain of Y49 at the start of α-helix H3. In the WT protein, Y49 stabilizes the H2−H3 interface by making contacts with I44 from H2 in both I and N states (Figure 3A,C), while in the structure of the intermediate Y49 forms nonnative interactions with A53 (Figure 3C). In contrast, in the L24A intermediate state the aromatic ring of Y49 flips toward helix H1 and forms a new nonnative interaction with A24. At the same time, removal of Y49 from the H2−H3 interface causes a change in tilt angle between helices H2 and H3, allowing them to come closer together to form a direct nonnative interaction between A53 and I44. It is important to emphasize that the orientation of Y49 found in the L24A I state is not possible in the WT protein because the side chains of Y49 and L24 would clash (Figure 3E). In addition to the differences described above, helix H3 contains an extra turn in the L24A I structure relative to the WT intermediate. This may be an artifact of the Rosetta force-fold that often extends elements of secondary structure in an effort to maximize hydrogen bonding.[18] In this context, it is noteworthy that the H3 boundaries as established by TALOS+[34] are R48−L55 and Y49−L55 for the WT and L24A I states, respectively. In summary, a comparison of the structures of WT and L24A intermediate states establishes that in addition to the nonnative interactions seen in the WT structure, additional nonnative contacts including A53−I44 and A24−Y49 are present in the mutant. As described below, these nonnative interactions are driven by the input experimental data (chemical shifts) and are not an artifact of the Rosetta force field used in the structure calculations.

Orientation of Y49 in the Intermediate Structure Is Driven by Input Chemical Shifts. A comparison of the three-dimensional structures of WT[35] and L24A FF domain folding intermediates establishes that the L24A mutation causes some rearrangements in side-chain packing due to a change in orientation of residue Y49, while the secondary structure and overall topology is maintained (Figure 3C,D). Since structure calculations of both intermediate states were based exclusively on backbone NMR chemical shifts and RDC data, additional analyses have been performed to ensure that it is these input experimental restraints that are responsible for the different Y49 conformations and not the force field of the Rosetta program.

In the first stage of structure calculation using the CS-Rosetta program,[18] 9-mer and 3-mer backbone fragments are selected by matching observed chemical shifts against those back-calculated from a database of known structures. Models are built by fragment assembly, and the range of angles present in the fragments represents a fundamental constraint on the topologies and angles sampled. Subsequently, side-chain packing occurs by a refinement process that identifies energetically favorable side-chain positions for a given input backbone structure. Note that the backbone is built without any detailed knowledge of side-chain packing.

As discussed in detail above, Y49 is a key player in generating the structural differences between WT and L24A I states; it is important, therefore, to ensure that experimental data are responsible for these distinct conformations. Figure 4A tabulates the measured chemical shifts for residues 48−50 in WT and L24A intermediate states and significant $^{1}H^{\alpha}$ and $^{13}C^{\alpha}$ chemical shift differences are observed for Y49, consistent with distinct backbone torsion angles for each conformer at this position.

A

	ϖ_I				
	^{15}N [ppm]	$^{1}H^{N}$ [ppm]	$^{13}C^{\alpha}$ [ppm]	$^{1}H^{\alpha}$ [ppm]	$^{13}C^{O}$ [ppm]
R48 L24A	118.38	8.34	57.23	4.06	177.49
R48 WT	117.98	9.00	57.68	4.20	177.83
Y49 L24A	117.88	7.71	58.17	4.60	176.52
Y49 WT	118.10	8.01	60.42	4.30	177.01
S50 L24A	115.73	7.87	60.66	4.28	175.37
S50 WT	114.87	7.95	61.30	4.31	175.96

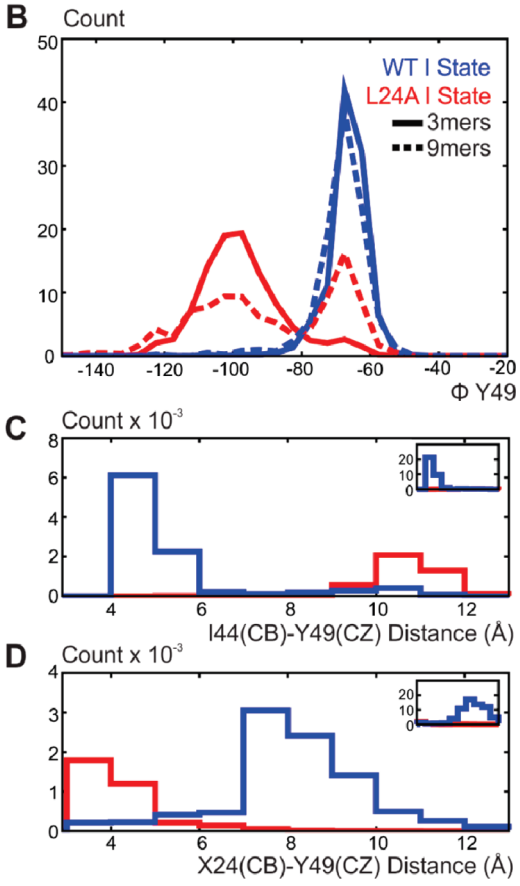


Figure 4. Experimental chemical shift data dictate the conformation of Y49. (A) Table of chemical shifts for residues 48−50 of L24A and WT intermediate states. (B) Histogram showing the distribution of φ angles for Y49 obtained from 9-mer (dashed line) and 3-mer (solid) fragments based on input chemical shifts of the WT (blue) and L24A (red) intermediate states. (C, D) Histograms of (C) I44(CB)−Y49(CZ) and (D) X24(CB)−Y49(CZ) distances obtained from structures of the intermediate states of WT (X = L) and L24A (X = A) FF domains, separated according to $\varphi_{Y49} < -80°$ (red, L24A I state conformation) or $\varphi_{Y49} > -80°$ (blue, WT I or native state conformation). (Insets) Corresponding distances in structures of the native state of L24A. Data are taken from the lowest-scoring 5% of 10^6 structural models generated for each state via the standard CS-Rosetta protocol.[36] Each of these structures is subsequently filtered against a set of 21 structures (the 10 lowest-energy conformers from WT and L24A I states and the NMR structure of WT N state[35]); so long as the structure in question is within 3 Å of at least one of the 21 structures, it is retained in the analysis. This process ensures that low-energy, misfolded structures are not included.

Notably, (φ, ψ) angles for this residue are equal to $(-55.4° \pm 5.5°, -41.6° \pm 7.8°)$ and $(-114° \pm 6.2°, 28.0° \pm 10.3°)$ for WT and L24A intermediates, respectively, based on analysis of calculated low-energy Rosetta models. In an effort to establish that such differences emerge from chemical shifts, we have analyzed the 9-mer and 3-mer fragments that are selected in the first stage of the computation. The distribution of Y49 φ angles so generated is illustrated in Figure 4B, with a very similar picture obtained when ψ angles are tabulated. Fragments for the WT I state contain φ values primarily in the nativelike α-helical range, −80° to −50°, while for the L24A intermediate state φ values largely extend into the nonnative range, −140° to −80°. It is clear that when WT I chemical shifts are used as input into CS-Rosetta calculations, there are virtually no fragments with Y49 φ angles that would be necessary for adoption of the L24A excited state structure ($\varphi \sim -115°$). It is also clear that the experimental chemical shift data are what "drives" the L24A intermediate structure into a distinct, nonnative conformation in the vicinity of Y49.

Of the nonnative interactions present in the L24A intermediate, the side-chain contact between residues 24 and 49 is critical. It is possible only because of the L24A mutation (Figure 3E) and it must be broken during the transition from state I to N. We have examined the effects of the Y49 φ torsion angle on the 24−49 interaction in both L24A and WT intermediates by measuring distances from Y49(CZ) to I44(CB) (a native contact) and from Y49(CZ) to residue 24(CB) (a nonnative contact). The Y49(CZ)−I44(CB) contact is 11.2 ± 0.6 and 4.7 ± 0.3 Å in L24A and WT I states, respectively, while the Y49(CZ)−residue 24(CB) distance is 4.4 ± 1.1 Å (L24A) and 9.9 ± 1.1 Å (WT). Contact distances have been compared across all I state CS-Rosetta models (for the lowest 5% in energy whose folds resemble those of either L24A or WT I states or the N state; see caption to Figure 4), by first separating structures based on nativelike Y49 φ angles, $-80° \le \varphi \le -40°$ (blue in Figure 4C,D) and nonnative angles, $-140° \le \varphi \le 80°$ (red). In almost all cases a nonnative φ angle between −140° and −80° (red curve) is incompatible with the native contact (Figure 4C) and provides a clear preference for the nonnative contact (Figure 4D). A very similar scenario occurs for the L24A native structure, as shown in the inset to Figures 4C,D. Here the 9-mer and 3-mer fragments for the native state localize to $-80° \le \varphi \le -40°$ for Y49 (data not shown)—that is, the native Y49−I44 interaction predominates—while the nonnative Y49−residue 24 contact is not observed. Again it is the input chemical shifts that determine whether native or nonnative interactions involving Y49 are obtained.

DISCUSSION

It is now clear that folding of even small, single-domain proteins can proceed through formation of intermediate states.[3−8,15] Characterization of these intermediate conformers provides insight into the nature of the folding process; however, in cases where the intermediates cannot be stabilized, their analysis is complicated by the fact that they are often only marginally populated and transiently formed. The FF domain from HYPA/FBP11 has been previously shown to fold via an on-pathway intermediate, $U \underset{k_{IU}}{\overset{k_{UI}}{\rightleftarrows}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftarrows}} N$, with fast formation of I from U (microsecond time scale), followed by a slower transition to the native state from the intermediate (millisecond time scale).[22,23] Under the experimental conditions used here with $p_U < p_I/5$ (see Figure S2, Supporting Information), the L24A FF domain folding pathway probed by CPMG NMR relaxation dispersion is well approximated by the I ↔ N interconversion, so it is expected that the chemical shifts and RDCs of the I state are little

contaminated by exchange with the unfolded ensemble. It is worth noting that, for fast exchange between U and I (microsecond time scale),[22,23] measured NMR parameters are population-weighted averages of values from each of the averaging states. Evidence that the U state contributes little to these parameters is provided by the fact that, as in the WT I state,[21] the positions and lengths of helices H1 and H2 are very similar to what is observed in the native structure, based on analysis of chemical shifts using the program TALOS+[34] (Figure 1A). In addition, RCI S^2 values, based on input chemical shifts, are again similar for both intermediate and native states in this region (Figure 1A). Clearly this would not be the case if there were significant contributions to the extracted I state chemical shifts from the U state.

In a recent study we have shown that the WT FF domain folding intermediate forms a stable, compact structure that contains both nativelike and nonnative interactions.[21] The structure so determined provides insight into why the final folding step to the native conformation is comparatively slow and rate-determining, since nonnative contacts in the intermediate state must be broken prior to formation of the native fold. Here we have extended our structural studies of FF domain folding by focusing on an L24A mutant that has an anomalously high Φ value, consistent with the presence of nonnative interactions in the folding transition-state ensemble.[24] In addition, a significantly reduced folding rate has been measured for L24A relative to the WT domain under conditions where the I state is destabilized,[24] and an approximate 5-fold decrease in k_{IN} is obtained (1397 s^{-1} for WT versus 321 s^{-1} for L24A) from analysis of CPMG relaxation data. If the rate-limiting step of FF domain folding involves breaking of nonnative interactions in the intermediate, then the slower k_{IN} value for L24A would be consistent with a larger number of such contacts relative to the WT protein.

In an effort to test this idea, chemical shifts and RDCs that report on bond vector orientations have been measured from relaxation dispersion data and used to determine an atomic resolution model for the invisible L24A I state based on an approach described previously.[21] As is the case for the WT intermediate, the structure of the L24A I conformer is stabilized by a combination of native and nonnative contacts. The nonnative interactions involving A17, A20, L52, and L55 reported previously in the context of the WT I state are also found for the L24A mutant. A notable difference, however, is that the network of nonnative interactions is more extensive for L24A than for WT. These additional interactions are formed because of a change in orientation of the side chain of Y49, as illustrated in Figure 3. In the I and N states of the WT domain, Y49 is localized to the outside surface of the protein, stabilizing the interface between helices H2 and H3. In contrast, in the L24A I state the aromatic group is flipped by 180° to extend into the interface between H1 and H3, forming nonnative contacts with A24. The nonnative A53−Y49 interaction in the WT intermediate is now replaced by the nonnative A53−I44 contact in the L24A I state. The comparison of WT and L24A intermediate structures provides compelling evidence that indeed "correcting" nonnative interactions is rate-limiting for FF domain folding.

Formation of non-native contacts in the I state of both WT and L24A FF domains slows folding. Similar kinetic traps have also been observed in studies of other all-helical proteins such as Im7,[4] Rd-apocyt-b_{562},[5] R16 and R17 domains of α-spectrin,[8] and EnHD.[7] In contrast, in other protein systems nonnative interactions can lead to enhanced folding rates. For example, G48 mutants of the SH3 domain from Fyn tyrosine kinase were found to fold with dramatically increased rates over WT, by as much as a factor of 10.[9] A strong correlation was found between β-sheet propensity and folding rates, suggesting that formation of stabilizing β-sheet structure in the transition state was responsible for the higher rate of folding. Interestingly, in this system, residues that are most stabilizing in the transition-state structure are most destabilizing in the native state and also cause the greatest reductions in in vitro functional activity. A similar situation was found for the Pin1 WW domain, where the exchange of a six-residue loop that is critical for function with a shorter sequence increased the folding rate by nearly an order of magnitude.[37] These examples make it clear that proteins have evolved for function and not folding speed.[3,38] In this regard, many of the residues in the FF domain that participate in nonnative secondary and tertiary interactions that stabilize the folding intermediate and decrease k_{IN} also play important structural roles in the native state that are critical to the stability of the protein and perhaps also to its function. For example, when residues A20 and L52 that make nonnative contacts are mutated to G and A, respectively, the native state is destabilized by close to 2.5 and 3 kcal/mol, respectively.[24]

The work presented here is important on a number of levels. First, it confirms the robustness of our approach for studying excited protein states, especially significant given that the field is still at an early stage of development. Second, with an additional atomic-resolution structure of an on-pathway FF domain folding intermediate that folds at a different rate than the WT domain, it is possible to begin to establish general principles of folding that cannot be done when only a single structure is available. Third, by comparing WT and L24A FF domain folding intermediates, the effects of single-site mutations on the structure of these invisible states can be studied, as we have shown here. While detailed structural studies of mutants of "native states" have appeared, there is little atomic-resolution structural information available on excited states that, in principle, could be much more affected by mutation. We have shown that for the L24A folding intermediate this is indeed the case, emphasizing the plasticity of states formed along the FF domain folding pathway.

In summary, structures of the invisible protein-folding intermediates of WT[21] and L24A FF domains provide strong evidence that, in this case, breakage of nonnative secondary and tertiary interactions is rate-limiting for folding. This study emphasizes the important role of nonnative interactions in controlling the rate of protein folding and in stabilizing folding intermediates, and it provides caution against the indiscriminate use of assumptions based on native-centric folding in the interpretation of experimental data. It also highlights the emerging role of CPMG relaxation dispersion NMR spectroscopy in structural studies of low-populated and transiently formed protein states at a level of detail that is not possible by use of other biophysical techniques.

MATERIALS AND METHODS

Relaxation dispersion data for ^{15}N,[17,25] $^{1}H^{N}$,[26] $^{13}C^{\alpha}$,[16] $^{1}H^{\alpha}$,[27] and $^{13}C^{O}$ [16,28] nuclei were recorded at magnetic field strengths of 11.7 and 18.8 T for samples of the L24A FF domain with specifically tailored isotope labeling schemes (U-^{15}N,^{13}C,^{2}H-labeling for ^{15}N, $^{1}H^{N}$, $^{13}C^{O}$ dispersion experiments, 20 °C; selective $^{13}C^{\alpha}$ labeling, full protonation for $^{13}C^{\alpha}$ experiments,[39] 30 °C; U-^{15}N,^{13}C,≈50%-^{2}H labeling for $^{1}H^{\alpha}$ experiments,[27] 30 °C. The buffers used were 50 mM sodium acetate, 100 mM NaCl,

pH = 5.7, 90% H_2O/10% ^2H_2O for ^{15}N, $^1H^N$, $^{13}C^O$ experiments or 100% ^2H_2O for $^{13}C^\alpha$, $^1H^\alpha$ experiments). Data were analyzed separately for each type of nucleus by use of a model of two-site conformational exchange, I ↔ N, as described elsewhere.[22] The chemical shifts of the folding intermediate, $\varpi_I = \varpi_N + \Delta\varpi_{IN}$, were calculated from $\Delta\varpi_{IN}$ values extracted from relaxation dispersion data, with signs determined as described previously[29,30,40] (see Supporting Information). In this way $\Delta\varpi_{IN}$ values have been obtained for 66 ^{15}N, 65 $^1H^N$, 60 $^{13}C^\alpha$, 58 $^1H^\alpha$, and 60 $^{13}C^O$ nuclei of the L24A FF domain. Additionally, ^{15}N relaxation dispersion data were measured for ^{15}N,^2H-labeled L24A and WT FF domain samples at 25 °C, allowing direct comparison of parameters for folding kinetics and thermodynamics ($k_{ex} = k_{NI} + k_{IN}$ and p_I).

$^1H^N-^{15}N$ RDC values of the folding intermediate were obtained from ^{15}N spin-state selective relaxation dispersion data[17] recorded at 11.7 and 18.8 T and measured on an ^{15}N,^2H-labeled L24A FF domain sample weakly aligned in a PEG($C_{12}E_5$)/hexanol liquid crystalline phase[31] (16 Hz residual ^2H_2O splitting, 20 °C). The data were analyzed to extract RDC differences between intermediate and native states, $|\Delta D_{IN}|$, from which RDCs in the intermediate state were calculated as $D_I = D_N + \Delta D_{IN}$. Relative signs of ΔD_{IN} and $\Delta\varpi_{IN}$ can be determined from fits of relaxation dispersion data as has been described previously,[17] so that signs of ΔD_{IN} can be determined once those for $\Delta\varpi_{IN}$ are known. Values of D_N were measured by the $^{15}N-^1H^N$ IPAP experiment.[41] Similar to the WT I state,[21] the intermediate state of the L24A FF domain interacts weakly with the alignment medium, resulting in elevated RDC values relative to those for the native conformation.

The secondary structure and dynamics of the folding intermediate were initially assessed by the TALOS+[34] and RCI[33] programs based on a nearly complete set of backbone chemical shifts obtained as described above. RCI-based S^2 values greater than 0.5 are obtained for all but a very few residues between W11 and K58. Therefore, this region, flanked by two residues on each side, was used in structure calculations of the intermediate state. The I state structure was generated via the standard CS-Rosetta protocol,[18] modified to include an empirical RDC energy term at the final stage of rescoring the calculated structural models.[21] In total, 2×10^4 models of the intermediate were generated, of which the 10 with lowest energies were analyzed, as described in some detail previously.[21]

■ ASSOCIATED CONTENT

Supporting Information. Complete ref 18; four figures showing experimental dispersion data, hydrogen exchange rates, and structures of the intermediate state of the L24A FF domain; and four tables of chemical shifts and residual dipolar couplings of the excited state. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
kay@pound.med.utoronto.ca

■ ACKNOWLEDGMENT

This work was supported through a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K. Dr. Julie Forman-Kay (Hospital for Sick Children, Toronto) is thanked for providing laboratory space and Dr. Ranjith Muhandiram for help with NMR experiments. T.L.R. acknowledges The European Molecular Biology Organization (ALTF 827-2006) and the CIHR for postdoctoral fellowships. L.E.K. is the recipient of a Canada Research Chair in Biochemistry. The coordinates of the L24A FF domain intermediate have been deposited in the PDB (219v).

■ REFERENCES

(1) Matouschek, A.; Kellis, J. T.; Serrano, L.; Fersht, A. R. *Nature* **1989**, *340*, 122.
(2) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937.
(3) Brockwell, D. J.; Radford, S. E. *Curr. Opin. Struct. Biol.* **2007**, *17*, 30.
(4) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 209.
(5) Feng, H. Q.; Zhou, Z.; Bai, Y. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5026.
(6) Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M. V.; Fersht, A. R. *Nature* **2005**, *437*, 1053.
(7) McCully, M. E.; Beck, D. A. C.; Fersht, A. R.; Daggett, V. *Biophys. J.* **2010**, *99*, 1628.
(8) Wensley, B. G.; Batey, S.; Bone, F. A. C.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J. *Nature* **2010**, *463*, 685.
(9) Di Nardo, A. A.; Korzhnev, D. M.; Stogios, P. J.; Zarrine-Afsar, A.; Kay, L. E.; Davidson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7954.
(10) Clementi, C.; Plotkin, S. S. *Protein Sci.* **2004**, *13*, 1750.
(11) Li, L.; Mirny, L. A.; Shakhnovich, E. I. *Nat. Struct. Biol.* **2000**, *7*, 336.
(12) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. *Nat. Struct. Biol.* **1994**, *1*, 149.
(13) Palmer, A. G.; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204.
(14) Korzhnev, D. M.; Kay, L. E. *Acc. Chem. Res.* **2008**, *41*, 442.
(15) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586.
(16) Hansen, D. F.; Vallurupalli, P.; Lundstrom, P.; Neudecker, P.; Kay, L. E. *J. Am. Chem. Soc.* **2008**, *130*, 2667.
(17) Vallurupalli, P.; Hansen, D. F.; Stollar, E.; Meirovitch, E.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18473.
(18) Shen, Y.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685.
(19) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615.
(20) Wishart, D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. *Nucleic Acids Res.* **2008**, *36*, W496.
(21) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312.
(22) Korzhnev, D. M.; Religa, T. L.; Lundström, P.; Fersht, A. R.; Kay, L. E. *J. Mol. Biol.* **2007**, *37*, 497.
(23) Jemth, P.; Gianni, S.; Day, R.; Li, B.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6450.
(24) Jemth, P.; Day, R.; Gianni, S.; Khan, F.; Allen, M.; Daggett, V.; Fersht, A. R. *J. Mol. Biol.* **2005**, *350*, 363.
(25) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Phys. Chem. B* **2008**, *112*, 5898.
(26) Ishima, R.; Torchia, D. A. *J. Biomol. NMR* **2003**, *25*, 243.
(27) Lundstrom, P.; Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Am. Chem. Soc.* **2009**, *131*, 1915.
(28) Lundstrom, P.; Hansen, D. F.; Kay, L. E. *J. Biomol. NMR* **2008**, *42*, 35.
(29) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 12352.

(30) Auer, R.; Hansen, D. F.; Neudecker, P.; Korzhnev, D. M.; Muhandiram, D. R.; Konrat, R.; Kay, L. E. *J. Biomol. NMR* **2010**, *46*, 205.
(31) Ruckert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793.
(32) Wishart, D. S.; Case, D. A. *Methods Enzymol.* **2001**, *338*, 3.
(33) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970.
(34) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213.
(35) Allen, M.; Friedler, A.; Schon, O.; Bycroft, M. *J. Mol. Biol.* **2002**, *323*, 411.
(36) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289.
(37) Jager, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10648.
(38) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
(39) Lundstrom, P.; Teilum, K.; Carstensen, T.; Bezsonova, I.; Wiesner, S.; Hansen, D. F.; Religa, T. L.; Akke, M.; Kay, L. E. *J. Biomol. NMR* **2007**, *38*, 199.
(40) Orekhov, V. Y.; Korzhnev, D. M.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126*, 1886.
(41) Yao, L. S.; Ying, J. F.; Bax, A. *J. Biomol. NMR* **2009**, *43*, 161.