THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

Cross-Validation of the Structure of a Transiently Formed and Low Populated FF Domain Folding Intermediate Determined by Relaxation Dispersion NMR and CS-Rosetta

Julia Barette,[†] Algirdas Velyvis,[†] Tomasz L. Religa,[†] Dmitry M. Korzhnev,[‡] and Lewis E. Kay*[,†,§]

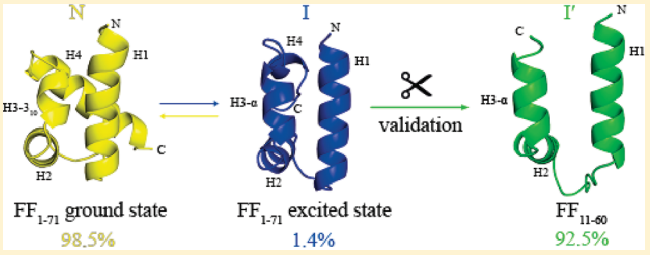
[†]Departments of Molecular Genetics, Biochemistry and Chemistry, The University of Toronto, Toronto, Ontario M5S1A8, Canada

[‡]Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, 263 Farmington Ave, Farmington, Connecticut 06030-3305, United States

[§]Hospital for Sick Children, Program in Molecular Structure and Function, 555 University Avenue, Toronto, Ontario M5G1X8, Canada

Supporting Information

ABSTRACT: We have recently reported the atomic resolution structure of a low populated and transiently formed on-pathway folding intermediate of the FF domain from human HYPA/FBP11 [Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L.E. *Science* **2011**, *329*, 1312−1316]. The structure was determined on the basis of backbone chemical shift and bond vector orientation restraints of the invisible intermediate state measured using relaxation dispersion nuclear magnetic resonance (NMR) spectroscopy that were subsequently input into the database structure determination program, CS-Rosetta. As a cross-validation of the structure so produced, we present here the solution structure of a mimic of the folding intermediate that is highly populated in solution, obtained from the wild-type domain by mutagenesis that destabilizes the native state. The relaxation dispersion/CS-Rosetta structures of the intermediate are within 2 Å of those of the mimic, with the nonnative interactions in the intermediate also observed in the mimic. This strongly confirms the structure of the FF domain folding intermediate, in particular, and validates the use of relaxation dispersion derived restraints in structural studies of invisible excited states, in general.



INTRODUCTION

Although some small proteins are thought to fold by a 2-state mechanism in which the unfolded state transitions in a highly cooperative manner to the folded conformer,[1,2] it is becoming increasingly clear that many proteins fold via processes involving formation of one or more metastable intermediates.[3−9] In principle, structural studies of such intermediates would provide a wealth of information about the folding process[7,8] and potentially also shed light as to how proteins misfold.[10] In practice, atomic resolution structural information is difficult to obtain because intermediates are often only fractionally and transiently populated, rendering them invisible to the traditional tools of structural biology. Recent developments[11] of Carr Purcell Meiboom Gill (CPMG) relaxation dispersion NMR spectroscopy[12,13] have provided an avenue for the study of these so-called excited protein states in cases where they are populated to at least 0.5% of the dominant (ground) state and exchange with it on the millisecond time-scale. In this approach, a variable number of refocusing pulses is applied during a spin−echo pulse train, modulating the effective transverse relaxation rates of NMR active nuclei (^{1}H, ^{13}C, and ^{15}N) that are part of the interconverting molecules. Fits of such rates as a function of the frequency of application of refocusing pulses to a model of chemical exchange leads to the extraction of the kinetic and thermodynamic parameters of the exchange process and of the chemical shift differences between interconverting states,[9,11] $|\Delta\varpi|$. Values of $|\Delta\varpi|$ can, in turn, be recast into shifts of the invisible excited state once the sign of $\Delta\varpi$ is known from additional experiments.[14−16] Backbone ^{15}N,[17,18] ^{1}H^{N},[19] ^{13}C$^{\alpha}$,[20] ^{13}C^{O},[21,22] and ^{1}H$^{\alpha}$[23] chemical shifts of the excited state can now be measured using a variety of different CPMG relaxation dispersion experiments. In addition, similar experiments have been developed for quantifying the orientation of bond vectors in the excited state via measurement of residual dipolar couplings (RDCs) on samples that are fractionally aligned.[24]

Chemical shifts and RDC values provide a powerful set of restraints for studies of the conformations of both ground and transiently formed excited states of proteins, even in the absence of other types of structural data.[8,25−29] Typically, a protocol is used whereby small fragments are selected from a database of structures based on the agreement with experimental data, with

Special Issue: Harold A. Scheraga Festschrift

Received: October 17, 2011
Revised: December 5, 2011
Published: December 07, 2011

models derived subsequently from fragment assembly. The approach has been cross-validated for a large number of small to medium sized proteins (approximately 150 residues or less) for which high resolution structural information is available from either traditional NMR or X-ray studies, and in the majority of cases, the structures produced are within 2 Å of those generated using traditional methods.[25,26] We have used one such database program, CS-Rosetta,[25] to produce models of an on-pathway folding intermediate of the wild-type FF domain from human HYPA/FBP11 (71 residues), whose ground state structure is a 4-helix bundle.[8] In this system, the intermediate is populated to approximately 2% and exchanges with the native state at a rate of 2000 s^{-1} (25 °C), well within the window of CPMG relaxation dispersion methodology so that a large number of chemical shift and bond vector orientation restraints could be obtained to produce structural models.[8]

Not surprisingly, the intermediate shares many aspects of structure with the native state.[8] For example, the orientations and lengths of the first two helices, H1 and H2, and the intervening loop are similar in both states. However, there are distinct differences as well. In particular, α-helix H3 is significantly longer in the intermediate, consisting of residues that comprise helix H3, the H3−H4 loop, and the beginning of α-helix H4 in the native structure. The lengthening of helix H3 generates a significant number of nonnative interactions that must be broken prior to formation of the native structure in what is the rate limiting step for folding in this system. It is thought that other helical bundle proteins also undergo a rate limiting reorganization of structure prior to assuming the final folded conformation.[30−33]

A great deal of information about the general principles of protein folding is available from atomic resolution models of intermediates that are generated along the folding trajectory. However, care must be taken to cross-validate the models that are produced. This is especially the case for the FF domain folding intermediate considered here since it is one of the very first examples of where an invisible excited state structure is produced from a relaxation dispersion NMR/database computational protocol. We show that truncation of the first ten and last eleven residues from the wild-type FF domain generates a fragment (FF_{11-60}), which adopts a highly populated conformer that is a good mimic of the wild-type folding intermediate. The structure of FF_{11-60} is then obtained using traditional NOE-based NMR methods[34] as a means of cross-validation of the CS-Rosetta-based intermediate structure obtained exclusively from relaxation dispersion derived restraints.[8] Good agreement between the FF_{11-60} and intermediate state conformations is obtained, with the nonnative interactions reproduced in the FF_{11-60} model, providing strong validation of the previously reported folding intermediate structure.[8]

MATERIALS AND METHODS

Sample Production. FF_{11-60} was produced from a modified pRSET vector (Invitrogen) encoding FF_{1-71} by using QuikChange PCR to delete residues 1−10 and 61−71 and to mutate residue 11 from W to G. The resulting construct expresses an N-terminal His-tagged lipo domain separated from the FF domain by a thrombin cleavage site.[35] The A51L mutant was likewise produced by QuikChange PCR from FF_{11-60}. FF_{11-60} and A51L-FF_{11-60} were expressed (with $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively) and purified as described.[35] Stereospecific assignment of pro-R and pro-S leucine δ and valine γ methyl groups made use of the method of Neri et al., using a sample prepared from a mixture of 10% $^{13}C_6$-glucose and 90% $^{12}C_6$-glucose.[36] Samples were 1−2 mM in protein dissolved in 20 mM sodium acetate, 50 mM NaCl, pH = 4.8, buffer containing 0.05% sodium azide, and 95% H_2O/5%D_2O.

NMR Spectroscopy. Backbone and side chain resonance assignments were performed at 25 °C on a 500 MHz spectrometer using a standard set of gradient and sensitivity enhanced triple-resonance NMR experiments.[37,38] Complete backbone assignments were obtained for all residues except G11, N12, K28, and R29. Stereospecific assignments of leucine and valine methyl groups were performed as described.[36] Interproton distance constraints were obtained from a simultaneous ^{15}N- and ^{13}C-edited NOESY-HSQC data set[39] recorded with a mixing time of 150 ms (500 MHz) and from a constant-time methyl−methyl NOESY experiment[40] (mixing time =150 ms, 800 MHz). NOEs from Ala residues were detected with high resolution using the pulse scheme illustrated in the Supporting Information (Figure S1; mixing time of 250 ms, 800 MHz) whereby NOE correlations of the form (ω^1H_{ALA},$\omega^{13}C_{ALA}$,ω^1H) were detected, where ω^1H_{ALA} and $\omega^{13}C_{ALA}$ are Ala methyl chemical shifts, and ω^1H is the chemical shift of a proximal proton.

Structure Calculations. The secondary structure and backbone dynamics of I′ were initially assessed using the TALOS+[41] and RCI[42] programs based on ^{15}N, $^1H^N$, $^{13}C^\alpha$, $^{13}C^O$, $^1H^\alpha$, and $^{13}C^\beta$ chemical shifts. Structural restraints included 58 φ and ψ torsion angles for regions predicted to be helical by TALOS+ as well as 121 NOE distance restraints (37 sequential $|i-j|=1$, 57 medium range $1 < |i-j| < 5$, and 27 long-range (all other) restraints; these were used as input into XPLOR-NIH[43,44] using script xplor-nih-2.27/eginput/protG/anneal.inp, modified to include the rama potential[45] for the ordered portion of the protein (residues 15−54). The final force constants for distance, torsion angle, and Ramachandran database potential of mean force restraints were 30 kcal mol^{-1}, 200 kcal mol^{-1}, and 1.0 kcal mol^{-1}, respectively. One hundred trial structures were calculated, and the ensemble of the 10 best structures with the lowest NOE and dihedral angle violations were chosen as representative (no NOE violation greater than 0.5 Å and no dihedral angle violation greater than 5°). PROCHECK software[46] was used to analyze the Ramachandran plots. Structural statistics are provided in the Supporting Information.

RESULTS AND DISCUSSION

The goal of the present work is to cross-validate the structure of the wild-type FF domain folding intermediate that we had solved previously on the basis of CPMG relaxation dispersion derived constraints and a CS-Rosetta structure determination protocol.[8] A number of experiments performed at the time of this initial work provided strong evidence but did not prove that the proposed intermediate state (I) structure was correct. First, using only backbone chemical shifts of the wild-type native state (N) in concert with CS-Rosetta, we were able to reproduce both the N state FF domain fold and the positions of the side chains, providing confidence in the protocol used for the calculations of I. Second, the structure of I was tested by making truncation mutations that were predicted to have little effect on I, while drastically affecting the stability of the native conformer. One such truncation involves removal of the C-terminal residues, 61−71, that form a large part of the α-helix H4 since these

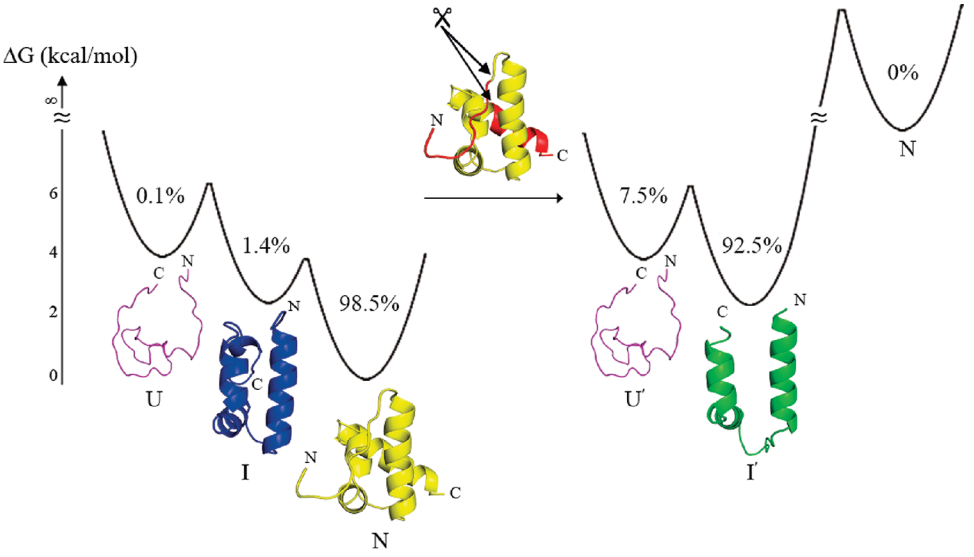


Figure 1. Populations and relative free energies of the unfolded (U), intermediate (I), and native states (N) of FF_{1-71} (left), as determined by stopped-flow fluorescence[6,35] and relaxation dispersion NMR studies, 25 °C.[8,47] The structure of the U state (purple) is a schematic based on a molecular dynamics simulation in the absence of restraints, while the structures of I (blue) and N (yellow) indicated by the ribbon diagrams, were determined by relaxation dispersion NMR/CS-Rosetta[8] (I; pdb accession code, 2KZG) and standard NMR[48] (N; pdb accession code, 1UZC) approaches, respectively. Truncation of N- (1−10) and C-terminal (61−71) residues (indicated in red on the structure above the arrow) creates FF_{11-60} (right) with a dominant population of I′ (whose structure determined in the present study is indicated in green). We have assumed that the relative energies between states I/U and I′/U′ are the same, and although U and U′ are assigned the same absolute free energies this, of course, need not be the case.

residues are critically important to the stability of the N state,[8,35] while they are likely of much less significance for I since they are partially disordered in this state. The resultant mutant, FF_{1-60}, folds with a single kinetic phase[8] and with a rate that is identical to that for the U to I transition in the wild-type protein,[6] suggesting that FF_{1-60} is a mimic of I. Moreover, while $^{1}H^{N}-^{15}N$ correlation spectra of FF_{1-60} established the presence of two separate conformers in slow exchange, the chemical shifts of one set were very similar to those of I, confirming that indeed helix 4 is not critical to the formation of the intermediate state, as expected on the basis of the I state structure. Unfortunately, the poor chemical shift resolution, in particular for the side chains, exacerbated by the extra set of correlations made further structural analysis on FF_{1-60} not possible.

Searching for an I State Mimic As a Quantitative Test of Structure. In an effort to produce an intermediate mimic with more favorable spectroscopic properties than FF_{1-60}, we compared backbone chemical shifts of both sets of correlations from the pair of slowly exchanging FF_{1-60} conformers. Although this analysis is still in a preliminary stage, we were able to ascertain that the major differences were in the N-terminus and that, further, the relaxation properties of one of the conformers were consistent with a dimeric structure. With this in mind, a second deletion was, therefore, introduced in which the first ten residues were removed and a W11G mutation added, to produce FF_{11-60}. As discussed in more detail below, $^{1}H^{N}-^{15}N$ HSQC spectra of FF_{11-60} were recorded showing a predominant single set of correlations, suggesting that this double truncated domain could form the basis for further structural studies.

The approach that we have taken to produce a potential I state mimic is schematized in Figure 1. The energy landscape of the wild-type FF domain is shown on the left-hand side with the relative stabilities of each of the states indicated (U, I, and N; 25 °C), established by stopped-flow fluorescence[35] and relaxation dispersion NMR.[8,47] Both the N- and C-termini have been cleaved to produce the landscape shown on the right, comprising an unfolded state, U′, and a predominant state, referred to as I′. In the diagram, we have assumed that the relative free energy difference between U′ and I′ is the same as that between U and I. Note that because I′ is highly populated, it can be studied using conventional NMR methods. We show below that I′ is a good structural mimic of the I state.

FF_{11-60} Is a Good Mimic of the FF Folding Intermediate. Prior to detailed studies of I′, it is first important to establish that this molecule is a good mimic of I. The most straightforward and rapid way of screening candidate molecules as good mimics is through a comparison of backbone chemical shifts with those of I. Figure 2a shows a $^{1}H^{N}-^{15}N$ HSQC correlation map of I′, with assignments of many of the cross-peaks indicated. These, along with assignments of side-chain ^{13}C and ^{1}H nuclei were obtained using a set of gradient and sensitivity enhanced triple resonance experiments,[37] supplemented with a simultaneous ^{15}N- and ^{13}C-edited NOESY-HSQC data set.[39] Correlations for residue N12 located at the N-terminus of the protein, as well as for K28 (C-terminus of helix H1) and R29 (N-terminus of the loop between H1 and H2), are not observed; however, all other backbone positions could be assigned. In addition to the intense correlations shown in the figure, there are much weaker peaks for which assignments are not available. Figure 2b compares ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ chemical shift differences between I′ and the native state of FF_{1-71} ($\Delta\varpi_{spec}$, Y axis) with chemical shift differences between the I and N states of FF_{1-71}, as obtained previously from analysis of relaxation dispersion data sets[8] ($\Delta\varpi_{disp}$, X axis). The line $\Delta\varpi_{spec} = \Delta\varpi_{disp}$ is included in the figure as a guide. It is clear that while there is a reasonably good correlation of chemical shifts, there is an offset between $\Delta\varpi_{spec}$ and $\Delta\varpi_{disp}$. Notably, $^{13}C^{\alpha}$ and $^{13}C^{O}$ resonances of I′ are more upfield than the corresponding

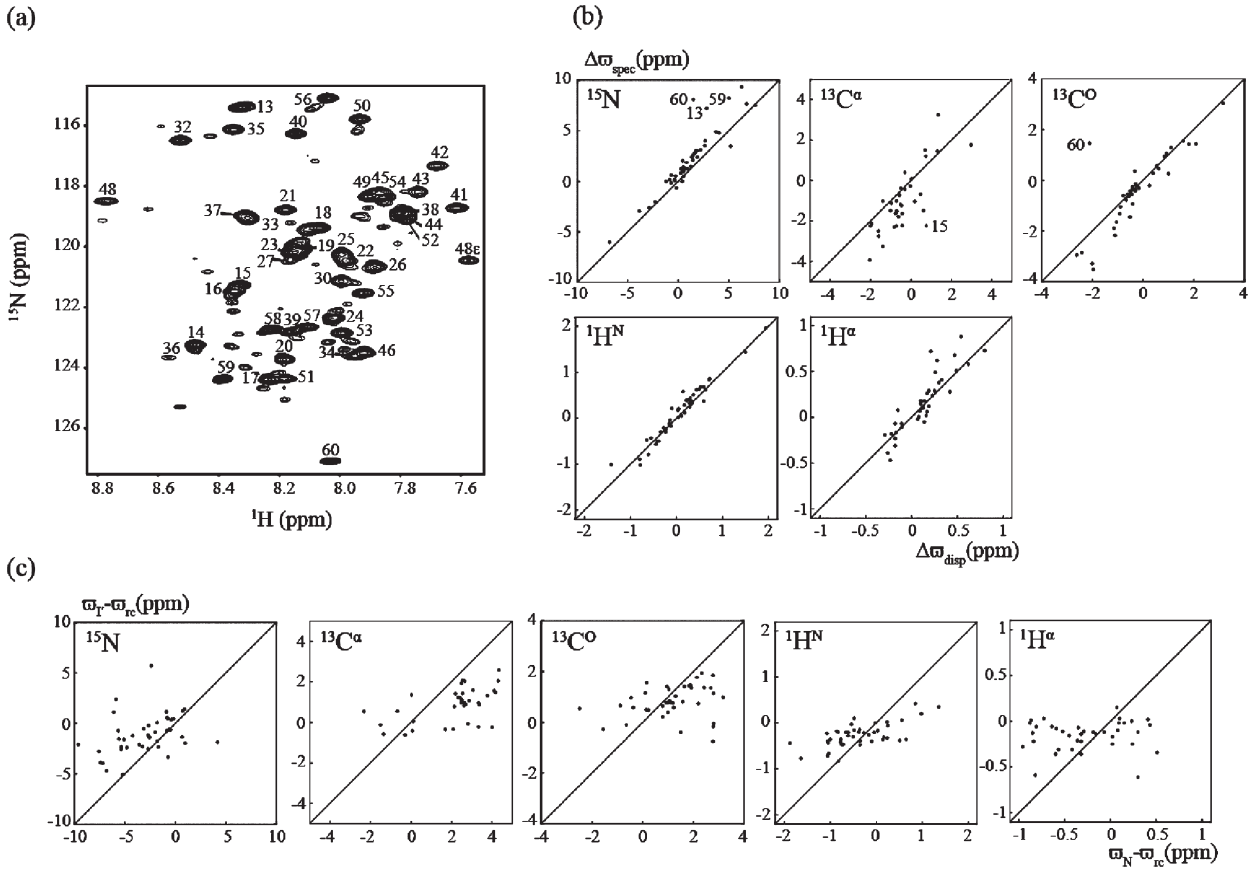


Figure 2. I′ state of FF_{11-60} is a good mimic of the FF_{1-71} folding intermediate. (a) $^{1}H^{N}-^{15}N$ HSQC correlation map of FF_{11-60}, with chemical shift assignments as indicated. A number of weak peaks are present that are unassigned. (b) Linear correlation plots of chemical shift differences between I′ and the FF_{1-71} native state ($\Delta\varpi_{spec}$, *Y* axis) vs the corresponding differences between FF_{1-71} I and N states ($\Delta\varpi_{disp}$, *X* axis) for backbone ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ spins. The FF_{1-71} I and N chemical shifts have been published previously,[8] while the I′ chemical shifts were obtained in the present study. The line $\Delta\varpi_{spec} = \Delta\varpi_{disp}$ is shown to guide the eye. Values of the rmsd between data sets have been calculated, after correcting for the offset (i.e., nonzero ($\langle\varpi_{spec} - \Delta\varpi_{disp}\rangle$ values), excluding outliers to the best fit line $y = x + b$; values of 0.69, 0.85, 0.44, 0.14, and 0.18 ppm are obtained for ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{O}$, $^{1}H^{N}$, and $^{1}H^{\alpha}$ spins. (c) Linear correlation plots of chemical shift differences between I′ and random coil (rc) shifts for the WT FF domain ($\varpi_{I'} - \varpi_{rc}$, *Y* axis) vs shift differences between FF_{1-71} native and random coil states ($\varpi_{N} - \varpi_{rc}$, *X* axis) for backbone ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ spins. FF_{1-71} random coil and native state chemical shifts were obtained from previously published values.[8]

peak positions in I ($\langle\Delta\varpi_{spec} - \Delta\varpi_{disp}\rangle_{C\alpha} = -0.59$ ppm, $\langle\Delta\varpi_{spec} - \Delta\varpi_{disp}\rangle_{CO} = -0.24$ ppm, where the angular brackets indicate averaging over all residues), while $^{1}H^{\alpha}$ resonances of I′ are more downfield ($\langle\Delta\varpi_{spec} - \Delta\varpi_{disp}\rangle_{H\alpha} = 0.05$ ppm) on average. These nuclei are sensitive reporters of secondary structure in proteins[49,50] (see below), and the correlation between $\Delta\varpi_{spec}$ and $\Delta\varpi_{disp}$ is, therefore, consistent with I′ and I possessing similar secondary structures. The signs of $\langle\Delta\varpi_{spec} - \Delta\varpi_{disp}\rangle_{X}$, $X \in \{^{13}C^{\alpha}, ^{13}C^{O}, ^{1}H^{\alpha}\}$ (and indeed also for ^{15}N and $^{1}H^{N}$, although these shifts are less diagnostic of secondary structure) suggest strongly that the extent of the α-helix secondary structure formation is less in the I′ state. A more detailed analysis, given below, establishes that the positions of the helices are very similar in I′ and I so that the doubly truncated variant is a good model of the wild-type FF domain folding intermediate that can be used to test the structure of I determined by CS-Rosetta. It is worth pointing out that, not surprisingly, the chemical shifts of I′ show no correlation with those of the N state of FF_{1-71}; Figure 2c.

I and I′ Have Similar Secondary Structures, but I′ Is More Dynamic. The backbone and $^{13}C^{\beta}$ chemical shifts of I′ have been input into the TALOS+ program[41] to predict residue-specific helical propensities; not surprisingly, I′ consists of 3 helices, H1−H3. The percent helicity vs residue profile for I′ is compared with that for I obtained in a similar manner; Figure 3a. Helices H1 and H3, that are adjacent to the truncation points, have lower helical content in I′, that most likely reflects partial destabilization of these elements caused by loss of adjacent residues. By contrast, the helical content of H2 is very similar in I′ and I. Notably, very similar boundaries are predicted for the two variants, with H1, H2, and H3 spanning residues 15−26, 36−44, and 48−54 in I′ compared to 14−27, 36−43, and 48−55 for I. The flexibility of the amides in I′ and I have been calculated from backbone chemical shifts using the RCI approach[42] and plotted as squared order parameters (S^2) in Figure 3b. Helices H1 and H3 are more dynamic in I′, with S^2 values between 0.70 and 0.80 (H1), 0.75−0.85 (H2), and 0.70−0.75 (N-terminus of H3) compared to 0.80−0.90 for the helices of I, with order parameters dropping off at the C-terminus of H3 for both variants. By means of comparison, we have also plotted RCI-based S^2 values for the N state of the wild-type FF domain.[8] The increased dynamics of I′ relative to N in part reflects rapid averaging between I′ and U (that is populated close to 10% in FF_{11-60}; Figure 1), that leads to

dx.doi.org/10.1021/jp209974f |*J. Phys. Chem. B* 2012, 116, 6637–6644

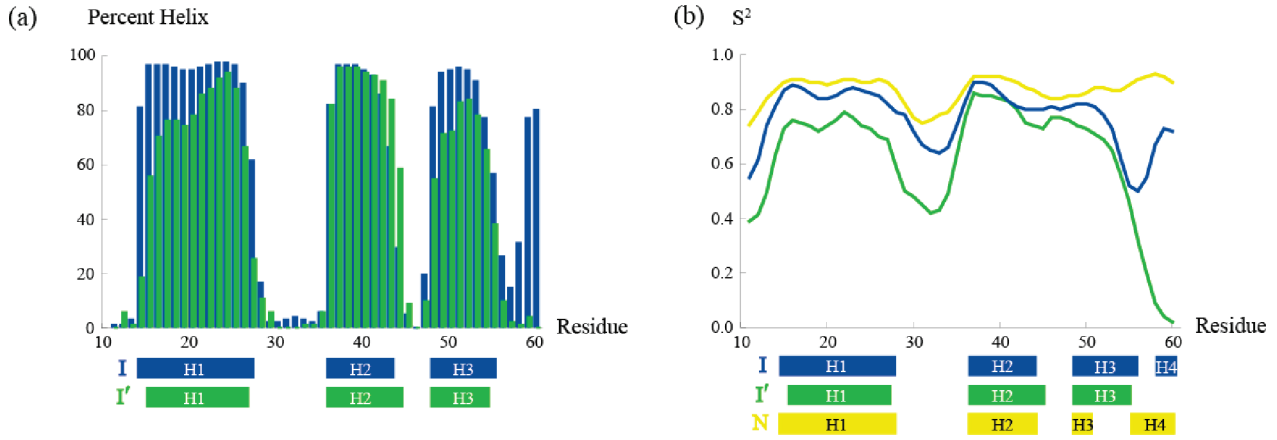
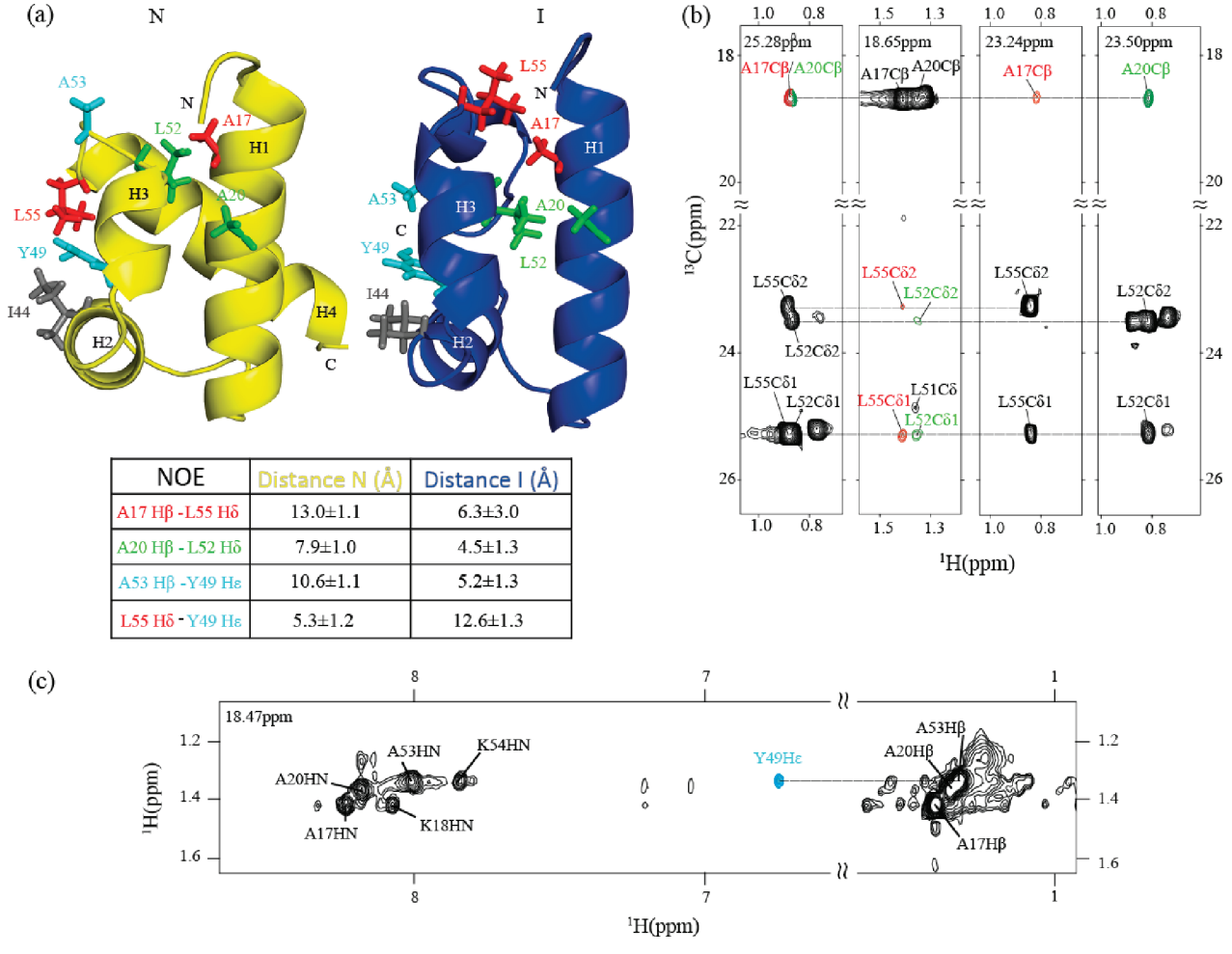


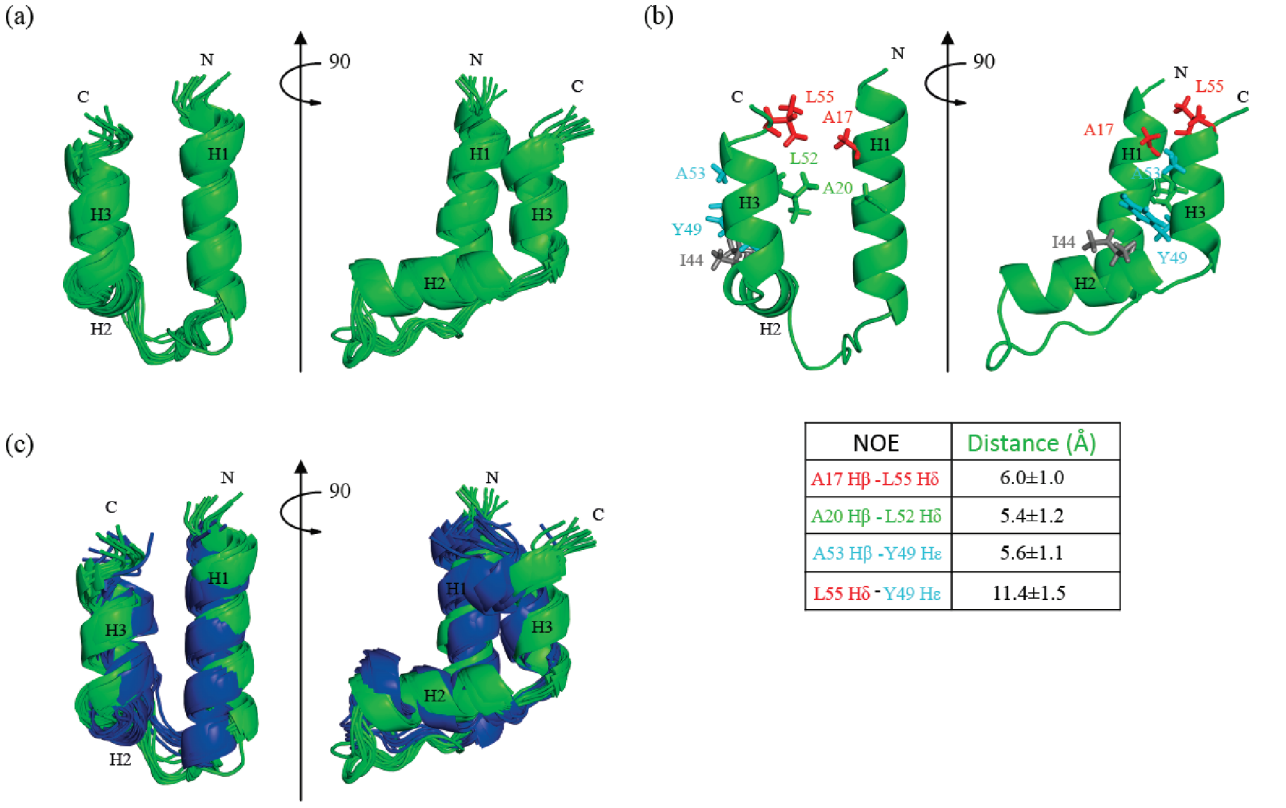
Figure 3. Comparison of per-residue helical propensities and backbone dynamics of I′ (green) with those of the FF_{1-71} folding intermediate, I (blue). (a) TALOS+-predicted helix propensities[41] based on backbone chemical shifts for I and backbone and $^{13}C^{\beta}$ shifts for I′. TALOS+ predicted helices are shown as rectangles and span residues 14−27, 36−43, and 48−55 for I and 15−26, 36−44, and 48−54 for I′. (b) RCI-predicted S^2 values for the backbone amide groups[42] of I′ (green), I (blue), and N (yellow).



NOE	Distance N (Å)	Distance I (Å)
A17 Hβ -L55 Hδ	13.0±1.1	6.3±3.0
A20 Hβ - L52 Hδ	7.9±1.0	4.5±1.3
A53 Hβ -Y49 Hε	10.6±1.1	5.2±1.3
L55 Hδ - Y49 Hε	5.3±1.2	12.6±1.3

Figure 4. Nonnative interactions predicted from the NMR relaxation dispersion/CS-Rosetta structure of I (see ref 8) are present in I′. (a) Structures of N (pdb accession code, 1UZC) and I (pdb accession code, 2KZG) states of FF_{1-71}. Colored in red, green, and cyan are three pairs of residues that form nonnative interactions in I, with the average and standard deviation of the distance between each residue pair in the native and intermediate states indicated in tabular form. Also indicated in the table is the Leu55H^{δ}−Tyr49H^{ε} distance in I and N. (b) F_1−F_3 strip-plots from a constant-time methyl−methyl NOESY spectrum[40] of FF_{11-60} (25 °C, 150 ms mixing time). NOE cross-peaks between L55 $H^{\delta1/2}$−A17 H^{β} are shown in red and between L52 $H^{\delta1/2}$−A20 H^{β} in green. (c) Strip-plot from a NOESY spectrum of FF_{11-60} correlating Ala $^{1}H^{\beta}/^{13}C^{\beta}$ chemical shifts with ^{1}H shifts of proximal protons (25 °C, 250 ms mixing time; see Supporting Information). An NOE cross-peak between A53 H^{β} and Y49 H^{ε} is shown in cyan.

dx.doi.org/10.1021/jp209974f |J. Phys. Chem. B 2012, 116, 6637–6644



NOE	Distance (Å)
A17 Hβ -L55 Hδ	6.0±1.0
A20 Hβ -L52 Hδ	5.4±1.2
A53 Hβ -Y49 Hε	5.6±1.1
L55 Hδ -Y49 Hε	11.4±1.5

Figure 5. Solution structure of the FF domain folding intermediate mimic, I′. (a) Superposition of the 10 lowest energy structures with no violations out of 100 calculated. The backbone heavy atom rmsd for the ensemble is 0.8 ± 0.2 Å. (b) Lowest energy structure of I′; highlighted in red, green, and cyan are three pairs of residues forming nonnative interactions in the I state, with the average and standard deviation of the distance between each pair in I′ indicated in the table. (c) Superposition of the ensemble of 10 representative lowest energy structures of I (blue; pdb code, 2KZG; rmsd of 1.1 ± 0.5 Å) with the 10 lowest energy structures of I′ (green). The backbone rmsd between pairs of structures from each of the ensembles is 2.1 ± 0.3 Å (all residues included). Only shown for I are those residues that are also present in I′.

a decrease in spectral resolution in relation to data sets recorded of the wild-type protein.

I′ Forms Nonnative Contacts Predicted by the CS-Rosetta Structure of I. As described above, a key difference in the structures of the N and I states of the wild-type FF domain is that a series of nonnative interactions are formed in I as a consequence of the elongated helix H3 in the intermediate;[8] Figure 4a. For example, the Y49-L55 contact in the N state is replaced by a Y49-A53 interaction in I (cyan), L55 in the I state is proximal to A17 (red), and residues L52-A20 (green) are closer in the intermediate then in the native structure. The distances between these amino acid pairs in both the N and I conformations are given in Figure 4a. The fact that I′ serves as a good structural mimic of the intermediate state provides an opportunity to rigorously test the CS-Rosetta structure of I by independently verifying the presence of these nonnative contacts. In principle, for a small domain of 50 residues, this should be easily accomplished by analysis of a ^{13}C-edited 3D NOESY data set. However, it is clear that I′ is much more dynamic than a typical native protein (see above), as might be expected for a domain that is only 1.5 kcal/mol more stable than its unfolded state (Figure 1). This in turn limits resolution in key regions of the ^{1}H−^{13}C correlation map. For example, A53/A20 are overlapped in ^{1}H−^{13}C HSQC spectra of I′, as are A17/A51, making it difficult to unambiguously assign a number of the potential nonnative NOE contacts. Since A51 (α-helix H3) is surface exposed, we have mutated this residue to Leu. The resulting $^{1}H^{N}-^{15}N$ fingerprint spectrum showed essentially no changes relative to the spectrum of A51-I′, with only minor perturbations in the region of the mutation. In addition to eliminating the A17/A51 ambiguity, the degeneracy of A53/A20 was lifted so that it was possible to resolve all of the Ala based NOEs. Figure 4b shows a number of strip-plots from a methyl−^{13}C-edited NOESY data set linking proximal methyl groups. Here, ^{13}C chemical shifts for both the origination and destination methyl groups were recorded in constant-time mode so that a high resolution map could be obtained.[40] A second data set was measured in which Ala ($^{1}H^{\beta}$,$^{13}C^{\beta}$) methyl chemical shifts were recorded in (t_1,t_2), with the chemical shifts of proximal protons measured in the direct dimension, Figure 4c (see Supporting Information, Figure S1). Notably, NOE correlations are observed between A17H$^{\beta}$/L55H$^{\delta}$, A20H$^{\beta}$/L52H$^{\delta}$ (Figure 4b), and A20H$^{\beta}$/Y49H$^{\varepsilon}$ (Figure 4c) and are all predicted from the structure of the wild-type intermediate, while none of these contacts were noted in studies of the native state FF domain,[48] as expected (BMRB and pdb accession codes, 5537 and 1UZC). An NOE connecting Y49/L55 was reported for the native structure that was not observed in data sets recorded of I′, again consistent with the expected distances in each state (Figure 4a). The observation of nonnative NOEs in spectra of I′, predicted from the structural model of I, provides very strong support of the CS-Rosetta-based structure of the invisible folding intermediate of the FF domain.

Solution Structure of I′ Validates the CS-Rosetta Structure of I. Having established that I and I′ have similar secondary structures

and that they both share the same nonnative contacts, we next solved the structure of I′ using the standard NOE-directed approach[34] (see Materials and Methods and Supporting Information, Table 1, for structural statistics). A superposition of the 10 lowest energy structures with no violations is shown in Figure 5a, with a single representative structure highlighting the nonnative interactions discussed above illustrated in Figure 5b (color coding is the same as in Figure 4). These nonnative distances measured in structures of I′ are tabulated in the figure. Notably, they are all within one standard deviation of the corresponding mean distances calculated for the CS-Rosetta structure of the folding intermediate (differences less than 1.5 Å), but they are at least 2 standard deviations away from the corresponding mean distances computed for the native state structure (differences ranging between 2.5−8.2 Å). Figure 5c shows the superposition of the ensemble comprising the 10 lowest energy I state structures in blue (pdb code, 2KZG; pair wise rmsd including all backbone heavy atoms = 1.1 ± 0.5 Å) with the corresponding ten best structures calculated for I′ in green (pair wise rmsd = 0.8 ± 0.2 Å). The backbone rmsd of the two structural ensembles excluding loops and unstructured termini (calculated over residues 14−28, 36−45, and 47−55) is 1.7 ± 0.3 Å (2.1 ± 0.3 Å when all residues are included). Typically, structures of proteins smaller than approximately 120 residues calculated using CS-Rosetta have a backbone rmsd of 1−2 Å relative to the experimentally determined X-ray or NMR structure.[25,26] The rmsd noted here, on the high end, likely reflects the fact that while I′ appears to be an excellent structural mimic of the wild-type folding intermediate, there are differences in the molecules that are compared (residues 11−60 vs 1−71) that could well lead to subtle changes in structure. For example, the absence of H4 in the deletion mutant, FF_{11-60}, may result in structural perturbations relative to the intermediate since in the structure of I there are contacts between H4 (that is only partially formed) and helices H1 and H2.[8] In turn, this could account for the slight differences in the orientation of helices between the two ensembles. Other differences may simply reflect (i) the quality of the I′ structure that is based on a reduced number of NOE constraints due to the dynamical nature of FF_{11-60}, which results in broadening of residues Leu24, Lys28, and Arg29 in particular and (ii) the relatively small number of restraints used to calculate I via the CS-Rosetta approach. Nevertheless, it is quite clear that the structures of I and I′ are very similar and that the intermediate conformation reported previously, which uses only CPMG-based restraints, is accurate.

CONCLUSIONS

The development of CPMG relaxation dispersion NMR spectroscopy, in concert with new and powerful database computational approaches for structure determination based on chemical shift restraints, opens up the possibility of obtaining atomic resolution structures of excited protein states[8,29] that play important roles in protein folding,[9] enzymology,[51−53] and molecular recognition.[54] As with any new methodology, the need for cross-validation is critical. Yet this is particularly challenging in studies of invisible states because the options for probing detailed structure in such systems are limited. One approach is to use the structure of the excited state as a starting point for the rational design of a stabilized form of this conformation through carefully crafted mutations that would in turn destabilize what was formerly the ground state. In this manner, the relative populations of the ground and excited states become inverted, the excited state becomes visible, and traditional structural biology methods can then be used for cross validation. This is a powerful approach because mutagenesis creates a stable version of the state of interest that is likely to be well behaved for biophysical characterization. An alternative approach is one where structures of the native and excited states are compared so as to come up with potential mutations that would have little effect on the structure of the higher energy conformer, while preventing the ground state from forming. In this case, the new ground state (that is, the former excited state) can again be studied by standard structural biology techniques, as we have demonstrated for I′ here. Such studies can be hampered, however, by the fact that the stability of this state remains essentially unchanged from that of the excited conformation, as illustrated in Figure 1.

Here, we have used this second method to generate a mimic of the wild-type FF domain folding intermediate, and we have solved its structure using standard multidimensional solution state NMR methods. Notably, the structures of I′ and I are similar, with the nonnative contacts predicted by the CS-Rosetta model (I) observed in the I′ conformation. This provides strong verification of our previously proposed I state structure. Further, the work demonstrates the utility in using a combined NMR relaxation dispersion/CS-Rosetta approach for studies of invisible excited protein states, providing structural data at a level of detail not possible using other biophysical techniques.

ASSOCIATED CONTENT

Supporting Information. Figure showing the pulse scheme used to record NOES between Ala $^1H^{\beta}$ and proximal proton spins; table showing structural statistics for the top 10% of structures calculated of I′. The structure of I′ has been submitted to the PDB (2LKS). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 416-978-0741. Fax: 416-978-6885. E-mail: kay@pound.med.utoronto.ca.

ACKNOWLEDGMENT

Dedicated to Professor Harold Scheraga on the occasion of his 90th birthday with great respect for his many seminal contributions to the field of protein chemistry. 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. J.B. and T.L.R. acknowledge funding support from the Natural Sciences and Engineering Research Council of Canada and the CIHR, respectively. L.E.K. is the recipient of a Canada Research Chair in Biochemistry.

REFERENCES

(1) Fersht, A. *Structure and Mechanism in Protein Science*; W.H. Freeman and Company: New York, 1999.
(2) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428–10435.
(3) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14121–14126.

(4) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M.; Alonso, D. O.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863–867.
(5) Capaldi, A. P.; Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. *Nat. Struct. Biol.* **2001**, *8*, 68–72.
(6) 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*, 6450–5.
(7) Religa, T. L.; Markson, J. S.; Mayor, U.; Freund, S. M.; Fersht, A. R. *Nature* **2005**, *437*, 1053–1056.
(8) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312–1316.
(9) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586–590.
(10) Chiti, F.; Dobson, C. M. *Nat. Chem. Biol.* **2009**, *5*, 15–22.
(11) Palmer, A. G.; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204–238.
(12) Carr, H. Y.; Purcell, E. M. *Phys. Rev.* **1954**, *54*, 630–638.
(13) Meiboom, S.; Gill, D. *Rev. Sci. Instrum.* **1958**, *29*, 688–691.
(14) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 12352–12360.
(15) Auer, R.; Neudecker, P.; Muhandiram, D. R.; Lundstrom, P.; Hansen, D. F.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **2009**, *131*, 10832–10833.
(16) Bouvignies, G.; Korzhnev, D. M.; Neudecker, P.; Hansen, D. F.; Cordes, M. H.; Kay, L. E. *J. Biomol. NMR* **2010**, *47*, 135–141.
(17) Loria, J. P.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1999**, *121*, 2331–2332.
(18) Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A. A.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 11341–11352.
(19) Ishima, R.; Torchia, D. *J. Biomol. NMR* **2003**, *25*, 243–248.
(20) Hansen, D. F.; Vallurupalli, P.; Lundstrom, P.; Neudecker, P.; Kay, L. E. *J. Am. Chem. Soc.* **2008**, *130*, 2667–2675.
(21) Ishima, R.; Baber, J.; Louis, J. M.; Torchia, D. A. *J. Biomol. NMR* **2004**, *29*, 187–98.
(22) Lundstrom, P.; Hansen, D. F.; Kay, L. E. *J. Biomol. NMR* **2008**, *42*, 35–47.
(23) Lundstrom, P.; Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Am. Chem. Soc.* **2009**, *131*, 1915–1926.
(24) Vallurupalli, P.; Hansen, D. F.; Stollar, E. J.; Meirovitch, E.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18473–18477.
(25) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685–4690.
(26) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.
(27) Wishart, D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. *Nucleic Acids Res.* **2008**, *36*, W496–502.
(28) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, *327*, 1014–1018.
(29) Vallurupalli, P.; Hansen, D. F.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 11766–71.
(30) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 209–216.
(31) Feng, H.; Zhou, Z.; Bai, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5026–5031.
(32) McCully, M. E.; Beck, D. A.; Fersht, A. R.; Daggett, V. *Biophys. J.* **2010**, *99*, 1628–1636.
(33) Wensley, B. G.; Batey, S.; Bone, F. A.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J. *Nature* **2010**, *463*, 685–688.
(34) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(35) Jemth, P.; Day, R.; Gianni, S.; Khan, F.; Allen, M.; Daggett, V.; Fersht, A. R. *J. Mol. Biol.* **2005**, *350*, 363–78.
(36) Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wüthrich, K. *Biochemistry* **1989**, *28*, 7510–7516.
(37) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
(38) Bax, A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 738–744.
(39) Pascal, S.; Muhandiram, D. R.; Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Magn. Reson. Series B* **1994**, *101*, 197–201.
(40) Zwahlen, C.; Gardner, K. H.; Sarma, S. P.; Horita, D. A.; Byrd, R. A.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 7617–7625.
(41) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213–223.
(42) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970–14971.
(43) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.
(44) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47–62.
(45) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1997**, *125*, 171–177.
(46) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
(47) Korzhnev, D. M.; Religa, T. L.; Lundstrom, P.; Fersht, A. R.; Kay, L. E. *J. Mol. Biol.* **2007**, *372*, 497–512.
(48) Allen, M.; Friedler, A.; Schon, O.; Bycroft, M. *J. Mol. Biol.* **2002**, *323*, 411–416.
(49) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647–1651.
(50) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492.
(51) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. *Science* **2006**, *313*, 1638–1642.
(52) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. *Nature* **2009**, *462*, 669–673.
(53) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913–916.
(54) Korzhnev, D. M.; Bezsonova, I.; Lee, S.; Chalikian, T. V.; Kay, L. E. *J. Mol. Biol.* **2009**, *386*, 391–405.