Transiently populated intermediate functions as a branching point of the FF domain folding pathway

Dmitry M. Korzhnev[a,1], Tomasz L. Religa[b,2], and Lewis E. Kay[b,c,1]

[a]Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center, Farmington, CT 06030; [b]Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8; and [c]Molecular Structure and Function, Hospital for Sick Children, Toronto, ON, Canada M5G 1X8

Edited by Peter G. Wolynes, Rice University, Houston, TX, and approved April 24, 2012 (received for review February 28, 2012)

Studies of protein folding and the intermediates that are formed along the folding pathway provide valuable insights into the process by which an unfolded ensemble forms a functional native conformation. However, because intermediates on folding pathways can serve as initiation points of aggregation (implicated in a number of diseases), their characterization assumes an even greater importance. Establishing the role of such intermediates in folding, misfolding, and aggregation remains a major challenge due to their often low populations and short lifetimes. We recently used NMR relaxation dispersion methods and computational techniques to determine an atomic resolution structure of the folding intermediate of a small protein module—the FF domain—with an equilibrium population of 2–3% and a millisecond lifetime, 25 °C. Based on this structure a variant FF domain has been designed in which the native state is selectively destabilized by removing the carboxyl-terminal helix in the native structure to produce a highly populated structural mimic of the intermediate state. Here, we show via solution NMR studies of the designed mimic that the mimic forms distinct conformers corresponding to monomeric and dimeric ($K_d = 0.2$ mM) forms of the protein. The conformers exchange on the seconds timescale with a monomer association rate of $1.1 \cdot 10^4$ M^{-1} s^{-1} and with a region responsible for dimerization localized to the amino-terminal residues of the FF domain. This study establishes the FF domain intermediate as a central player in both folding and misfolding pathways and illustrates how incomplete folding can lead to the formation of higher-order structures.

excited protein states | protein folding intermediate

It is increasingly clear that protein aggregation can be initiated from locally unfolded and/or misfolded conformations that become accessible due to thermal fluctuations from the native state (1–4). Such conformers, therefore, may play critical roles in the formation of molecular assemblies implicated in aggregation-related diseases (1–3). However, these species are often sparsely and transiently populated (4, 5), challenging their detailed characterization using conventional tools of structural biology. Their role in aggregate formation and in protein misfolding in general is, therefore, not well understood.

With the development of nuclear magnetic resonance (NMR) relaxation dispersion (RD) methods (5, 6) it has become possible to study transiently formed conformers populated at a level of 0.5% or higher (excited states), provided that they interconvert on the millisecond timescale with a populated conformation that can be observed in NMR spectra (ground state). The thermodynamics and kinetics of the exchange reaction can be accessed by RD methodology. In addition, backbone ^{1}H, ^{15}N, ^{13}C chemical shifts (5–7), residual dipolar couplings (RDCs) (8), and residual chemical shift anisotropies (RCSAs) (9) of the excited state can be measured, providing a rich source of structural information. By combining the RD measurements (5, 6) with computational approaches for protein structure determination from limited NMR data (10–12), it has become possible to generate atomic resolution models of excited protein states (13–15).

RD NMR studies of protein folding have been particularly fruitful (5). One example is the wild-type FF domain from human HYPA/FBP11, referred to as FF_{1-71} in what follows (13, 14, 16–18). This domain was shown by Fersht and coworkers to fold via an on-pathway intermediate that forms rapidly on the microsecond timescale and rearranges into the native state within a few milliseconds (16, 17). Because the folding intermediate has a fractional equilibrium population of ≈3% and a millisecond lifetime (25 °C), it cannot be directly observed in NMR spectra. Yet, its structure was recently elucidated using RD NMR (13, 14). The intermediate is remarkably well structured and can be thought of as representing an alternative, less favorable fold of the FF domain (13). Although native-like topology is preserved in this conformation, there are significant nonnative interactions that prevent formation of the C-terminal α-helix of the native four-helix bundle structure. These contacts must be broken prior to forming the native conformer, explaining why folding from the intermediate state is rate-limiting.

RD NMR-derived structures of folding intermediates open the way for designing rational strategies for their isolation (13). In this context, structural differences between intermediate and native states can be used to identify amino acids and/or regions of structure that contribute differently to the stabilities of these states. Through mutations and/or truncations the native state can be selectively destabilized so that the intermediate becomes the predominant conformation in solution. Such an approach led to the isolation of the FF domain folding intermediate (13). Here, truncation of 11 C-terminal residues that are partially disordered in the intermediate state, but form a stabilizing α-helix in the native protein, resulted in a stable variant, FF_{1-60}, with NMR spectra and folding kinetic profiles closely resembling those of the folding intermediate of FF_{1-71}. A similar approach has been employed in the isolation of folding intermediates of apo-cytochrome b_{562} (19), engrailed homeo domain (20), and the Im7 protein (21), albeit using different strategies for mutant design.

Once isolated, folding intermediates can be further investigated by conventional biophysical techniques. In one such application they can be used to cross-validate RD NMR-derived structural models of the intermediate state, a topic of particular interest at present because RD-derived structures are only beginning to emerge. Notably, high-resolution NMR studies of a truncated variant, FF_{11-60}, that mimics the folding intermediate of the full-length domain have established the accuracy of the RD-based model (22). However, beyond cross-validation of

Author contributions: D.M.K. and L.E.K. designed research; D.M.K. and T.L.R. performed research; D.M.K. and T.L.R. analyzed data; and D.M.K. and L.E.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: korzhniev@uchc.edu or kay@pound.med.utoronto.ca.

[2]Present address: Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH 44106.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201799109/-/DCSupplemental.

RD NMR-derived structures, "trapped" folding intermediates provide the means to explore transitions to other states on the folding energy landscape that are "hidden" in NMR studies of natively folded proteins by the presence of the ground state. We provide one such example using $FF_{1–60}$, showing that the FF domain folding intermediate interconverts on a seconds timescale between monomeric and dimeric forms. As such, the intermediate serves as a branch point in the folding pathway, with a conformation that facilitates both folding to the native state and formation of nonnative dimeric structures.

Results and Discussion

The FF Domain Folding Pathway. The N-terminal FF domain of human HYPA/FBP11 is a 71-residue, four-helix module (23) whose folding mechanism has been extensively studied by equilibrium and kinetic unfolding/refolding experiments, protein engineering methods, and NMR spectroscopy (13, 14, 16–18). In their original studies, Fersht and coworkers established that $FF_{1–71}$ folds to the native state, *N*, via an on-pathway intermediate, *I*, in two kinetic phases (16, 17), as shown in Fig. 1*A*. We have subsequently examined the folding behavior of the wild-type FF domain and a number of mutational variants by RD NMR spectroscopy (13, 14, 18), confirming the multistate nature of FF domain folding. A high-resolution structure of the *I* state (from restraints that include RD NMR-derived ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{13}C^{O}$ chemical shifts and $^{1}H^{N}$-^{15}N RDCs) was obtained using CS-ROSETTA (13) (Fig. 1*A*). It is noteworthy that all of the RD dispersion data, recorded in the temperature range 20–35 °C, could be well fit to a two-state model of exchange, $N \leftrightarrow I$, because the population of the unfolded state, *U*, under the conditions of our experiments is approximately an order of magnitude smaller than *I* and the $I \leftrightarrow U$ interconversion is on the microsecond timescale, which could not be detected in our RD experiments. Any additional conformational states, if present, were not observed in RD studies of the wild-type domain.

In an effort to probe the energy landscape still further we have designed a truncated variant of the wild-type protein, $FF_{1–60}$, that closely mimics the folding intermediate of the full-length domain (13). This variant lacks the C-terminal region containing native α-helix H4 that is critical for the structural integrity of the *N* state. Thus, $FF_{1–60}$ cannot adopt the native FF domain structure. Notably, $FF_{1–60}$ folds in a single kinetic phase with a rate of approximately 10^5/s, corresponding to that of the *U* to *I* transition in the full-length domain, $FF_{1–71}$ (16). Spectra of the truncated variant are of reasonable quality and have been assigned using standard triple-resonance NMR experiments (13). Of interest, two sets of signals were observed for many residues corresponding to a pair of conformations in solution (Fig. 1*B*). Backbone resonance assignments for one of the two states, "state *M*," are very similar to those of the folding intermediate of $FF_{1–71}$ (Fig. 1*C*), establishing that the two are structurally very similar. By contrast, the chemical shifts of the second form of $FF_{1–60}$, "state *D*," are distinct from those of the *I* and *N* states of $FF_{1–71}$ (Fig. 1*C*, *Inset*). Moreover, backbone resonance assignments for state *D* extend only to Met42, with resonances from the C-terminal portion of the domain, Ile43-Gln60, missing in NMR spectra, presumably due to microsecond–millisecond timescale exchange with additional conformational states. The chemical shifts for the flexible N-terminal residues extending up to Lys7 for both *M* and *D* states are indistinguishable. The presence of an alternative form of $FF_{1–60}$ in solution, distinct from the *I* state, suggests that the folding intermediate can access nonnative conformational states off the folding pathway of the FF domain (Fig. 1*A*).

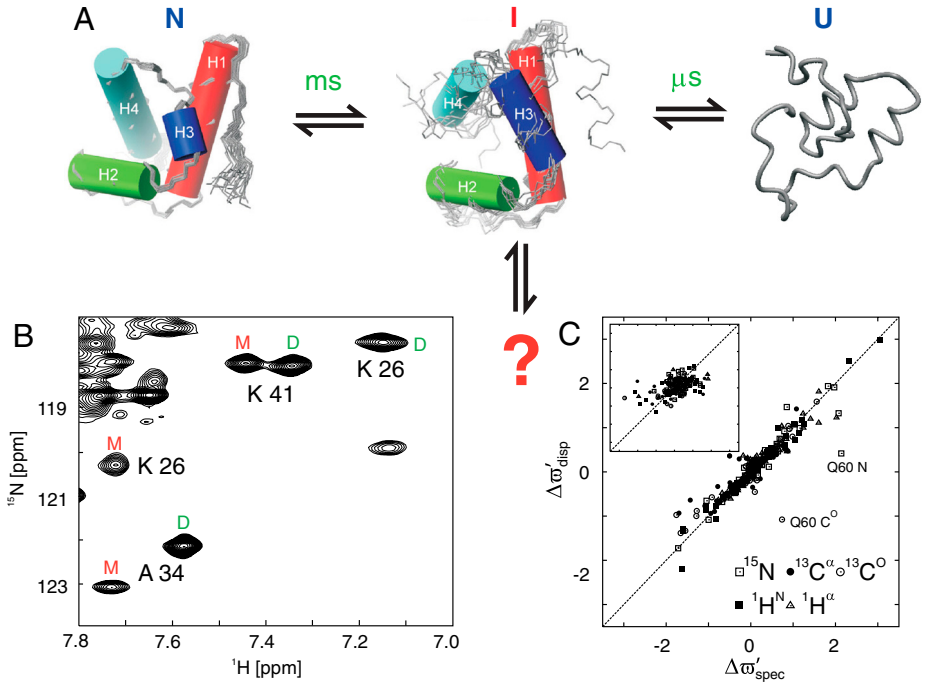

Fig. 1. (*A*) The folding pathway of the FF domain consists of a fast microsecond transition from *U* to *I* and a slower millisecond conversion from *I* to *N*. Structural ensembles of *I* [Protein Data Bank (PDB) ID code 2KZG] and *N* (PDB ID code 1UZC) were calculated as described elsewhere (13, 23). (*B*) Selected region of the $^{1}H^{N}$-^{15}N HSQC spectrum of the truncated variant, $FF_{1–60}$, a mimic of the folding intermediate of the full-length domain (11.7 *T*, 25 °C). Two sets of signals are observed, derived from two separate conformations of the protein, denoted as *M* and *D*. (*C*) Backbone $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{O}$ chemical shift differences between *I* and *N* states of the full-length FF domain ($FF_{1–71}$) obtained from RD NMR data (13) ($\Delta\varpi'_{disp} = \Delta\varpi_{disp}/\varpi_{std,i}$), plotted vs. chemical differences between $FF_{1–60}$ (*M* form) and the native state $FF_{1–71}$ ($\Delta\varpi'_{spec} = \Delta\varpi_{spec}/\varpi_{std,i}$). $\varpi_{std,i}$ is a nucleus- and residue-specific normalization value that corresponds to the range of shift values (1 SD) that are observed in a database of protein chemical shifts (www.bmrb.wisc.edu) for the nucleus/residue in question. A similar correlation for the *D* state of $FF_{1–60}$ is shown (*Inset*).

Mimic of the Folding Intermediate Undergoes Slow Dimerization. We have carried out further NMR studies of FF_{1-60} to understand the origin of the second set of correlations that have been observed in spectra in the hope that this would provide an avenue for further exploration of the FF domain energy landscape. The fact that separate sets of signals are observed for the *M* and *D* states of FF_{1-60} indicates that the interconversion of these forms is slow on the NMR chemical shift timescale. Further quantification can be obtained by ^{15}N longitudinal exchange measurements (24, 25), provided that the rate of interconversion is on the order of (or somewhat faster than) the spin-lattice relaxation rates of the backbone ^{15}N probes that are exploited in the experiment (i.e., 1–2 per s). By recording the ^{15}N chemical shift prior to a mixing period of duration *T*, the resulting correlation spectrum will contain four peaks per amide group: two autopeaks, *MM* and *DD*, at chemical shifts (ϖ_{NM}, ϖ_{HM}) and (ϖ_{ND}, ϖ_{HD}), originating from magnetization that is not transferred between states during *T*; and two exchange cross-peaks, *MD* and *DM*, at chemical shifts (ϖ_{NM}, ϖ_{HD}) and (ϖ_{ND}, ϖ_{HM}), derived from magnetization transferred between states during the mixing period (24, 25). Fig. 2*A* shows a region of the $^1H^N$-^{15}N exchange spectrum recorded on a 0.34 mM FF_{1-60} sample ($T = 0.34$ s, 25 °C) with exchange cross-peaks connecting autopeaks of amide groups in states *M* and *D*, indicating that the two forms of FF_{1-60} interconvert on the seconds timescale.

Fig. 2*B* shows the mixing time dependencies of auto- and cross-peak volumes for Ala34, with corresponding traces for Ser35 illustrated in Fig. S1. These were fit to a model of exchange,

$$f_M \underset{k_{DM}}{\overset{k_{MD}}{\rightleftarrows}} f_D,$$

where f_M and f_D are the fractional populations of states *M* and *D*, respectively, as described in detail by Tollinger et al. (25) and summarized in *SI Materials and Methods*. Although well-resolved autopeaks for over twenty residues of the *M* and *D* states are observed in $^1H^N$-^{15}N correlation spectra of FF_{1-60}, significantly fewer amide groups produce a set of four non-overlapped auto- and cross-peaks that are required to obtain robust estimates of k_{MD}, k_{DM}. These were obtained from fits of five complete sets of time profiles, with $k_{MD} = 3.05 \pm 0.95$ s^{-1} and $k_{DM} = 2.16 \pm 0.76$ s^{-1}. In addition to the rates, per residue ^{15}N R_1 values were also obtained (see below), with rates for the *D* state systematically smaller than for the *M* conformer that is a mimic of the folding intermediate.

Values of ^{15}N R_1 and R_2 relaxation rates were measured (11.7 *T*, 25 °C) as described in *Materials and Methods* and plotted in Fig. 2*C* (circles). The primary source of motion leading to ^{15}N relaxation is rotational diffusion of the molecule that can be described by a single rotational correlation time, τ_R, in the limit of isotropic rotation (τ_R is proportional to the molecular volume) (26). Relaxation rates for ^{15}N are also sensitive to very rapid picosecond–nanosecond intramolecular motions of the H^N-N vector, while conformational exchange on the microsecond–millisecond timescale can lead to enhancement of ^{15}N R_2 values (26). Systematic differences in ^{15}N R_1 and R_2 rates of *M* (Fig. 2*C*, red

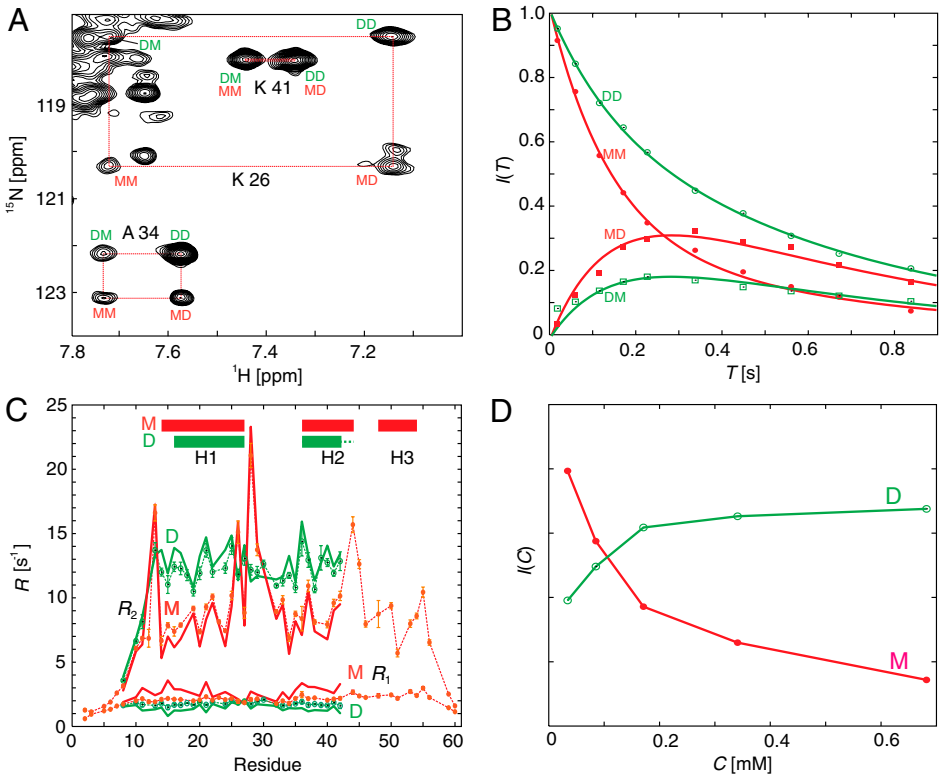

Fig. 2. (*A*) Selected region of a $^1H^N$-^{15}N magnetization exchange spectrum (24, 25) of 0.34 mM FF_{1-60} recorded with a mixing time $T = 0.34$ s (11.7 *T*, 25 °C), showing two sets of autopeaks (*MM* and *DD*) connected by exchange cross-peaks (*MD* and *DM*) for Ala34. (*B*) Mixing time, *T*, dependencies of autopeak (circles) and exchange cross-peak (boxes) volumes for Ala34 from magnetization exchange experiments. The solid lines are generated from a least-squares fit of the exchange data for Ala34 to Eq. **S2** in *SI Materials and Methods*, resulting in $k_{MD} = 3.34 \pm 0.17$ s^{-1} and $k_{DM} = 1.97 \pm 0.09$ s^{-1}; averaging over five residues with a full set of four non-overlapped auto- and cross-peaks gives $k_{MD} = 3.05 \pm 0.95$ s^{-1} and $k_{DM} = 2.16 \pm 0.76$ s^{-1}. (*C*) Apparent backbone ^{15}N R_1 and R_2 relaxation rates as a function of residue for correlations derived from *M* (red circles) and *D* (green circles) states of FF_{1-60} obtained by single-exponential fits of peak volumes in a series of two-dimensional $^1H^N$-^{15}N correlation spectra measured in conventional ^{15}N R_1 and $R_{1\rho}$ experiments (28, 29). As discussed in the text, the apparent ^{15}N R_1 and R_2 rates are affected by slow exchange between *M* and *D* states; intrinsic relaxation rates (solid lines) are calculated using the procedure described in *SI Materials and Methods*. Secondary structural elements are indicated at the top of the panel. Helix H2 of state *D* may extend further than Met42 (denoted by dashed line after residue 42); however, correlations for Ile43 and beyond were missing in spectra of the dimeric form. (*D*) Cross-peak volumes for Ala34 measured from $^1H^N$-^{15}N HSQC spectra of FF_{1-60} recorded as a function of protein concentration. The volumes are normalized by total protein concentration.

circles) and *D* (Fig. 2*C*, green circles) establish that molecules in these states tumble with different rates. The slower R_2 and faster R_1 rates for amide ^{15}N spins of the *M* state are in keeping with expectations for a smaller protein that tumbles with a faster overall rate. Note that several regions of the *M* state, including residues Thr13 (N-cap proceeding helix H1), Lys26-Arg29 (end of H1, H1-H2 loop), and Ile44,Asn45 (end of H2, H2-H3 loop), have enhanced ^{15}N R_2 values, indicative of microsecond–millisecond conformational dynamics.

Values of τ_R can be calculated from ^{15}N R_2/R_1 ratios provided that (*i*) internal motions of the backbone H^N-N vectors are in the extreme narrowing limit (typically less than tens of picoseconds) and/or significantly restricted, and (*ii*) there is no contribution to ^{15}N R_2 from conformational exchange (27). In principle, measured R_1 and R_2 decay curves are biexponential because they are contaminated by the interconversion process, with contributions from intrinsic relaxation rates in both of the exchanging states as well as from k_{MD} and k_{DM} (*SI Materials and Methods*). Intrinsic ^{15}N R_1 rates can be obtained from fits of the time dependencies of auto- and cross-peaks in the longitudinal exchange experiment (24, 25) (Fig. 2*B*); however, as described above, complete time profiles are available for only a small fraction of the residues. We have therefore extracted the intrinsic ^{15}N R_1 and R_2 rates (Fig. 2*C*, solid lines) from the apparent rates obtained from monoexponential fits of decay curves in conventional ^{15}N R_1 and $R_{1\rho}$ experiments (28, 29) using the correction procedure described in *SI Materials and Methods*. After applying the correction, ^{15}N R_2/R_1 ratios for residues 14 to 42 that belong to well-ordered regions of *M* and *D* were used to calculate τ_R for each of the two states, excluding residues with ^{15}N R_2/R_1 ratios outside ± 1 standard deviation from the mean. Values of 5.2 ± 1.2 ns and 10.8 ± 1.1 ns were obtained for states *M* and *D*, respectively, consistent with a slow interconversion between monomeric and dimeric forms of the FF_{1-60} domain.

To obtain additional confirmation that the exchange equilibrium involves a monomer-dimer interconversion, we have carried out a dilution series whereby peak volumes in ^{1}H-^{15}N HSQC spectra from the *M* and *D* states were measured as a function of protein concentration. As expected, peak volumes normalized to protein concentration, decrease for the monomer, and increase for the dimer as a function of concentration, as shown in Fig. 2*D* for Ala34 and Fig. S1 for Ser35. With the "players" involved in exchange now established, the pseudo-first-order rate constants k_{MD} and k_{DM} measured from fits of the magnetization exchange profiles described above can be recast in terms of the rates, k_{on} and k_{off}, corresponding to the dimerization, $2M \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} D$. Values of $k_{on} = 1.10 \pm 0.69 \cdot 10^4$ $M^{-1}\,s^{-1}$; $k_{off} = 2.16 \pm 1.52$ s^{-1} (see *Materials and Methods*), and a dissociation constant $K_d = 0.20 \pm 0.05$ mM are obtained. Note that the measured k_{on} rate is several orders of magnitude slower than that expected for diffusion-controlled protein association (30), approximately 10^6 $M^{-1}\,s^{-1}$, suggesting a rate-limiting step for dimerization that involves at least some rearrangement of structure.

Structural Differences Between the Monomer and Dimer. Backbone NMR chemical shifts are very sensitive probes of local protein conformation (31). A comparison of such shifts between *M* and *D* states can therefore inform on the structural changes accompanying dimerization. Fig. 3*A* plots the cumulative change in the backbone $^1H^N$, ^{15}N, $^1H^\alpha$, $^{13}C^\alpha$, and $^{13}C^O$ chemical shifts (32), $\Delta\varpi'_{DM} = \sqrt{\frac{1}{N}\sum_i (\frac{\Delta\varpi_i}{\Delta\varpi_{std,i}})^2}$, as a function of residue, where $\Delta\varpi_i$ is chemical shift difference between the monomeric and dimeric forms of FF_{1-60}, $\varpi_{std,i}$ is a nucleus- and residue-specific normalization constant corresponding to one standard deviation of the chemical shift range listed in the BMRB database, *N* is the number of backbone chemical shifts available for a given residue, and index *i* runs from 1 to *N*. Close to zero $\Delta\varpi'_{DM}$ values are expected in regions of the protein with little conformational change between the states. In contrast, a stretch of $\Delta\varpi'_{DM}$ values on the order of 1 or higher points to a structural rearrangement. It is clear from Fig. 3*A* that the largest chemical shift differences upon dimer formation are localized to a region that includes residues Glu15-Lys18 at the beginning of α-helix H1 of the monomeric form. Elevated $\Delta\varpi'_{DM}$ values are also observed for residues Leu25-Arg29 at the end of helix H1, with the H1-H2 loop and α-helix H2 displaying smaller chemical shift changes. We are unable to assess the extent of chemical shift changes at the end of α-helix H2, the H2-H3 loop, and for α-helix H3 of FF_{1-60} because resonances from this region of the dimer are missing in NMR spectra.

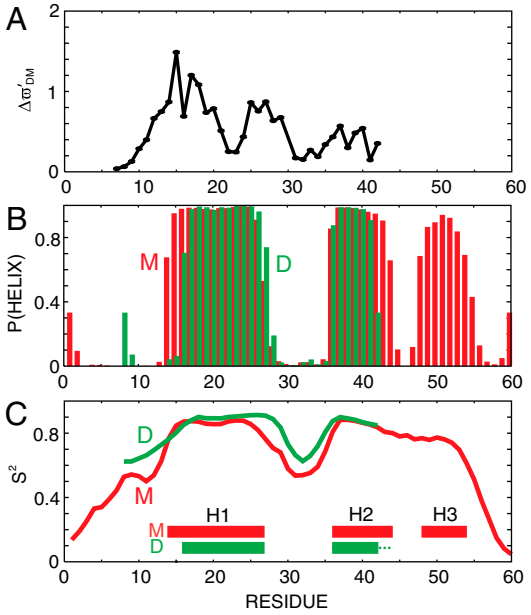

Fig. 3. (*A*) Backbone chemical shift differences between monomeric (M) and dimeric (D) forms of FF_{1-60}, $\Delta\varpi'_{DM}$, as described in the text. Values of $\Delta\varpi'_{DM}$ are shown for residues Thr8-Met42; the chemical shifts of the first seven residues are indistinguishable in the two forms, while those from the C terminus (residues Ile43-Gln60) are missing in NMR spectra of the dimer. (*B*) TALOS-plus predicted (33) α-helix probability, *P* (helix), plotted vs. residue for states *M* (red) and *D* (green, residues Thr8-Met42 only). (*C*) RCI-predicted (34, 35) order parameters, S^2, for the backbone amide groups of *M* (red) and *D* (green).

Fig. 3*B* compares the secondary structures of the monomeric (red) and dimeric (green) (residues 8–42 only) forms of FF_{1-60} predicted from measured backbone chemical shifts using the TALOS+ program (33). As expected, the monomer comprises three α-helices encompassing residues Lys14-Glu27 (H1), Trp36-Ile44 (H2), and Arg48-Lys54 (H3), closely matching the corresponding helix positions in the folding intermediate of the full-length domain. Similar secondary structure is predicted for Thr8-Met42 of the dimer, including helices H1 (Glu16-Glu27) and H2 (Trp36-Met42). It is noteworthy that α-helix H1 is shorter by two residues in the dimer, starting at Glu16 as opposed to Lys14. This is in line with the observation that Glu15, Ala17, and Lys18 have the largest backbone chemical shift changes upon dimer formation (Fig. 3*A*).

The backbone chemical shifts were also used to predict conformational flexibility of the two forms of FF_{1-60} using the random coil index (RCI) approach (34, 35). RCI-based order parameters, S^2, for the backbone amide groups of *M* (red) and *D* (green) are plotted vs. residue in Fig. 3*C*. Note that extreme S^2 values of 1 and 0 correspond to fully restricted and unconstrained internal

motion of the backbone H^N-N group, respectively. Values of S^2 for the monomeric form are nearly identical to those of *I*, obtained in a previous study (13). Interestingly, the dimeric form of $FF_{1–60}$ is even less flexible than the monomer. Although helix H1 in the dimer is shorter by two residues (at the N terminus), the region from Thr8-Glu16 proceeding H1 is more ordered in the dimer than in the monomer. Similarly, elevated S^2 values are observed at the end of helix H1 of the dimer (residues Leu25-Arg29), pointing to restricted mobility in this region as well. The RCI method is sensitive to a timescale ranging from picosecond–nanosecond to approximately hundreds of microseconds (35). Restricted mobility in this time window does not exclude the possibility of slower conformational dynamics on a microsecond–millisecond timescale; although reasonably high S^2 values are noted for Lys26-Arg29 and Ile44,Asn45 (>0.7), these residues do undergo slow conformational exchange as established by elevated ^{15}N R_2 rates (Fig. 2*C*).

Chemical shift changes between *M* and *D* states (Fig. 3*A*) and the S^2 vs. residue profile (Fig. 3*C*) provide a strong indication as to which regions of $FF_{1–60}$ might participate in dimer formation. The large $\Delta\varpi'_{DM}$ values and higher-order parameters for residues N-terminal of helix H1 in the dimer relative to those in the monomer potentially point to this region as a site for dimerization that increases in order upon dimer formation. To test this hypothesis we have analyzed FF domain variants prepared with truncated C termini (residues Ala61-Lys71) and with varying truncations at the N terminus: $FF_{1–60}$, $FF_{7–60}$, $FF_{9–60}$, and $FF_{11–60}$ (the first residue in each of the constructs is mutated to Gly). For the first three variants ($FF_{1–60}$, $FF_{7–60}$, and $FF_{9–60}$), $^1H^N$-^{15}N HSQC spectra are very similar, with two sets of signals corresponding to states *M* and *D* of the protein, while only a single set of signals is observed in NMR spectra of $FF_{11–60}$ that belong to the monomeric form. We have previously determined the structure of $FF_{11–60}$ using a standard NOE-based approach, establishing that it is a very good structural mimic of the folding intermediate of the full-length domain (22). The finding that truncation of the N-terminal residues of the FF domain, and specifically elimination of Thr10 and Trp11, leads to a shift in the monomer-dimer equilibrium toward the monomer form suggests that these residues are responsible for stabilization of the dimer. It is likely that residues N-terminal to helix H1 also contribute to stability of the monomer because $FF_{11–60}$ is significantly more flexible than $FF_{1–60}$ based on lower RCI-derived S^2 values (22).

Concluding Remarks. The development of relaxation dispersion NMR spectroscopy (5, 6) opens the possibility for detailed structural studies of sparsely populated, transiently formed intermediates that have been recalcitrant to study using standard biophysical approaches. Key to the success of this method is that the invisible, excited state is probed via the visible, ground state that gives rise to high-quality spectra. The methodology is limited, however, to the study of states that interconvert on the millisecond timescale and to exchange reactions involving one or two excited conformers. Thus, only a small portion of the energy landscape can be explored. One approach to increasing the number of states accessible to study would be to perturb the ground state so that it is no longer populated. In so doing, a new ground state is formed (formerly a higher-energy state) that can then be used to probe further exchange events involving additional conformers using a variety of different NMR experiments that are sensitive to dynamics spanning a window of over 12 orders of magnitude (36). This is the strategy taken here. Atomic resolution structures of both the native state and a folding intermediate of $FF_{1–71}$, the latter determined by Carr–Purcell–Meiboom–Gill relaxation dispersion, have guided the design of a new ground state, $FF_{1–60}$, that corresponds to the intermediate structure. This new variant is used to further probe regions of the folding/unfolding landscape that would otherwise be hidden by the native state.

Here, we have demonstrated that $FF_{1–60}$ undergoes slow dimerization on the seconds timescale with a K_d of 0.2 mM, and with kinetics that are significantly slower than diffusion-limited. Further truncation mutations of $FF_{1–60}$ involving the amino terminus establish that dimerization proceeds through interactions involving the first 11 residues of the domain, in particular Thr10 and Trp11. Both monomer and dimer are relatively rigid on the picosecond–nanosecond to microsecond timescale and are significantly more dynamic on the slower microsecond–millisecond timescale, leading to exchange line broadening in NMR spectra. Notably, several regions of the monomer have enhanced ^{15}N R_2 rates, while resonances from the C-terminal region of the dimer are missing in NMR spectra, presumably due to microsecond–millisecond exchange with additional conformational states.

The observed monomer-dimer equilibrium and microsecond–millisecond dynamics of $FF_{1–60}$ are consistent with the FF domain folding intermediate serving as a branch point leading to either folding to the native conformation or formation of further misfolded states. Although we have no direct evidence that the dimeric form of $FF_{11–60}$ further aggregates to form higher oligomeric species, the fact that the C-terminal third of the dimer cannot be observed in NMR spectra strongly suggests that this region of the protein samples additional conformational states. In this regard it is of interest to note that in many of the systems studied to date for which models of aggregation/fibrillation have been obtained, the first step is dimerization, followed by rearrangements to accommodate higher-order oligomerization (3). One such example involves β2-microglobulin that forms fibrils in dialysis-related amyloidosis (37). The mechanism for the assembly of such fibrils involves the interconversion of the native conformation of the protein to one that is native-like, N^*, by Cu^{2+} that is present at high concentrations in hemodialysate. N^* then forms dimers that subsequently assemble into fibrils. Partially unfolded or misfolded conformers that form via thermal fluctuations from the ground state are thought to be implicated in a number of aggregation-related diseases (1–3). The approach presented here provides a way of characterizing such states. The finding that a sparsely populated intermediate of the FF domain folding pathway can initiate dimerization and the characterization of the resultant nonnative dimeric state is an important step toward unraveling the role of folding intermediates in both protein folding and misfolding pathways.

Materials and Methods

Protein Samples. Uniformly ^{15}N- and $^{15}N/^{13}C$-labeled samples of truncated variants of the first FF domain from HYPA/FBP11 were produced as described previously (13, 17). The plasmid-encoding $FF_{1–60}$ (residues 1–60) was obtained using the QuikChange protocol (Stratagene) by introducing a stop codon after position 60 in the gene encoding the full-length domain cloned into a modified pRSET vector (13). The plasmids for FF domain variants with truncated N termini, $FF_{7–60}$, $FF_{9–60}$, and $FF_{11–60}$, were produced from a plasmid encoding $FF_{1–60}$ as described elsewhere (22). All NMR experiments were performed on samples with protein concentrations of approximately 0.3 mM, 50 mM sodium acetate, 100 mM NaCl, pH = 4.9, 25 °C.

Backbone Resonance Assignment. The backbone ^{15}N, $^1H^N$, $^{13}C^\alpha$, $^1H^\alpha$, and $^{13}C^O$ resonance assignments for the *M* and *D* states of $^{15}N/^{13}C$-labeled $FF_{1–60}$ were obtained using a standard set of triple-resonance NMR experiments (38) as described previously (13). All data were recorded on a Varian Inova spectrometer, 11.7 *T*, 25 °C.

Magnetization-Exchange Measurements. Nitrogen-15 longitudinal exchange experiments (24, 25) were recorded on a 0.34 mM $FF_{1–60}$ sample, 14.0 *T*, 25 °C. A series of 10 two-dimensional $^1H^N$-^{15}N correlation spectra was measured with mixing times *T* for relaxation and exchange of ^{15}N longitudinal magnetization ranging from 0.014 to 0.84 s. Volumes of auto- and exchange cross-peaks were quantified for five residues (Glu15, Ala17, Phe21, Ala34, and Ser35) for which well-resolved correlations (corresponding to *MM*, *DD*, *MD*, and *DM* peaks; see text) were obtained. The extracted first-order rate constants, k_{MD} and k_{DM}, were used to calculate k_{on} and k_{off} according

to $k_{on} = [k_{MD}(k_{MD} + k_{DM})]/[2k_{DM}P_0]$, $k_{off} = k_{DM}$, where P_0 denotes the total protein concentration. The dimer dissociation constant K_d was calculated as k_{off}/k_{on}. Analysis of data is as described in *SI Materials and Methods*.

Nitrogen-15 Relaxation Measurements. Measurements for ^{15}N R_1 and $R_{1\rho}$ were performed on a 0.32 mM FF_{1-60} sample, 11.7 *T*, 25 °C, as described previously (28, 29). The apparent rotating-frame $R_{1\rho}$ and transverse R_2 relaxation rates were obtained by single exponential fits of peak volumes in a series of two-dimensional $^1H^N$-^{15}N correlation spectra measured as a function of relaxation delay *T*. Transverse relaxation rates R_2 were calculated from R_1 and $R_{1\rho}$ using the equation $R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta$, where $\theta = \arctan(\omega_{SL}/\Delta\omega)$, $\Delta\omega$ is the resonance offset from the spin-lock carrier, and ω_{SL} is the spin-lock field strength in the ^{15}N $R_{1\rho}$ experiments ($\omega_{SL} = 1.82$ KHz). Intrinsic relaxation rates were calculated from the apparent R_1 and R_2 values as described in *SI Materials and Methods*.

Dilution Series. The $^1H^N$-^{15}N HSQC spectra of FF_{1-60} were recorded as a function of protein concentration ranging from 0.68 to 0.034 mM. Volumes of non-overlapped peaks corresponding to *M* and *D* forms of FF_{1-60} were quantified and normalized to the total protein concentration.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant P30GM092369 and University of Connecticut Health Center startup funds (to D.K.) and a Canadian Institutes of Health Research (CIHR) Grant (L.E.K.). T.L.R. was funded through a postdoctoral fellowship from CIHR. L.E.K. holds a Canada Research Chair in Biochemistry.

1. Dobson CM (2003) Protein folding and misfolding. *Nature* 426:884–890.
2. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
3. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5:15–22.
4. Brockwell DJ, Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* 17:30–37.
5. Korzhnev DM, Kay LE (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Acc Chem Res* 41:442–451.
6. Palmer AG, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* 339:204–238.
7. Hansen DF, Vallurupalli P, Lundström P, Neudecker P, Kay LE (2008) Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *J Am Chem Soc* 130:2667–2675.
8. Vallurupalli P, Hansen DF, Stollar E, Meirovitch E, Kay LE (2007) Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci USA* 104:18473–18477.
9. Vallurupalli P, Hansen DF, Kay LE (2008) Probing structure in invisible protein states with anisotropic NMR chemical shifts. *J Am Chem Soc* 130:2734–2735.
10. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615–9620.
11. Shen Y, et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685–4690.
12. Wishart DS, et al. (2008) CS23D: A web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496–W502.
13. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE (2010) A transient and low-populated protein folding intermediate at atomic resolution. *Science* 329:1312–1316.
14. Korzhnev DM, et al. (2011) Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: An NMR relaxation dispersion study. *J Am Chem Soc* 133:10974–10982.
15. Bouvignies G, et al. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477:111–114.
16. Jemth P, et al. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc Natl Acad Sci USA* 101:6450–6455.
17. Jemth P, et al. (2005) The structure of the major transition state for folding of an FF domain from experiment and simulation. *J Mol Biol* 350:363–378.
18. Korzhnev DM, Religa TL, Lundström P, Fersht AR, Kay LE (2007) The folding pathway of an FF domain: Characterization of an on-pathway intermediate state under folding conditions by ^{15}N, $^{13}C^{\alpha}$ and ^{13}C-methyl relaxation dispersion and $^1H/^2H$-exchange NMR spectroscopy. *J Mol Biol* 37:497–512.
19. Feng HQ, Zhou Z, Bai YW (2005) A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci USA* 102:5026–5031.
20. Religa TL, Markson JS, Mayor U, Freund SMV, Fersht AR (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* 437:1053–1056.
21. Whittaker SBM, Spence GR, Grossmann JG, Radford SE, Moore GR (2007) NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* 366:1001–1015.
22. Barette J, Velyvis A, Religa TL, Korzhnev DM, Kay LE (2012) Cross-validation of the structure of a transiently formed and low populated FF domain folding intermediate determined by relaxation dispersion NMR and CS-Rosetta. *J Phys Chem B*, in press.
23. Allen M, Friedler A, Schon O, Bycroft M (2002) The structure of an FF domain from human HYPA/FBP11. *J Mol Biol* 323:411–416.
24. Farrow NA, Zhang OW, Forman-Kay JD, Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical-exchange rates of systems in slow equilibrium. *J Biomol NMR* 4:727–734.
25. Tollinger M, Skrynnikov NR, Mulder FAA, Forman-Kay JD, Kay LE (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* 123:11341–11352.
26. Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY (2001) NMR studies of Brownian tumbling and internal motions in proteins. *Prog Nucl Magn Reson Spectrosc* 38:197–266.
27. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* 28:8972–8979.
28. Farrow NA, et al. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33:5984–6003.
29. Korzhnev DM, Skrynnikov NR, Millet O, Torchia DA, Kay LE (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* 124:10743–10753.
30. Schreiber G (2002) Kinetic studies of protein-protein interactions. *Curr Opin Struct Biol* 12:41–47.
31. Wishart DS (2011) Interpreting protein chemical shift data. *Prog Nucl Magn Reson Spectrosc* 58:62–87.
32. Vallurupalli P, Hansen DF, Kay LE (2008) Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc Natl Acad Sci USA* 105:11766–11771.
33. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223.
34. Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J Am Chem Soc* 127:14970–14971.
35. Berjanskii MV, Wishart DS (2008) Application of the random coil index to studying protein flexibility. *J Biomol NMR* 40:31–48.
36. Mittermaier AK, Kay LE (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34:601–611.
37. Eichner T, Kalverda AP, Thompson GS, Homans SW, Radford SE (2011) Conformational conversion during amyloid formation at atomic resolution. *Mol Cell* 41:161–172.
38. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34:93–158.