Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding

Philipp Neudecker, Patrik Lundström, and Lewis E. Kay*
Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ABSTRACT Characterization of the mechanisms by which proteins fold into their native conformations is important not only for protein structure prediction and design but also because protein misfolding intermediates may play critical roles in fibril formation that are commonplace in neurodegenerative disorders. In practice, the study of folding pathways is complicated by the fact that for the most part intermediates are low-populated and short-lived so that biophysical studies are difficult. Due to recent methodological advances, relaxation dispersion NMR spectroscopy has emerged as a particularly powerful tool to obtain high-resolution structural information about protein folding events on the millisecond timescale. Applications of the methodology to study the folding of SH3 domains have shown that folding proceeds via previously undetected on-pathway intermediates, sometimes stabilized by nonnative long-range interactions. The relaxation dispersion approach provides a detailed kinetic and thermodynamic description of the folding process as well as the promise of obtaining an atomic level structural description of intermediate states. We review the concerted application of a variety of recently developed NMR relaxation dispersion experiments to obtain a "high-resolution" picture of the folding pathway of the A39V/N53P/V55L Fyn SH3 domain.

INTRODUCTION

An understanding of the processes by which an unfolded protein acquires its native state conformation remains a very significant and challenging biophysical problem, even 50 years after the seminal investigations of Anfinsen showed that proteins such as ribonuclease fold spontaneously in a matter of several seconds or faster (1). The timescale of protein folding is remarkable considering the vast number of conformations that potentially are accessible to a polypeptide chain. Moreover, globular proteins are only marginally stable, with free energies of folding on the order of $\Delta G_{U\rightarrow F} = G_F - G_U \approx -50$ kJ/mol, which corresponds to the energy contribution from only a handful of favorable noncovalent interactions out of the thousands found in a globular protein (2). Yet, protein folding appears to be a highly cooperative event. In fact, most experimental kinetic studies of independently folding protein units or ''domains'' are consistent with a surprisingly simple two-state transition $U \leftrightarrow F$ connecting the unfolded, U, and native, F, states (3–5). In this approximation an understanding of the folding pathway at a structural level involves the characterization of only a single transition state (TS), TS(UF), which can be probed via the effects of point mutations on the folding kinetics in an approach called Φ-value analysis (4,6), or via a related method termed Ψ-value analysis (7). However, experimental detection of additional states along the folding pathway such as intermediates I is often very difficult because they are both low and transiently populated (4,8). The use of denaturants in stopped-flow mixing and equilibrium unfolding experiments can further lead to destabilization of compact intermediates, resulting in their rapid pre-equilibration in kinetic experiments during the instrument dead-time of typically a few milliseconds (9). Accordingly, many recent studies, often invoking additional methods such as ultra-fast mixing (10) or temperature jump (11) to resolve the rapid pre-equilibration stage, have identified more complex folding pathways with intermediate states, suggesting that folding energy landscapes tend to be more rugged than previously anticipated (8). Even with these methods to resolve fast kinetics, however, there remain rather few studies where Φ-values have been measured for the two transition state ensembles and the intermediate of even relatively ''simple'' three-state folding reactions (12,13). In this regard, native-state hydrogen exchange experiments are particularly powerful in providing detailed structural information about one or more I states on a per-residue basis, even in cases when the intermediates are populated at extremely low levels that would be invisible to the kinetic experiments mentioned above (14–16).

The development of modern multidimensional, multinuclear NMR spectroscopy has proven extremely valuable to the study of protein folding, and has played an important role in many biophysical studies including the native-state hydrogen exchange method mentioned above (14–16). NMR has emerged as a particularly powerful technique to identify and quantify folding intermediates because an abundance of site-specific probes can be studied simultaneously in a single one- or multidimensional NMR spectrum. Experiments can be recorded as kinetic measurements that monitor the return to equilibrium after an initial perturbation (such as a change in denaturant concentration, pH, temperature, or solvent protonation). For example, very slow folding events can be monitored directly in real-time (17), and recently developed fast acquisition methods allow the recording of

Submitted November 1, 2008, and accepted for publication December 12, 2008.

*Correspondence: kay@pound.med.utoronto.ca

Editor: Edward H. Egelman.

0006-3495/09/03/2045/10 $2.00

doi: 10.1016/j.bpj.2008.12.3907

two-dimensional NMR experiments for real-time folding studies with a time resolution of a few seconds (18). For folding events on the more common millisecond to second timescale, proton/deuterium amide hydrogen exchange NMR spectroscopy studies have proven particularly successful in identifying intermediates (14–16). In favorable cases the site-specific resolution afforded by hydrogen exchange NMR spectroscopy provides a rational basis for stabilizing an excited intermediate state through protein engineering so that it becomes the ground state and hence accessible to conventional high-resolution structure determination, as demonstrated in several elegant studies on redesigned apo-cytochrome b_{562} (19), En-HD (20), T4 lysozyme (21) and RNase H (22).

It has also long been recognized that NMR spectroscopy is a very powerful method to quantify chemical exchange kinetics of systems in equilibrium (23). Although the basic experimental ideas were formulated over a half-century ago (24) and applications to very simple biological systems have appeared for many years, detailed studies of protein conformational exchange had to await the development of new experiments and labeling schemes (25–28). In recent years our laboratory has studied equilibrium protein folding reactions to elucidate the folding pathways of a number of small single domain proteins. One such system, which has been studied by a large number of laboratories, is the SH3 domain, a module that folds into a five-stranded β-sandwich structure (29). Notably, studies that include calorimetric, equilibrium and kinetic folding/unfolding experiments are consistent with a two-state folding mechanism, with no evidence for the formation of partially folded intermediates (5,30). By contrast, equilibrium NMR experiments that provide kinetic, thermodynamic and structural information on the folding reactions of a number of mutants of the Abp1p and Fyn SH3 domains can only be interpreted in terms of the formation of a low-populated intermediate, which must be on the folding pathway (9,31,32). In the following review, we briefly highlight some of the qualitative features of the NMR methodology that led to the identification of intermediates of Fyn SH3 domain folding that were ''invisible'' to other more established techniques, and then present a synopsis of a number of NMR studies that have provided detailed atomic level insight into the folding mechanism of the Fyn SH3 domain.

NMR relaxation dispersion experiments

Before describing the underlying physical basis of the NMR methodology that is used to study equilibrium conformational exchange processes, we first discuss how chemical exchange events can affect NMR spectra in general. Consider the exchange between the predominantly populated (ground) state, A, and an excited state, B, $A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B$, with exchange rate $k_{ex,AB} = k_{AB} + k_{BA}$. The states are populated according to the Boltzmann distribution

$$\frac{p_B}{p_A} = \frac{k_{AB}}{k_{BA}} = e^{-\frac{\Delta G_{A\rightarrow B}(T)}{RT}}, \quad (1)$$

where p_A and p_B are the fractional populations of states A and B, respectively, R is the gas constant, and T the absolute temperature. The temperature dependence of the free energy difference between states is given by

$$\begin{aligned}\Delta G_{A\rightarrow B}(T) &= G_B(T) - G_A(T)\\ &= \Delta H_{A\rightarrow B}(T_m) + \Delta C_p(A\rightarrow B) \times (T - T_m)\\ &\quad -T\Delta S_{A\rightarrow B}(T_m) - T\Delta C_p(A\rightarrow B)\ln\tfrac{T}{T_m},\end{aligned} \quad (2)$$

where $\Delta H_{A\rightarrow B}(T_m)$, $\Delta S_{A\rightarrow B}(T_m)$, and $\Delta C_p(A\rightarrow B)$ are the enthalpy, entropy, and heat capacity differences, respectively, between states B and A at an arbitrary reference temperature, T_m (4). Assuming transition-state theory, the temperature dependence of the rate of transition from state A to state B, separated by the rate-limiting transition state TS(AB), is given by the Eyring equation:

$$k_{AB} = \kappa\frac{k_B T}{h}e^{-\frac{\Delta G_{A\rightarrow TS(AB)(T)}}{RT}}, \quad (3)$$

where κ is a transmission coefficient, and k_B and h are the Boltzmann and Planck constants, respectively (33). Suppose that an NMR-active probe on the protein (such as a ^{15}N or ^{1}HN spin, for example) resonates at distinct frequencies (chemical shifts) in the ground and excited states, denoted by ω_A and ω_B (in units of rad/s), respectively, or alternatively by ϖ_A and ϖ_B (in ppm). Chemical shift differences $\Delta\omega_{AB} = \omega_B - \omega_A$ normally range from 0 to $\approx$6000 rad/s, depending, of course, on the NMR probe, the nature of the exchange process, and the spectrometer frequency at which the experiment is recorded. If the exchange rate, $k_{ex,AB}$, is very slow compared to $\Delta\omega_{AB}$ ($k_{ex,AB} < \approx 0.05 \times \Delta\omega_{AB}$; slow exchange regime) then the NMR spectrum contains separate, well-defined resonances for each of the two states. In cases where $k_{ex,AB}$ is on the order of 1/s–10/s the exchange kinetics is most easily quantified via magnetization exchange experiments (25,26). Conversely, if the exchange rate $k_{ex,AB}$ is similar to the chemical shift difference, $\approx$100/s to 2000/s (intermediate exchange regime), then the exchange causes significant line-broadening of signals from both states. The low-populated excited state resonance usually becomes too broad to be observed but the line-broadening of the ground state resonance still reports on the exchange with the invisible excited state conformer and such line-broadening can be quantified to extract out the kinetics and thermodynamics of the exchange event as well as the excited state chemical shifts, which can be interpreted in terms of structure.

Many NMR experiments can be understood without recourse to a rigorous quantum-mechanical description, including those to study equilibrium chemical exchange processes. Upon insertion of the NMR sample into a large external magnetic field the sum over all magnetic dipoles in the sample produces net magnetization, which aligns

with the field (taken to be the z axis). Subsequent application of a radio-frequency pulse at the appropriate frequency places the magnetization into the x-y plane, as indicated in Fig. 1 *A*, where a 90°_y pulse rotates magnetization to the x axis. In the absence of further pulses the magnetization subsequently rotates (precesses) around the z axis with a characteristic angular frequency (chemical shift) ω, so that over time it will accrue a phase $\phi = \omega t$ with respect to the x axis. Let us first consider the case of a protein that exists in two discrete conformations which do not interconvert, and follow the course of a single NMR-active nucleus with different chemical shifts, ω_A and ω_B, in the two states. Magnetization at sites with different chemical shifts will rotate with different frequencies and thus become increasingly defocused with time. Now, consider the case where after an evolution period of time $T/2$ an additional pulse is applied, one that inverts the accrued phase of the magnetization at each site (a so-called 180°_x pulse), regardless of precession frequency. During the successive delay $T/2$ the phase differences between magnetization from sites A and B decrease so that at time T the phase difference is zero and both magnetization components are returned to the x axis (Fig. 1 *B*). The NMR spectrum for such a nonexchanging system is shown in the inset to Fig. 1 *B*, where the intensities of each of the resonance lines provide a faithful readout of the relative populations of the two states. Next, consider the case where exchange between ground (A) and excited (B) states, $A \leftrightarrow B$, causes the precession frequency of an NMR spin to vary stochastically between ω_A and ω_B. The stochastic nature means that the accrued phases for magnetization originating on different molecules in general are not the same and hence the signal, which is the vector sum of the magnetization from all of the molecules in the sample, broadens in the frequency domain and decreases in intensity (*inset* to Fig. 1 *C*). For the same reason, application of a 180°_x pulse in the middle of the evolution period does not, in general, lead to a complete refocusing of magnetization (Fig. 1 *C*). However, as more and more refocusing pulses are applied during T, refocusing improves progressively as long as the rate of application of pulses is on the order of or faster than the exchange rate (Fig. 1 *D*). The idea behind this so-called Carr-Purcell-Meiboom-Gill (CPMG) (34,35) relaxation dispersion (RD) experiment is to monitor the signal intensity $I(\nu_{CPMG})$ from spin probes of the observable ground state, A, as a function of the repetition rate of the refocusing pulses, defined as $\nu_{CPMG} = 1/(2\tau)$, where τ is the separation between the pulses. $I(\nu_{CPMG})$ is then converted into an effective transverse relaxation rate that quantifies the rate of decay of signal, $R_{eff}(\nu_{CPMG}) = -\ln(I(\nu_{CPMG})/I_0)/T$, where I_0 is a reference signal intensity. These relaxation dispersion profiles $R_{eff}(\nu_{CPMG})$ (Fig. 2) depend in a nonlinear fashion on the rates of the exchange process (k_{AB} and k_{BA}, or equivalently $k_{ex,AB}$ and p_B) and on the absolute chemical shift differences $|\Delta\varpi_{AB}|$ according to the Bloch-McConnell equations (24). The complete set of exchange parameters can be

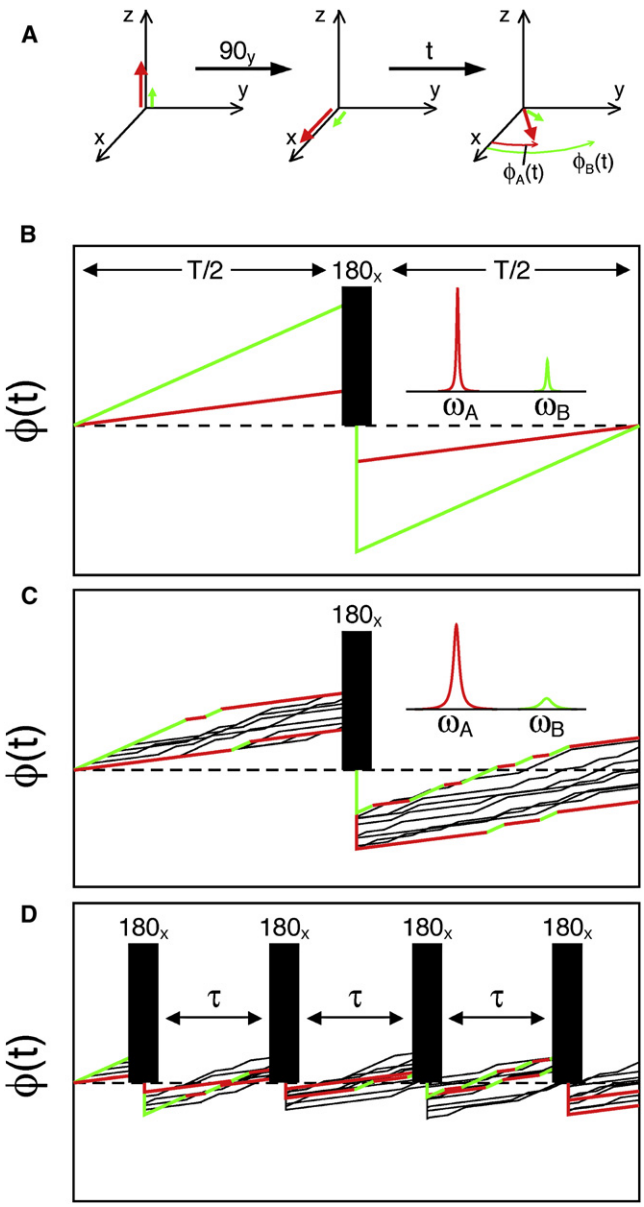

FIGURE 1 NMR spectroscopy of a single probe attached to a protein molecule that exists in two discrete conformations highlighted in red and green. The probe resonates with different characteristic NMR frequencies, ωA and ωB, in each of the two states. (*A*) A pulse is applied to equilibrium magnetization (along z) that places it along the x axis. The magnetization subsequently rotates around the z axis at its characteristic angular frequency, ω, so that after a time t the accrued phase with respect to the x axis is $\phi = \omega t$. (*B*) In the absence of chemical exchange the evolution of magnetization with different precession frequencies is refocused when a 180°_x pulse is applied in the middle of a fixed delay T. The NMR spectrum of the system is shown as an inset. (*C*) In the presence of chemical exchange the precession frequency stochastically jumps between ω_A (red) and ω_B (green) with a different trajectory for each molecule. Ten such trajectories are shown with two highlighted in color to indicate the timecourse of the jumps. The exchange process leads to line-broadening, as can be seen in the spectrum shown in the inset. (*D*) The application of successively increasing numbers of 180°_x pulses during the same delay improves refocusing and hence minimizes the effects of chemical exchange.

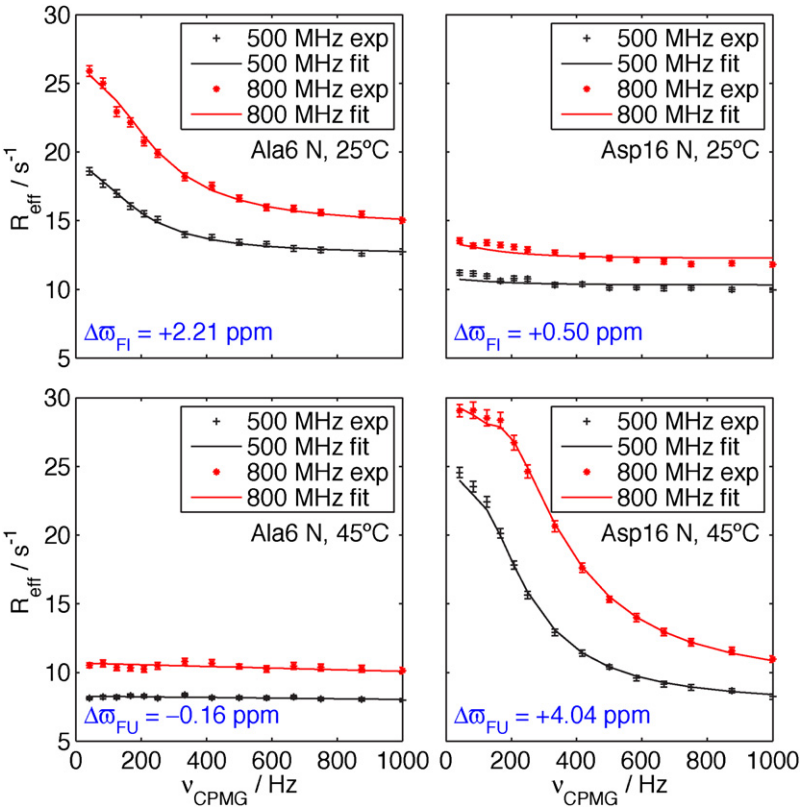

FIGURE 2 ^{15}N RD profiles for Ala-6 (*left*) and Asp-16 (*right*) of the A39V/N53P/V55L Fyn SH3 domain at 25°C (*top*) and 45°C (*bottom*). Continuous lines are the predictions from fitting a global three-state folding model U↔I↔F to all data recorded in 5°C increments between 10°C and 45°C. Adapted from Neudecker et al. (9).

obtained by least-squares fitting of the RD profiles, provided that the exchange is in the intermediate regime. The sign of $\Delta\varpi_{AB}$ can also be obtained experimentally (28), and hence the chemical shifts of the invisible excited state according to $\varpi_B = \varpi_A + \Delta\varpi_{AB}$. New isotope-labeling schemes have been developed along with improved CPMG RD pulse sequences so that accurate excited state chemical shifts can be extracted for backbone ^{1}HN, ^{15}N, carbonyl ^{13}CO, ^{1}Hα, ^{13}Cα and methyl group resonances (25–28,36).

RD experiments are very sensitive to excited states populated to as little as $\approx$0.5%, corresponding to a free energy difference between exchanging states of $<\approx$13 kJ/mol, so long as these states exchange on the millisecond timescale and the exchange process is accompanied by sizable chemical shift changes $|\Delta\varpi_{AB}|$ (37). A significant number of protein folding equilibria satisfy these criteria. It is worth noting that one of the major advantages of the RD NMR method over approaches that monitor folding/unfolding kinetics as a function of denaturant concentration is that the NMR studies can be performed in the absence of denaturant or other perturbations that might affect intermediate states.

Choice of the A39V/N53P/V55L Fyn SH3 domain to probe protein folding

Over the past decade a large number of folding studies of SH3 domains have emerged that have made use of a wide variety of biophysical techniques, concluding that the folding reaction could be described by a two-state U↔F mechanism with no detectable accumulation of intermediates (5,30). NMR RD experiments cannot be performed on the wild-type Fyn SH3 domain because the folding exchange rate $k_{ex,FU}$ is too small and the stability of the domain too high; however, experiments performed on a variety of G48 mutants of the Fyn and Abp1p SH3 domains designed to have folding rates and stabilities suitable for RD analysis while preserving the native fold all showed very strong evidence of an on-pathway intermediate I (31,32). Is this on-pathway intermediate merely a consequence of the choice of the site of mutation (position 48) or does it correspond to a state sampled during folding of wild-type SH3 domains in general? A number of arguments suggest that it is the latter. First, the structural features of the intermediate state for Fyn and Abp1p folding, as established by chemical shifts measured from RD, agree well with the picture of the transition state ensemble of SH3 domain folding emerging from Φ-value analysis using a variety of wild-type SH3 domains (31,32,38–40). Second, computational studies on the src SH3 domain show that folding proceeds through an intermediate state with similar properties to those of the Fyn SH3 domain mutants deduced by RD (41). Finally, we have examined an additional pair of mutants, N53P/V55L and A39V/N53P/V55L Fyn SH3, which preserve Gly-48 and are amenable to RD experiments. Both of these mutants were shown to fold via a low-populated on-pathway intermediate with many of the structural properties of the

intermediates of G48 mutants characterized previously, but with some differences as well, as described in detail below (9,32). Of these two mutants, the A39V/N53P/V55L Fyn SH3 domain was of particular interest since i), at temperatures below 30°C the three-state folding process U↔I↔F established by RD measurements was effectively reduced to a two-state exchange reaction I↔F, thus making the intermediate particularly accessible experimentally and ii), a thorough analysis of the ^{15}N chemical shifts of the intermediate, ϖ_I, revealed the presence of pronounced nonnative long-range interactions (9). The tremendous success of protein folding studies by methods that are interpreted primarily in the context of the formation of native contacts such as Φ-value analysis (4,6) and native-centric computer simulations (42) might lead one to suggest that folding is predominantly driven by native interactions and that evolution has largely selected against nonnative contacts that could cause kinetic traps. However, many recent experimental and theoretical studies have provided increasing evidence that nonnative interactions have a significant role in protein folding (8,15) and can even speed up the process; in fact, the A39V/N53P/V55L Fyn SH3 domain is a case where nonnative interactions involving residue 53 lead to an increase in the folding rate (43). The exact role of nonnative interactions in folding intermediates is still poorly understood and of considerable interest, especially because folding intermediates are suspected to be the precursors for the amyloid fibrils associated with many neurodegenerative diseases (44).

An on-pathway folding intermediate

RD experiments are very sensitive to the presence of intermediate states, partly because a large number of sites in the protein can be studied simultaneously. In principle, each site monitors the exchange event and for the case of a three-state folding reaction, U↔I↔F, site-specific chemical shift differences $|\Delta\varpi_{FI}|$ and $|\Delta\varpi_{FU}|$ are obtained. Just by coincidence, some sites will have large $|\Delta\varpi_{FI}|$ but $|\Delta\varpi_{FU}|$ close to 0 and these predominantly probe the I↔F exchange event, whereas others will have $|\Delta\varpi_{FI}|$ close to 0 but large $|\Delta\varpi_{FU}|$ so that the U↔F process dominates. As a consequence, fitting the RD profiles of every resonance individually assuming a two-state model will result in a very inhomogeneous distribution of exchange rates and excited state populations, and when all residues are fit simultaneously to a global two-state folding model a very elevated fitting residual χ^2 is obtained. Such findings should be taken as an indication that the exchange process is more complex than two-state (31,32,37). In the case of the A39V/N53P/V55L Fyn SH3 domain identification of the intermediate was particularly clear. Some residues, like Ala-6 (Fig. 2), exhibit large ^{15}N relaxation dispersions at room temperature but not at elevated temperatures, whereas others such as Asp-16 (Fig. 2) show exactly the opposite behavior (9). In the context of a two-site exchange model this suggests that the excited state in exchange with the native (ground) state, F, is different at low and high temperatures. As expected, the ^{15}N chemical shifts reconstructed from the RD profiles for the most prominent excited state at elevated temperatures are very close to random coil values (45), so this state was readily identified as the unfolded state U, with a fractional population, $p_U = 3.0\%$ at 45°C (9). At 30°C its population drops to $p_U = 0.4\%$, and at 25°C and below p_U becomes negligible. At these temperatures the RD profiles are well fit by a simple two-state exchange model I↔F with a different excited state I than at higher temperatures, with amide ^{1}HN and ^{15}N chemical shifts that are not random coil like, but rather well-dispersed, suggesting that this state is folded, although not natively so (see below). The population of I, p_I, is 2.7% at 10°C, and decreases to 1.9% at 45°C. Exchange between F and I becomes fast at elevated temperatures ($k_{ex,FI} = 11081$/s at 45°C), but the value of p_I remains significant and a three-state model is necessary to fit the data at 30°C and above. It is worth noting that the information content of the RD profiles allows a discrimination between on- (U↔I↔F) and off- (I↔F↔U or F↔U↔I) pathway intermediates since exchange rates and chemical shift values are extracted simultaneously in fits of the data; the best fit was obtained for the model with I on-pathway (9). Thus, the equilibrium RD NMR experiments establish that the A39V/N53P/V55L Fyn SH3 domain folds under native conditions according to a three-state mechanism, U↔I↔F, which effectively reduces to I↔F at 25°C and below.

Comparison between NMR relaxation dispersion and stopped-flow kinetics data

The NMR data argue very strongly that the A39V/N53P/V55L Fyn SH3 domain folds via an on-pathway folding intermediate, yet no such intermediates have been observed to date in extensive studies of SH3 domains using kinetic methods (5,30). To examine this issue further we have measured the folding/unfolding kinetics of the A39V/N53P/V55L Fyn SH3 domain as a function of denaturant concentration by stopped-flow, and the resulting chevron plot was indeed fully consistent with the two-state folding model, U↔F (9). Prediction of the chevron plot on the basis of parameters obtained from the three-state folding model determined from the RD experiments established that the kinetics from both methods are quantitatively consistent. Although the presence of the intermediate state is predicted to slow down folding under native conditions by a very small amount compared to strictly two-state folding, the resulting ''roll-over'' in the chevron plot is too small for reliable detection even on stopped-flow devices with very short dead-times, especially since even small amounts of denaturant speed up the F↔I pre-equilibrium to the point where the transition from U to I becomes exclusively rate-limiting for folding (9). Thus, the stopped-flow fluorescence

measurements report only on the U↔I exchange event. All of the SH3 domains that we have studied to this point by RD NMR fold via similar on-pathway intermediates (9,31,32,37), which most likely cannot be detected by conventional stopped-flow measurements. This suggests that intermediates may be a more general feature of SH3 domain folding than anticipated based on experiment, in agreement with theoretical predictions (41,46). Under typical stopped-flow conditions the early transition state TS(UI) becomes rate-limiting. Thus, in the context of the three-state folding pathway U↔I↔F, the abundance of Φ-values that are available in the literature for SH3 domains obtained from stopped-flow kinetics analyzed assuming a two-state folding pathway should be interpreted as probing the early transition state TS(UI) (9).

The thermodynamics of the folding pathway

Fits of the three-state folding model U↔I↔F to RD profiles of the A39V/N53P/V55L Fyn SH3 domain generated well-defined values for all four kinetic rates (k_{FI}, k_{IF}, k_{IU}, k_{UI}) as well as populations of all three states ($p_I = k_{FI}/k_{IF} \times p_F$, $p_U = k_{IU}/k_{UI} \times p_I$, $p_F + p_I + p_U = 100\%$). The corresponding energy level diagram (Fig. 3), derived from Eqs. 1–3, shows the enthalpy/entropy compensation typical for protein folding (4). Both rate-limiting transition states TS(UI) and TS(IF) are entropically destabilized. Interestingly, the I state is more enthalpically favored and entropically disfavored than F, suggesting that I has a fully collapsed, rigid structure. This result is corroborated by studies over a wide temperature range (10–45°C), which allowed quantification of the heat capacity differences ΔC_p between I, TS(IF) and F from the curvature of the Eyring plots according to Eqs. 2 and 3. The heat capacity of I is surprisingly low and is identical to that of F within experimental error (9). Heat capacities are considered to reflect solvent-accessible surface areas (47) so that the similar values obtained for I and F argue that I likely has a fully collapsed, albeit nonnative structure (see below). Along these lines, one would expect that the nonnative hydrophobic core of I would be broken and partially rehydrated to allow formation of the native core, and indeed the transition state TS(IF) has a significantly higher heat capacity than either of the two states that it connects (9).

High-resolution structural information about the folding intermediate

As described above, the excited state spectral parameters that are directly accessible from RD experiments are the NMR chemical shifts, which in turn are well-known to reflect protein primary, secondary and, to a lesser extent, tertiary structure (45). These chemical shifts are routinely recast in terms of dihedral angle restraints in high-resolution structural studies of ground state structures of proteins (45,48,49) but they also provide valuable structural information about invisible excited states (28). Residual anisotropic interactions

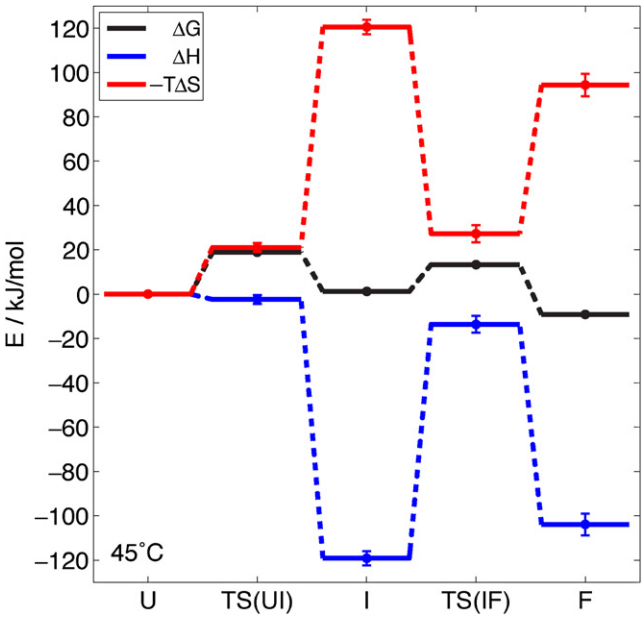

FIGURE 3 Profiles of free energy, ΔG (*black*), and the associated enthalpic ΔH (*blue*) and entropic $-T\Delta S$ (*red*) contributions to ΔG along the folding pathway for the A39V/N53P/V55L Fyn SH3 domain at 45°C. Values of the thermodynamic parameters are referenced with respect to the unfolded state U, which is arbitrarily assigned values of 0. TS(UI) and TS(IF) denote the rate-limiting transition states between states U, I and I, F respectively. Heat capacity values of $\Delta C_p(U \rightarrow I) = \Delta C_p(U \rightarrow F) - (0.15 \pm 0.17)$ kJ/mol/K and $\Delta C_p(U \rightarrow TS(IF)) = \Delta C_p(U \rightarrow F) + (1.46 \pm 0.11)$ kJ/mol/K are extracted from the temperature dependence of k_{FI} and k_{IF} according to Eqs. 2 and 3. Adapted from Neudecker et al. (9).

such as residual dipolar couplings also manifest in shifts of spectral lines and RD experiments have been developed for their measurement as well (28). This is a particularly exciting development because such interactions provide orientational restraints that have been shown to also be very useful for structural studies of excited conformers (50).

The chemical shifts of the folding intermediate of the A39V/N53P/V55L Fyn SH3 domain can be determined very accurately from fitting the two-state model I↔F to RD data recorded at low temperatures (9), bypassing the complexity of a three-state fitting procedure (31,51) required at higher temperatures. As seen in Fig. 4, changes in backbone amide ^{15}N chemical shifts between F and I states are highly localized to the terminal β-sheet (strands β_1 and β_5) and the 3_{10}-helical turn immediately preceding strand β_5, whereas the ^{15}N chemical shifts of the central β-sheet (strands β_2, β_3, β_4) differ much less between the two states. Backbone amide ^{1}HN and especially carbonyl ^{13}C shifts, the latter being less sensitive to side-chain effects (45), show very similar profiles (P. Neudecker, P. Lundström, and L. E. Kay, unpublished). The chemical shifts of the intermediate in the NH_2- and COOH-terminal regions deviate significantly from random coil values (45). For example, the amide group of Glu-5 shows strongly upfield ^{1}HN and downfield ^{15}N shift values in the intermediate for the A39V/N53P/V55L Fyn

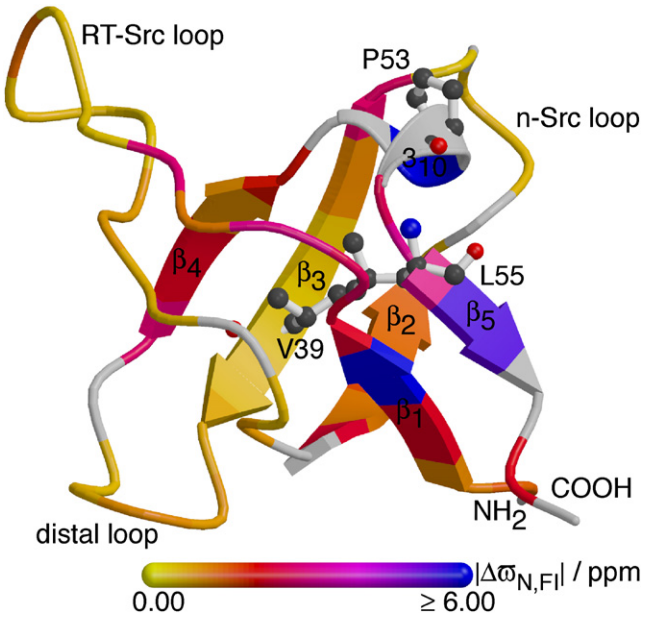

FIGURE 4 Schematic representation of the secondary structure of a homology model of the A39V/N53P/V55L *Gallus gallus* Fyn SH3 domain colored according to the backbone amide ^{15}N chemical shift differences between resonances in intermediate and folded states, $|\Delta\varpi_{N,FI}|$, from 0.0 ppm (*yellow*) to 6.0 ppm (*blue*), or gray if $\Delta\varpi_{N,FI}$ could not be determined due to missing resonance assignments, resonance overlap, or if the exchange contributions to transverse relaxation were of insufficient magnitude to quantify shift differences. The mutated residues Val-39, Pro-53, and Leu-55 are shown in ball-and-stick representation. Values of $|\Delta\varpi_{N,FI}| >$ 1.5 ppm (*red to blue*) are localized primarily to a region including the proximal NH_2- and COOH-termini comprising strands β_1 and β_5, indicating that the rest of the domain essentially retains a native-like backbone conformation in the intermediate state. Adapted from Neudecker et al. (9).

SH3 domain (9), but not for the G48M and G48V Fyn SH3 mutants that had been studied previously (31,52). Since the primary sequence of strand β_1 in all these mutants is identical the differences in shifts observed for the A39V/N53P/V55L Fyn SH3 domain must reflect nonlocal interactions. One explanation is that the additional hydrophobicity introduced by the mutation N53P (note that position 53 is a solvent-exposed side chain in the native state, Fig. 4) stabilizes a nonnative hydrophobic cluster, which involves residues from the NH_2- and COOH-termini, in the intermediate state. This interpretation is corroborated by computer simulations of the Fyn SH3 domain, where increasing the hydrophobicity at position 53 leads to a higher contact probability with the NH_2-terminal residues in the transition state ensemble (43). Favorable nonnative interactions in the transition state between a hydrophobic side chain introduced at position 53 and the side chains of Leu-3 and Phe-4 could indeed be demonstrated experimentally using double mutant thermodynamic cycles based on stopped-flow kinetics interpreted on the basis of a two-state model (43). Interestingly, these favorable nonnative interactions were found to speed up folding $\approx$2-fold in both the simulations and in the stopped-flow experiments (43). This nonnative hydrophobic cluster is not present in the G48M or G48V Fyn SH3 domains, whose termini appear to be much more flexible and partially hydrated (31). We are currently in the process of producing atomic models of the nonnative I state using a combination of extensive chemical shift (^{1}HN, ^{15}N, ^{13}CO, ^{1}Hα, ^{13}Cα) and residual anisotropic interactions that manifest upon weak alignment, measured through a variety of RD experiments (25–28,36). We have established that the cosolvents used for alignment preserve the kinetics and thermodynamics of the three-state folding pathway.

NMR-based Φ-value analysis of a three-state folding pathway

As described above analysis of RD data can provide detailed information about intermediate states, so long as they are populated at a level of $\approx$0.5% or more. Insight into the nature of TS ensembles is more difficult to obtain since they are of much higher energy and hence much lower population. Such information is available, however, from Φ-value analysis. Here the effects of mutations on folding/unfolding rates are monitored and recast in terms of changes in free energies that can, with care, be interpreted in terms of formation of structure in the TS at the site of mutation (4,6). Since RD experiments yield kinetic rates, and hence via Eqs. 1–3 the free energy profile that describes the folding event, they can be used as the kinetic method for Φ-value analysis in much the same way as conventional approaches (53). The principle of Φ-value analysis is illustrated in Fig. 5 *A*. A point mutation at a particular site in the protein will change the stability of the protein by an amount $\Delta\Delta G_{U\rightarrow F} = \Delta G_{U\rightarrow F}(\text{mutant}) - \Delta G_{U\rightarrow F}(\text{wt})$ and in a similar manner modulate the stability of the intermediate and transition states (A) relative to the unfolded state by $\Delta\Delta G_{U\rightarrow A} = \Delta G_{U\rightarrow A}(\text{mutant}) - \Delta G_{U\rightarrow A}(\text{wt})$. The Φ-value of state A is defined as

$$\Phi_A = \frac{\Delta\Delta G_{U\rightarrow A}}{\Delta\Delta G_{U\rightarrow F}}. \quad (4)$$

If the mutated side chain is in its native environment in state A the mutation will affect the stability of A in the same way as F, resulting in $\Phi_A \approx 1$. By contrast, if the mutated side chain is disordered it will affect A in the same way as U, so $\Delta\Delta G_{U\rightarrow A} \approx 0$ and $\Phi_A \approx 0$. Φ-values therefore reflect the formation of native contacts along the folding pathway.

In agreement with a large body of Φ-values for SH3 domains that have been interpreted on the basis of a two-state folding transition the Φ-values for the early transition state TS(UI) of the A39V/N53P/V55L Fyn SH3 domain derived from NMR RD experiments (53) are highly polarized (Fig. 5). The central β-sheet, probed by the mutations **R40T** and **T47S** (for clarity, mutations for Φ-value analysis are highlighted in boldface), is already formed in TS(UI) and is maintained throughout the folding pathway, as indicated by high Φ-values (Fig. 5 *B*). By contrast, the terminal

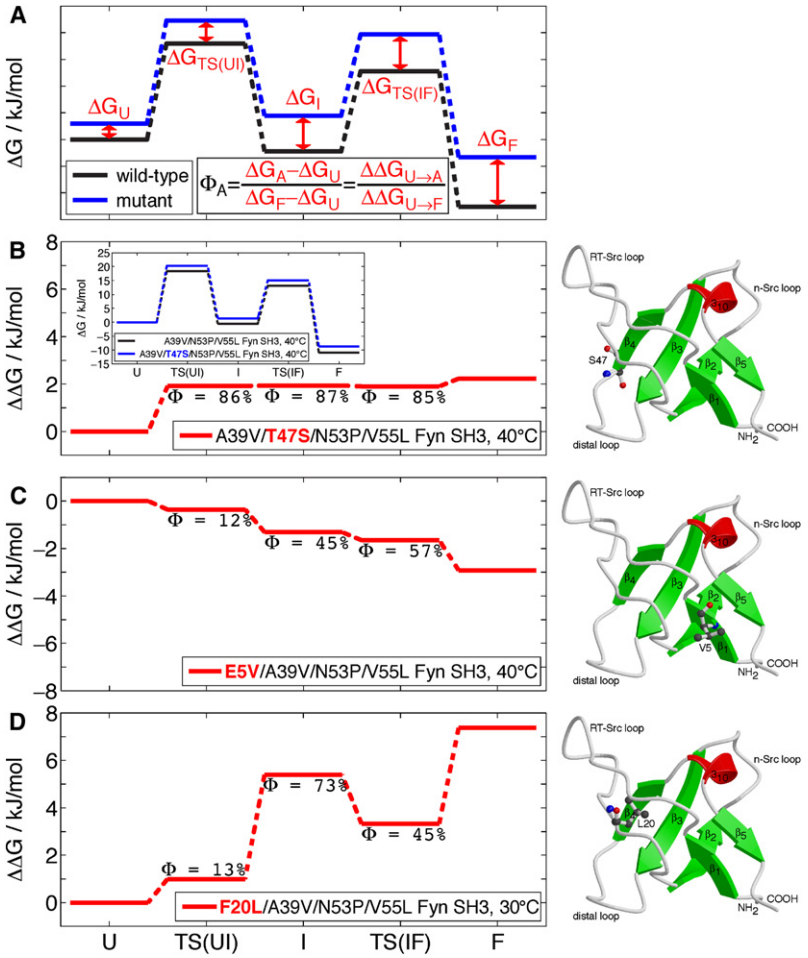

FIGURE 5 Changes in free energy for state A, $\Delta G_A = G_A(\text{mutant}) - G_A(\text{wt})$, used to calculate Φ_A (*A*), along with changes in free energy upon mutation, $\Delta\Delta G$, along the folding pathway for several mutants of the A39V/N53P/V55L Fyn SH3 domain (*B–D*). Values of *G* are referenced with respect to the unfolded state U, which is arbitrarily assigned a value of 0. TS(UI) and TS(IF) denote the rate-limiting transition states between states U, I, and I, F respectively. Side chains mutated for Φ-value analysis are indicated in boldface and highlighted in ball-and-stick representation on the right-hand side (same homology model and view as in Fig. 4). (*B*) Inset shows the pair of ΔG profiles ("pseudo-wt" and mutant) at 40°C from which $\Delta\Delta G$ values are obtained, with the free energies of U states both assigned arbitrarily to 0 (strictly true only if $\Delta G_U = 0$). Note that, unlike **T47S** (*B*) and **F20L** (*D*), **E5V** (*C*) is a stabilizing mutation ($\Delta\Delta G_{U\rightarrow F} < 0$), but this is irrelevant for Φ-value analysis. Adapted from Neudecker et al. (53).

β-sheet, probed by the mutations **L3A** and **E5V**, does not fully adopt its native conformation until after the late transition state TS(IF), as indicated by low Φ-values (Fig. 5 *C*).

Calculated Φ_I values also corroborate the picture of the intermediate state that emerges from analysis of chemical shifts. The largest changes in chemical shifts that accompany the hydrophobic core mutation **F20L** are in strand β_1, which does not make contact with Phe-20 in the folded state. Conversely, the **E5V** mutation perturbs the peak positions of Leu-18 and Ser-19 in the intermediate. These chemical shift perturbations provide strong evidence of nonnative hydrophobic contacts in the intermediate. The high Φ_I associated with **F20L** but significantly lower $\Phi_{TS(IF)}$ value likely reflects these nonnative contacts in the intermediate, which must be broken in the late transition state before formation of the native fold (Fig. 5 *D*).

A distinct advantage of RD NMR over other kinetic methods for Φ-value analysis is the availability of the chemical shifts of the ground and excited states as sensitive probes to monitor any (undesired) structural changes introduced by the mutation. In the case of the A39V/N53P/V55L Fyn SH3 domain the mutations produced only very small changes in chemical shifts, so that the Φ-values could be interpreted with confidence (53).

Comparison of relaxation dispersion NMR spectroscopy with other biophysical methods

The above discussion has focused on the use of RD NMR spectroscopy to probe the folding pathway of the A39V/N53P/V55L Fyn SH3 domain. The methodology employed is particularly sensitive to exchange events so long as they involve "excited" states populated at a level of 0.5% or higher that interconvert with a "visible" ground state on the millisecond timescale and there is a difference in chemical shifts between the exchanging states. Detailed kinetic, thermodynamic and structural data can be obtained and structures of intermediate states can in principle be determined irrespective of whether such states are native-like or not. A powerful aspect of the NMR technique is that experiments can be performed in the absence of denaturants that can affect structures of U or I states, and that it is often possible to establish whether the intermediate is on- or off-pathway, as has been the case in folding studies of a number of small domains (9,31,32,54).

RD methodology is complementary to other kinetic techniques such as stopped-flow mixing (4,10,11), temperature jump (11), as well as hydrogen exchange experiments

(14–16). With these other methods it is possible to detect and characterize excited states with free energies that far exceed those of the ''low-lying'' states that can be probed by RD NMR. The picture of protein folding that emerges from all of these techniques is similar. For example, stopped-flow derived Φ-values for SH3 domain folding (38–40) paint a picture of a TS ensemble with native-like structure in strands β_2 to β_4 and the n-Src loop that is similar to what emerges from analysis of chemical shifts of the intermediate state measured via RD NMR (9,31,32,37). RD NMR studies of an FF domain, a small four-helix bundle module, establish the presence of an intermediate state where helices 1 to 3 are at least partially formed, whereas helix 4 is not (54). Hydrogen exchange experiments on the folded state corroborate the fact that helix 4 is not formed in the intermediate state and the relative stabilities of helices 1–3 vs. 4 based on hydrogen exchange rates agrees with expectations from RD NMR (54). The NMR results are consistent with a Φ-value analysis of the rate-limiting transition state between I and F (55). Native state hydrogen exchange experiments have provided a wealth of data on protein folding intermediates and have established that for many proteins folding proceeds through cooperatively folding units serving as templates for further structure formation, sometimes called foldons (16). The folding intermediates of SH3 and FF domains that have been characterized by RD NMR studies (see above) also establish that formation of foldons is an important step in the folding process for these modules as well. In this regard the conclusions from RD NMR experiments support a growing list of experimental and theoretical results that many proteins, even relatively small ones, do not seem to fold along a multitude of equivalent pathways on a smooth funnel-like energy landscape, but rather by preferred stepwise pathways with specific, highly polarized transition states and intermediates, where parts of the protein have already adopted native-like conformations (4,8–11,14–16,37,53,54) and where missing favorable interactions are frequently substituted by nonnative contacts (8,9,15,16,43,53).

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K. P.N. and P.L. hold postdoctoral fellowships from the CIHR and the CIHR Training Grant on Protein Folding in Health and Disease, respectively.

REFERENCES

1. Anfinsen, B. C. 1973. Principles that govern the folding of protein chains. *Science*. 181:223–230.
2. Jaenicke, R. 1991. Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*. 30:3147–3161.
3. Jackson, S. E., and A. R. Fersht. 1991. Folding of Chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*. 30:10428–10435.
4. Fersht, A. R. 1998. Structure and Mechanism in Protein Science. Freeman, New York.
5. Capaldi, A. P., and S. E. Radford. 1998. Kinetic studies of β-sheet protein folding. *Curr. Opin. Struct. Biol.* 8:86–92.
6. Fersht, A. R., A. Matouschek, and L. Serrano. 1992. The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224:771–782.
7. Sosnick, T. R., B. A. Krantz, R. S. Dothager, and M. Baxa. 2006. Characterizing the protein folding transition state using Ψ analysis. *Chem. Rev.* 106:1862–1876.
8. Brockwell, D. J., and S. E. Radford. 2007. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17:30–37.
9. Neudecker, P., A. Zarrine-Afsar, W. -Y. Choy, D. R. Muhandiram, A. R. Davidson, et al. 2006. Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* 363:958–976.
10. Roder, H., K. Maki, and H. Cheng. 2006. Early events in protein folding explored by rapid mixing methods. *Chem. Rev.* 106:1836–1861.
11. Nölting, B. 2006. Protein Folding Kinetics, 2nd ed. Springer, Berlin.
12. Khan, F., J. I. Chuang, S. Gianni, and A. R. Fersht. 2003. The kinetic pathway of folding of barnase. *J. Mol. Biol.* 333:169–186.
13. Nölting, B., R. Golbik, J. L. Neira, A. S. Soler-Gonzalez, G. Schreiber, et al. 1997. The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. USA*. 94:826–830.
14. Krishna, M. M. G., L. Hoang, Y. Lin, and S. W. Englander. 2004. Hydrogen exchange methods to study protein folding. *Methods*. 34:51–64.
15. Bai, Y. 2006. Protein folding pathways studied by pulsed- and native-state hydrogen exchange. *Chem. Rev.* 106:1757–1768.
16. Englander, S. W., L. Mayne, and M. M. G. Krishna. 2007. Protein folding and misfolding: mechanism and principles. *Q. Rev. Biophys.* 40:287–326.
17. Zeeb, M., and J. Balbach. 2004. Protein folding studied by real-time NMR spectroscopy. *Methods*. 34:65–74.
18. Schanda, P., V. Forge, and B. Brutscher. 2007. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*. 104:11257–11262.
19. Feng, H., J. Takei, R. Lipsitz, N. Tjandra, and Y. Bai. 2003. Specific non-native hydrophobic interactions in a hidden folding intermediate: implications for protein folding. *Biochemistry*. 42:12461–12465.
20. Religa, T. L., J. S. Markson, U. Mayor, S. M. V. Freund, and A. R. Fersht. 2005. Solution structure of a protein denatured state and folding intermediate. *Nature*. 437:1053–1056.
21. Kato, H., H. Feng, and Y. Bai. 2007. The folding pathway of T4 lysozyme: The high-resolution structure and folding of a hidden intermediate. *J. Mol. Biol.* 365:870–880.
22. Zhou, Z., H. Feng, R. Ghirlando, and Y. Bai. 2008. The high-resolution NMR structure of the early folding intermediate of the *Thermus thermophilus* ribonuclease H. *J. Mol. Biol.* 384:531–539.
23. Deverell, C., R. E. Morgan, and J. H. Strange. 1970. Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Mol. Phys.* 18:553–559.
24. McConnell, H. M. 1958. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* 28:430–431.
25. Palmer, A. G., C. D. Kroenke, and J. P. Loria. 2001. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* 339:204–238.
26. Palmer, A. G. 2004. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* 104:3623–3640.
27. Palmer, A. G., M. J. Grey, and C. Wang. 2005. Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. *Methods Enzymol.* 394:430–465.
28. Hansen, D. F., P. Vallurupalli, and L. E. Kay. 2008. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR*. 41:113–120.
29. Noble, M. E. M., A. Musacchio, M. Saraste, S. A. Courtneidge, and R. K. Wierenga. 1993. Crystal structure of the SH3 domain in human

Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *EMBO J.* 12:2617–2624.

30. Plaxco, K. W., J. I. Guijarro, C. J. Morton, M. Pitkeathly, I. D. Campbell, and C. M. Dobson. 1998. The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry.* 37:2529–2537.
31. Korzhnev, D. M., X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, et al. 2004. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature.* 430:586–590.
32. Korzhnev, D. M., P. Neudecker, A. Zarrine-Afsar, A. R. Davidson, and L. E. Kay. 2006. Abp1p and Fyn SH3 domains fold through similar low-populated intermediate states. *Biochemistry.* 45:10175–10183.
33. Eyring, H. 1935. The activated complex and the absolute rate of chemical reactions. *Chem. Rev.* 17:65–77.
34. Carr, H. Y., and E. M. Purcell. 1954. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* 94:630–638.
35. Meiboom, S., and D. Gill. 1958. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* 29:688–691.
36. Lundström, P., D. F. Hansen, P. Vallurupalli, and L. E. Kay. 2009. Accurate measurement of Alpha proton chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy. *J. Am. Chem. Soc.* 131:1915–1926.
37. Korzhnev, D. M., and L. E. Kay. 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.
38. Martinez, J. C., and L. Serrano. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6:1010–1016.
39. Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski, et al. 1999. Experiment and theory highlight role of native state topology in SH3 domain folding. *Nat. Struct. Biol.* 6:1016–1024.
40. Northey, J. G. B., A. A. Di Nardo, and A. R. Davidson. 2002. Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* 9:126–130.
41. Guo, W., S. Lampoudi, and J. -E. Shea. 2003. Posttransition state desolvation of the hydrophobic core of the src-SH3 protein domain. *Biophys. J.* 85:61–69.
42. Taketomi, H., Y. Ueda, and N. Gō. 1975. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7:445–459.
43. Zarrine-Afsar, A., S. Wallin, A. M. Neculai, P. Neudecker, P. L. Howell, et al. 2008. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. USA.* 105:9999–10004.
44. Canet, D., A. M. Last, P. Tito, M. Sunde, A. Spencer, et al. 2002. Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Biol.* 9:308–315.
45. Wishart, D. S., and D. A. Case. 2001. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* 338:3–34.
46. Ollerenshaw, J. E., H. Kaya, H. S. Chan, and L. E. Kay. 2004. Sparsely populated folding intermediates of the Fyn SH3 domain: Matching native-centric essential dynamics and experiment. *Proc. Natl. Acad. Sci. USA.* 101:14748–14753.
47. Myers, J. K., C. N. Pace, and J. M. Scholtz. 1995. Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* 4:2138–2148.
48. Cavalli, A., X. Salvatella, C. M. Dobson, and M. Vendruscolo. 2007. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. USA.* 104:9615–9620.
49. Shen, Y., O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, et al. 2008. Consistent blind protein structure generation from NMR chemical shifts data. *Proc. Natl. Acad. Sci. USA.* 105:4685–4690.
50. Vallurupalli, P., D. F. Hansen, and L. E. Kay. 2008. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. USA.* 105:11766–11771.
51. Neudecker, P., D. M. Korzhnev, and L. E. Kay. 2006. Assessment of the effects of increased relaxation dispersion data on the extraction of 3-site exchange parameters characterizing the unfolding of an SH3 domain. *J. Biomol. NMR.* 34:129–135.
52. Korzhnev, D. M., P. Neudecker, A. Mittermaier, V. Y. Orekhov, and L. E. Kay. 2005. Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: An application to the folding of a Fyn SH3 domain mutant. *J. Am. Chem. Soc.* 127:15602–15611.
53. Neudecker, P., A. Zarrine-Afsar, A. R. Davidson, and L. E. Kay. 2007. Φ-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. USA.* 104:15717–15722.
54. Korzhnev, D. M., T. L. Religa, P. Lundström, A. R. Fersht, and L. E. Kay. 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 13C-methyl relaxation dispersion and $^{1}H/^{2}H$-exchange NMR spectroscopy. *J. Mol. Biol.* 372:497–512.
55. Jemth, P., R. Day, S. Gianni, F. Khan, M. Allen, 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.