Article

Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding

Dmitry M. Korzhnev, and Lewis E. Kay

Acc. Chem. Res., **2008**, 41 (3), 442-451 • DOI: 10.1021/ar700189y • Publication Date (Web): 15 February 2008

ACCOUNTS
of chemical research

Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding

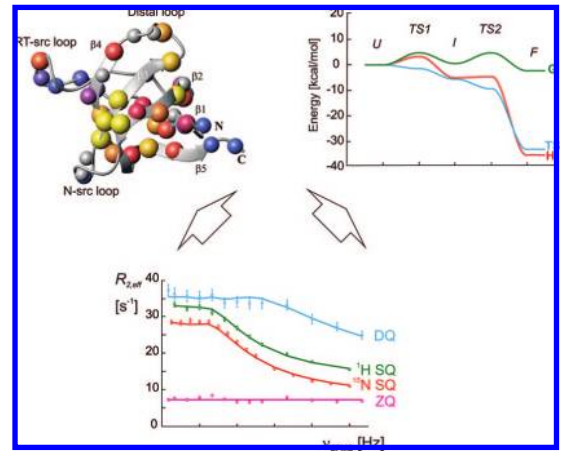
DMITRY M. KORZHNEV AND LEWIS E. KAY*

Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

RECEIVED ON AUGUST 23, 2007

CONSPECTUS

Biological function depends on molecular dynamics that lead to excursions from highly populated ground states to much less populated excited states. The low populations and the transient formation of such excited states render them invisible to the conventional methods of structural biology. Thus, while detailed pictures of ground-state structures of biomolecules have emerged over the years, largely through X-ray diffraction and solution nuclear magnetic resonance (NMR) spectroscopy studies, much less structural data has been accumulated on the conformational properties of the invisible excited states that are necessary to fully explain function. NMR spectroscopy is a powerful tool for studying conformational dynamics because it is sensitive to dynamics over a wide range of time scales, extending from picoseconds to seconds and because information is, in principle, available at nearly every position in the molecule. Here an NMR method for quantifying millisecond time scale dynamics that involve transitions between different molecular conformations is described. The basic experimental approach, termed relaxation dispersion NMR spectroscopy, is outlined to provide the reader with an intuitive feel for the technology. A variety of different experiments that probe conformational exchange at different sites in proteins are described, including a brief summary of data-fitting procedures to extract both the kinetic and thermodynamic properties of the exchange process and the structural features of the invisible excited states along the exchange pathway. It is shown that the methodology facilitates detection of intermediates and other excited states that are populated at low levels, 0.5% or higher, that cannot be observed directly in spectra, so long as they exchange with the observable ground state of the protein on the millisecond time scale. The power of the methodology is illustrated by a detailed application to the study of protein folding of the small modular SH3 domain. The kinetics and thermodynamics that describe the folding of this domain have been characterized through the effects of temperature, pressure, side-chain deuteration, and mutation, and the structural features of a low-populated folding intermediate have been assessed. Despite the fact that many previous studies have shown that SH3 domains fold via a two-state mechanism, the NMR methods presented unequivocally establish the presence of an on-pathway folding intermediate. The unique capabilities of NMR relaxation dispersion follow from the fact that large numbers of residues can be probed individually in a single experiment. By contrast, many other forms of spectroscopy monitor properties that are averaged over all residues in the molecule or that make use of only one or two reporters. The NMR methodology is not limited to protein folding, and applications to enzymatic catalysis, binding, and molecular recognition are beginning to emerge.



Introduction

It is becoming increasingly clear that a detailed understanding of molecular function requires not only high-resolution structural information of the sort that can be obtained through X-ray diffraction or NMR studies but also a quantitative characterization of how structure changes with time. In

Published on the Web 02/15/2008 www.pubs.acs.org/acr
10.1021/ar700189y CCC: $40.75

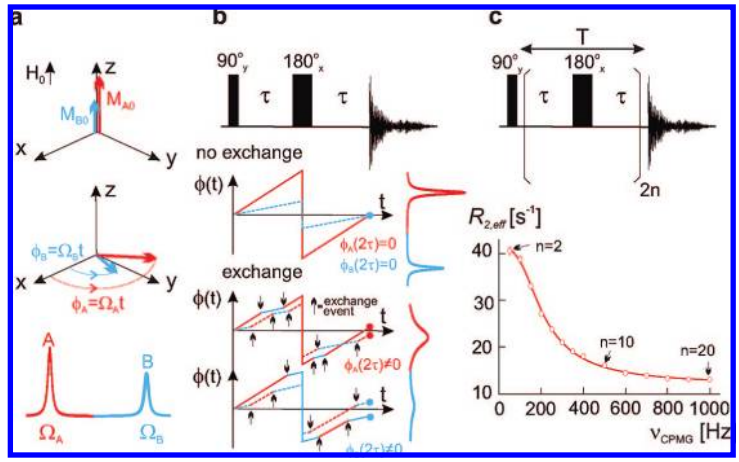


FIGURE 1. The physical basis underlying CPMG relaxation dispersion NMR spectroscopy, as described in the text.

the case of proteins, in particular, dynamics have been implicated in many aspects of function, including molecular recognition and signaling,[1] folding,[2,3] enzymatic catalysis,[4,5] and allostery.[6] Many of the structural changes involve conversion to low-populated and transient conformers that are extremely difficult to study using conventional biophysical tools. Herein we describe a solution-based NMR technique that can provide atomic resolution studies of millisecond time scale conformational transitions in proteins, termed Carr−Purcell−Meiboom−Gill (CPMG) relaxation dispersion NMR.[7] Advances in the development of ^{1}H, ^{15}N, and ^{13}C NMR relaxation dispersion experiments that probe millisecond conformational exchange at nearly every site in suitably labeled proteins have recently allowed detailed *quantitative* studies of processes involving *multiple low-populated states*.[2,5,8,9] In cases where exchange is between a ground state and "excited" states that are populated at levels of 0.5% or higher with rates of exchange on the order of a few hundred to several thousand per second, the relaxation data measured for multiple nuclei often allow complete *kinetic* and *thermodynamic* characterization of the exchange process, even though the resonances of the "excited" states cannot be directly observed in NMR spectra. Additionally, relaxation dispersion data report NMR chemical shift differences between states, providing *structural* information on the low-populated "excited" species.[2]

In what follows, we review CPMG relaxation dispersion methods for studies of millisecond time scale dynamics in proteins that have been developed over the past few years.[10−17] The basic principles behind such experiments are described, along with a brief survey of the different approaches that have become available, allowing measurement of dynamics at different positions in proteins. Subsequently, an application of the methodology to the characterization of the folding pathway of the SH3 domain,[1] which has served as a model system in many laboratories, is described. Studies of folding/unfolding as a function of temperature,[2,18,19] pressure,[20] mutation,[2,21,22] and levels of protein deuteration[23] have provided a detailed picture of the folding pathway, along with a description of an "invisible" intermediate that is formed during the folding process. The work provides an example of the power of relaxation dispersion for the study of "excited" states that can include but are not restricted to those formed during protein folding.

CPMG-Based NMR Relaxation Dispersion Methods

The Experimental Approach. A basic understanding of this class of experiment builds upon the idea of a "spin−echo" that was first described by Erwin Hahn in 1950 and has since become important in much of NMR and indeed in other spectroscopies as well. Although a complete description of many NMR experiments requires quantum mechanics, much insight can be obtained from a classical picture in which the macroscopic nuclear magnetization at equilibrium is represented by a vector that aligns along an axis (Z) that is collinear with the magnetic field, $\mathbf{H}_0$, Figure 1a. Application of a radio frequency pulse at the appropriate frequency rotates the magnetization vector away from the Z-axis, toward the $X-Y$ plane, and if a 90°_y pulse is applied, the magnetization is placed along the X-axis. Subsequently, magnetization rotates about the Z-axis with a characteristic angular frequency Ω; in the example considered in Figure 1a, there are two such components (red and blue), each with discrete resonance frequencies. Over time each magnetization component will accrue a phase, φ, with respect to the X-axis, given by $\varphi = \Omega t$. In the basic spin−echo experiment, Figure 1b, a $\tau-180^\circ_x-\tau$ block is added after the 90°_y pulse, followed by signal acquisition. During the first τ element, magnetization in each of the red and blue states

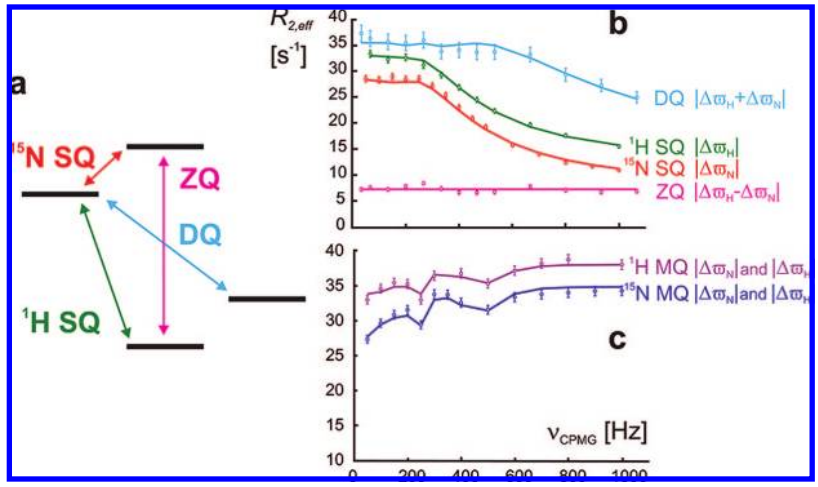


FIGURE 2. (a) Energy-level diagram for a two-spin ^{1}H−^{15}N spin system showing four transitions that can be used to probe microsecond to millisecond exchange events along with experimental data (b, c) from the amide group of Glu11 measured in an ^{15}N/^{2}H-labeled sample of the G48M Fyn SH3 domain[27] (shown with the same colors as the corresponding transitions). Each relaxation dispersion profile is annotated with the chemical shift difference(s) to which it is sensitive.

acquires phase $\varphi_A = \Omega_A\tau$ and $\varphi_B = \Omega_B\tau$, respectively, so that a plot of φ vs t is linear, with slope Ω (Figure 1b, middle). At $t = \tau$, a "refocusing" 180°_x pulse is applied that inverts the phase of the magnetization and the phase then "builds-up" again during the second τ period so that the net phase is zero at $t = 2\tau$, giving rise to an ordering of magnetization along the X-axis referred to as a spin−echo. Now suppose that the red and blue magnetization components derive from a single NMR active probe (a ^{1}H, ^{13}C, or ^{15}N nucleus, for example) on a molecule that exchanges stochastically between two distinct conformations. If the exchange of conformations leads to a difference in the magnetic environment of the probe, corresponding to resonance frequencies Ω_A, Ω_B, then a plot of φ vs t might look like one of those shown in Figure 1b, bottom, where because each magnetization component generally acquires a different phase before and after the refocusing pulse, $\varphi(2\tau) \neq 0$. Different $\varphi(t)$ profiles would be observed for different molecules since the process of exchange is random, and because the net signal is given by the sum of $\varphi(2\tau)$ over all molecules, the intensity would be reduced relative to the case of no exchange.

CPMG-type dispersion experiments exploit multiple-echo refocusing pulse sequences with the prototype experiment for studies of an isolated single spin consisting of a 90°_y pulse followed by an even number of repeats of the $\tau-180^\circ_x-\tau$ block (Figure 1c). The frequency of application of 180°_x pulses in the CPMG sequence, $1/(4\tau)$, is called the CPMG frequency, ν_{CPMG}. Qualitatively, if the rate of molecular exchange is slow compared with ν_{CPMG}, magnetization is nearly completely restored along the X-axis after each block, so the amplitude of consecutive spin−echoes is little affected by the exchange process. In contrast, as the exchange events become on the order of ν_{CPMG}, the CPMG sequence becomes less effective at refocusing leading to a decay of the spin−echo amplitude. In general, *relaxation dispersion* is the dependence of the rate of magnetization decay during the CPMG sequence, $R_{2,eff}$, on ν_{CPMG}. In practice, $R_{2,eff}$ rates are quantified from peak intensities in NMR spectra recorded with different numbers of repeats of the $\tau-180^\circ_x-\tau$ CPMG block, as described previously.[17] In most cases, the molecule of interest exists in one highly populated ground state, occasionally visiting a number of exited minor states, and only the resonances of the ground state are observed.

NMR relaxation dispersion profiles $R_{2,eff}(\nu_{CPMG})$ provide sensitive measures of conformational/chemical exchange processes allowing extraction of populations of exchanging states (p_n), rates of transitions between states (k_{mn}), and absolute values of frequency differences between states $|\Delta\omega_{mn}| = |\omega_m - \omega_n|$ (or, equivalently, chemical shift differences in ppm $|\Delta\tilde{\omega}_{mn}|$) for each pair of exchanging states m and n. All of the parameters mentioned above can be obtained by a least-squares fit of experimental relaxation dispersion data (i) to approximate expressions (which are available for two-state exchange only) or (ii) by solving the Bloch−McConnell equations numerically for a given chemical exchange model.[7] In most cases, exchange contributions to $R_{2,eff}$ decrease with increasing ν_{CPMG} (see Figure 1). In the absence of exchange or in the case of $\Delta\tilde{\omega}_{mn} = 0$, $R_{2,eff}$ rates are independent of ν_{CPMG}, so that the measured relaxation dispersion profiles are flat.

Relaxation Dispersion Experiments for Backbone Amide Groups and Beyond. New strategies for biosynthetic labeling of proteins with NMR active isotopes and concomi-

tant multinuclear, multidimensional NMR techniques have facilitated the investigation of conformational exchange in many sites in proteins using ^{15}N, ^{13}C, and ^{1}H nuclei as probes. Although based on the original spin−echo and CPMG approaches that were designed for studies involving *isolated single spins* (Figure 1), most NMR experiments for protein applications are more complicated due to manipulations of magnetization derived from heteronuclear *multiple-spin systems*. From the complexity emerges certain advantages, however, since it is possible to design experiments that exploit different NMR transitions (qualitatively, different types of signal) to probe conformational exchange. For example, Figure 2a shows an energy-level diagram for the backbone $^{1}H-^{15}N$ spin pair, highlighting with arrows the transitions that can be exploited as sensitive probes of microsecond to millisecond exchange. Relaxation dispersions profiles that are based on these different transitions,[10−12] Figure 2b, or linear combinations of transitions,[13] Figure 2c, are described by the same functional dependence on the exchange parameters[13] but have different shapes because they are sensitive to either proton or nitrogen chemical shift differences between states, $\Delta\tilde{\omega}_{H,mn}$ or $\Delta\tilde{\omega}_{N,mn}$, respectively, or both.

CPMG-type relaxation dispersion experiments that exploit different spin coherences in multiple-spin systems have also been designed for a number of other protein groups. For example, ^{13}C CPMG experiments were proposed for quantification of exchange using $^{13}C^{\alpha}$ (ref 3) and carbonyl[14] nuclei as well as methyl groups,[15,16] and ^{15}N schemes have been proposed for measurements involving side-chain $^{15}NH_2$ moieties.[17] In contrast to relaxation dispersion experiments for the study of backbone amide groups that can be measured in uniformly ^{15}N-labeled protein samples, experiments that exploit ^{13}C as a probe generally require more elaborate isotope labeling, as described in the primary literature.[24−26]

Model Selection. Extraction of kinetic and thermodynamic parameters describing conformational exchange events, including chemical shift changes between the exchanging states, is based on fits of $R_{2,eff}(\nu_{CPMG})$ to theoretical models. However, often the underlying process giving rise to the observed exchange event is not known *a priori*, and hence the appropriate exchange model is unclear. In many cases, the "appropriateness" of a given model is based on goodness of fit criteria; nevertheless, high-quality data fits are not proof of the validity of a given model. The majority of CPMG-type dispersion studies of proteins published to date have involved data analysis under the assumption of two-state exchange, even though the underlying dynamics may be more complex.

Typically, relaxation dispersion profiles from a single probe (for example, a given ^{15}N or ^{13}C nucleus) are not sufficiently sensitive for discrimination between two-state and more complex exchange models,[2,27] unless the exchange process is described by kinetic constants that differ by an order of magnitude or more.[8] Characterization of complex multisite dynamics is possible, at least in some cases, from a combined analysis of relaxation dispersion data involving multiple probes within the context of a *global exchange model* (all data fit to a single model to extract kinetic parameters describing the exchange process). In many cases, it is also useful to globally fit the data recorded over a range of experimental conditions, such as temperature,[2,18,19] pressure,[20] or substrate concentration (for binding reactions).[9] In temperature- or pressure-dependent studies, the robustness of data fitting can be improved by assuming that (i) the temperature or pressure dependence of exchange rate constants follows transition-state theory and (ii) the chemical shift differences between states are independent of temperature or pressure.[7]

Studying Protein Folding by Relaxation Dispersion NMR

Why NMR? In the past few years CPMG-type relaxation dispersion methods have emerged as an ideal tool for studies of folding/unfolding transitions in marginally stable wild-type proteins and destabilized protein mutants that fold on the millisecond time scale, so long as the "excited" states that participate in the folding reaction, unfolded (U) and intermediate (I) states, are 1−3 kcal/mol higher in free energy than the "ground" folded state (F). Under these conditions, the fractional populations of U and I are in the range of 0.5−10%. Typically, most ^{1}H, ^{15}N, and ^{13}C nuclei exhibit substantial chemical shift changes upon protein folding that give rise to large dispersion curves, thus providing multiple independent NMR probes of the folding process. Since the folding event involves all residues in a protein, relaxation dispersion profiles for all nuclei and spin coherences measured in different CPMG-type experiments report on the same exchange process and, therefore, can be analyzed together within the frame of global exchange models. Combined analysis of extensive data sets measured at several magnetic fields or under multiple sets of conditions involving different temperatures or pressures or both, for example, allows effective discrimination between two-state and more complex folding models.[2,18−20,23,27] Because chemical shift differences are extracted simultaneously with rates during the fitting process and because chemical shifts provide unique spectroscopic signatures of each of the exchanging states, it is possible to establish

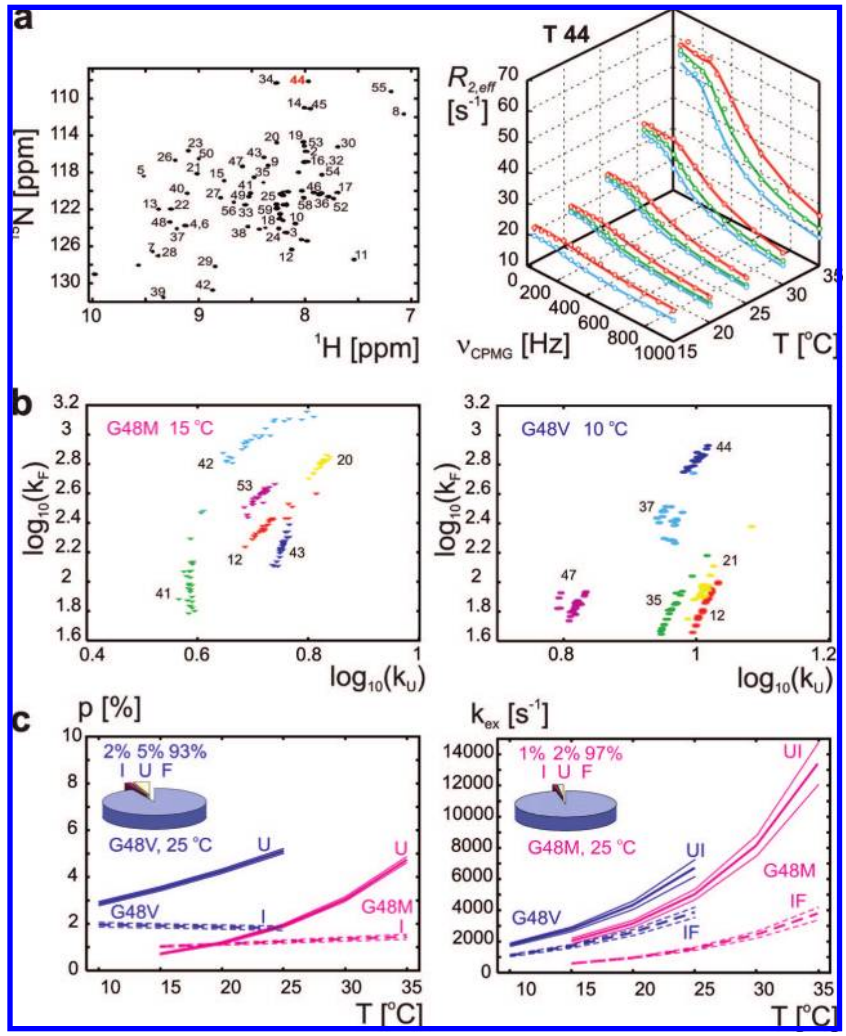


FIGURE 3. (a) ^{1}H−^{15}N correlation map of ^{15}N-labeled, protonated G48M Fyn SH3, 25 °C, along with typical ^{15}N relaxation dispersion profiles as a function of temperature and magnetic field strength (18.8 T, red; 14.1 T, green; 11.7 T, blue). Shown with solid lines are best fits to a global three-state exchange model. (b) Distributions of the apparent two-state folding and unfolding rates, k_f and k_u, for individual amide groups of G48M and G48V mutants of the Fyn SH3 domain obtained as described previously.[2] (c) Temperature dependencies of populations of I and U states, p_I and p_U, and rates of interconversion between states, $k_{ex,IF}$ (IF) and $k_{ex,UI}$ (UI) for the G48M and G48V Fyn SH3 domains from global fits of ^{15}N relaxation dispersion data to a three-state exchange model, U ↔ I ↔ F. Portions of figure from Korzhnev et al.[2]

whether an intermediate is on- or off-pathway, in the case of a three-state folding process, even for I states that are populated at levels as low as 0.5%. This is much more difficult to ascertain using other spectroscopic methods where both rates and amplitudes of changes must be measured and where spectroscopic profiles of each of the states must be known (fluorescence yields, extinction coefficients, etc). A further advantage is that the NMR experiment can be performed in aqueous solution, unlike other methods that require quantification of exchange rates as a function of denaturant.

Folding of SH3 Domains. SH3 domains[1] are small (about 60 residue) protein modules that fold into five-stranded β-sandwiches composed of two orthogonal β-sheets. They have been studied in detail using numerous equilibrium and kinetic biophysical techniques,[28] showing that they fold reversibly in a two-state manner. For example, stopped-flow fluorescence studies establish that the wild-type Fyn SH3 domain folds with a rate of 80 per second and with a stability of 4.2 kcal/mol.[21] A series of mutations at position Gly48 have been shown to both destabilize the protein and accelerate its folding rate by approximately 10-fold.[21] These mutants have unfolding free-energies (ΔG_{UF}) of 1−3 kcal/mol and folding rates of up to 1000 per second, well within the window that is amenable for study by NMR relaxation dispersion methods. An initial comparative analysis of seven Gly48 mutants that were examined by both NMR and stopped-flow methods showed that ΔG_{UF} values measured by the two approaches differed by <0.15 kcal/mol on aver-

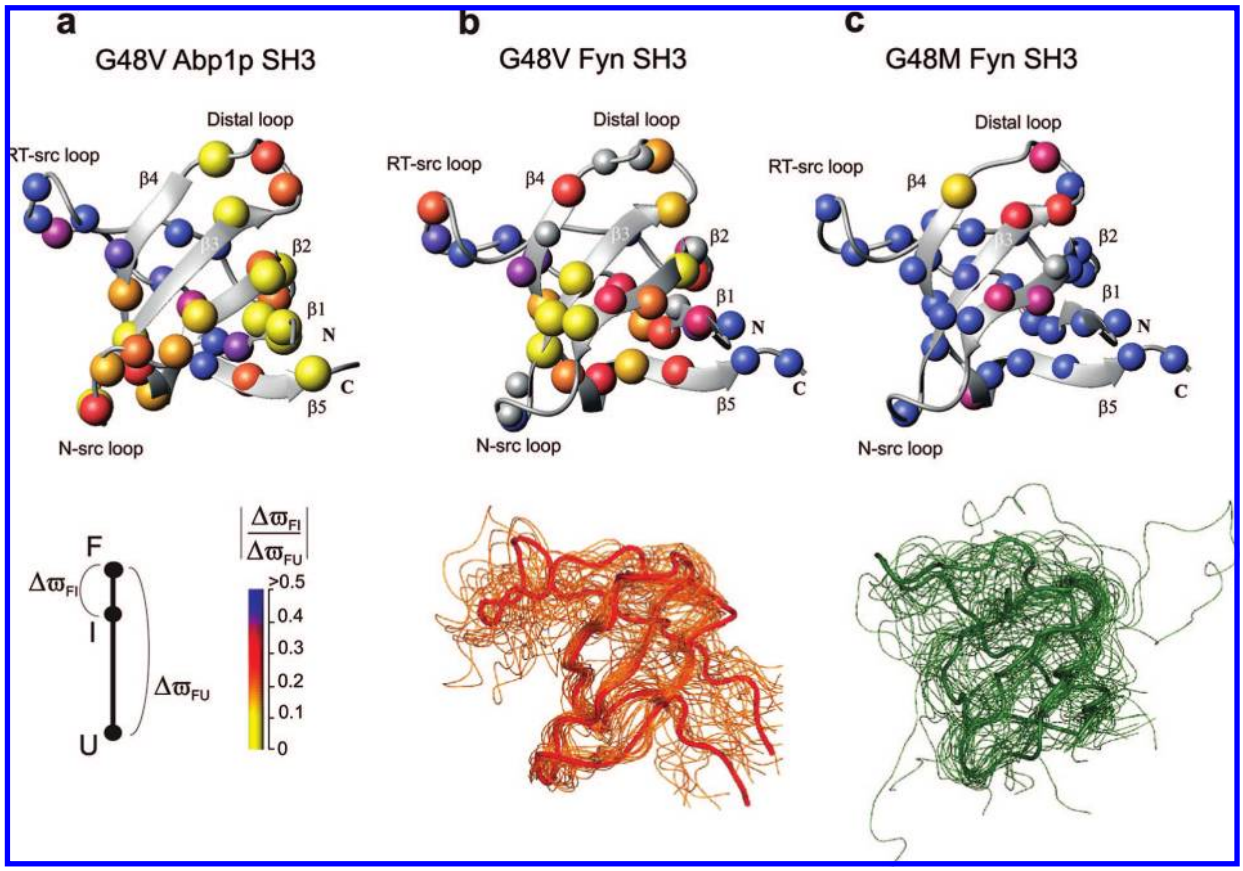


FIGURE 4. Ratios of ^{15}N chemical shift differences $|\Delta| = |\Delta\tilde{\omega}_{FI}/\Delta\tilde{\omega}_{FU}|$ for (a) G48V Abp1p SH3,[18] (b) G48V Fyn SH3, and (c) G48M Fyn SH3 domains[2] color coded on structures of the wild-type proteins, where each sphere in the diagram is a backbone amide nitrogen position (top plots). Bottom plots in panels b and c show ribbon representations of 25 structures that model the ensembles of the I states of G48V and G48M Fyn SH3 calculated based on experimental Δ ratios. Modified from Korzhnev et al.[2,18]

age,[21] establishing the accuracy of the NMR technique. However, as we show below, the NMR method can provide insight beyond that available using conventional methodologies.

We have analyzed a pair of Fyn SH3 domain mutants, G48M and G48V.[2] Figure 3a shows the $^{1}H-^{15}N$ correlation map of the G48M domain where each correlation corresponds to a backbone amide site. Notably, only correlations are observed from the folded state of the protein; however "invisible" excited states are nonetheless present since large ^{15}N relaxation dispersion profiles are observed for over 40 residues in data sets recorded at three magnetic fields and a range of temperatures, as shown for Thr44 of the G48M Fyn SH3 domain. Dispersion profiles like those in Figure 3a were initially fit on a per-residue basis to a model of two-state exchange; if such a model is appropriate, then very similar k_f (k_u) values would be obtained for each residue. That this is not the case is illustrated in Figure 3b, where k_f rates differ by as much as an order of magnitude for residues in both G48M and G48V. Not surprisingly, unsatisfactory fits were obtained from a global analysis of temperature-dependent CPMG dispersion data from all residues using a two-state model. By contrast, a global three-state model,

$$U \underset{k_{IU}}{\overset{k_{UI}}{\rightleftharpoons}} I \underset{k_{FI}}{\overset{k_{IF}}{\rightleftharpoons}} F$$

involving the formation of an on-pathway intermediate state, I, fit all of the data well (solid lines in dispersion profiles, Figure 3a), with the chemical shifts extracted for the U state agreeing with random-coil values, establishing that the exchange event is indeed a folding/unfolding equilibrium. Figure 3c shows the populations of each state as a function of temperature, along with the temperature dependence of k_{ex} ($UI = k_{UI} + k_{IU}$, $IF = k_{IF} + k_{FI}$) for the two mutants.

The observation via relaxation dispersion NMR of a folding intermediate that was not detected using other more established approaches illustrates the power of the dispersion methodology and the important role that NMR can play in studies of protein folding. This fact notwithstanding one could argue that the observed I state is of limited interest because it has been detected in mutants involving a highly conserved residue. To address this issue and to establish that

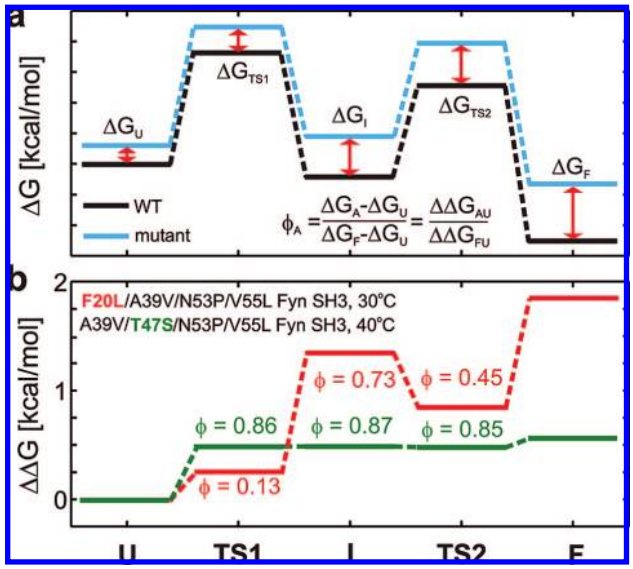


FIGURE 5. (a) Calculation of φ values for intermediate and transition states along a three-state folding pathway, U ↔ I ↔ F, from changes in free energies upon mutation, ΔG_A (here A denotes one of F, TS1, I, TS2, or U). (b) Changes in free-energy differences $\Delta\Delta G_{AU}=\Delta G_A - \Delta G_U$ for various states along a three-state folding pathway of A39V/N53P/V55L Fyn SH3 caused by T47S (green) and F20L (red) mutations, along with $\varphi_A = \Delta\Delta G_{AU}/\Delta\Delta G_{FU}$ values for corresponding intermediate and transition states.

the I state folding intermediate is not unique to the Fyn SH3 domain, we performed additional temperature-dependent ^{15}N studies on a pair of SH3 domains,[18,19] including (i) an A39V/N53P/V55L triple mutant of the Fyn SH3 module where the conserved Gly48 residue is retained and (ii) a fast-folding G48V mutant of a stabilized form of the Abp1p SH3 domain where 64% of the sequence is distinct from that of the Fyn SH3 module. Relaxation dispersion experiments indicate conclusively that both of these systems fold through an on-pathway intermediate populated at a level of 1−2% at 40 °C, supporting the notion that the U ↔ I ↔ F pathway is a conserved feature of SH3 domain folding.

Structures of Low-Lying Excited States from Chemical Shifts. Chemical shifts are very sensitive probes of structure, and there are exciting recent developments in the use of such shifts in *de novo* structure calculations of folded proteins.[29] Chemical shifts are also useful in providing a qualitative description of structure formation in the excited states, U and I, that are obtained from fits of dispersion data. In one approach, chemical shifts in the intermediate I are referenced to those in states F and U by calculating $\Delta = \Delta\tilde{\omega}_{FI}/\Delta\tilde{\omega}_{FU}$. If a set of residues that are either close in primary sequence or proximal in the native state have similar chemical shifts in I and F ($\Delta = 0$), then it is highly probably that they have a native-like arrangement in the I state. By contrast, if a continuous stretch of residues has similar chemical shifts in I and U ($\Delta = 1$), one can conclude that this region is disordered in the intermediate state. For example, Figure 4 shows ^{15}N $|\Delta|$ ratios for G48V Abp1p SH3[18] (a) and G48 mutants of the Fyn SH3 domain[2] (b, c) mapped onto structures of the corresponding folded proteins. It is also possible to generate conformational ensembles of I using a biased molecular dynamics protocol[30] where the experimental ^{15}N Δ ratios are used as restraints, as illustrated for both G48M and G48V Fyn SH3 (Figure 4b,c, bottom).[2] Both of the "representations" in Figure 4b,c show structure formation in the $\beta3-\beta4$ hairpin (distal loop) and in a portion of strands $\beta2$, $\beta3$, and $\beta4$, although there are clear differences in the extent of "native-like" structure that has been formed in each intermediate.

NMR Based φ-Value Analysis of Intermediate and Transitions States. A very powerful method for assessing the formation of interactions along protein folding pathways involves quantifying how the energy of each state varies upon mutation by φ-value analysis.[31] In this approach the one-dimensional free-energy landscape that characterizes the folding reaction is obtained for the wild-type and a series of mutant proteins, and the ratio $\varphi_A = \Delta\Delta G_{AU}/\Delta\Delta G_{FU}$ is calculated (see Figure 5a). If the stabilities of states A (= TS1, I, and TS2) and F are equally affected by mutation ($\varphi_A = 1$) this provides strong evidence for the formation of native-like contacts in state A at the site of mutation, while if $\varphi_A = 0$, then the site is as disordered in state A as it is in the U state. Values of $\varphi_A \neq 1$ are more difficult to interpret but can be analyzed rigorously providing that the appropriate mutations are made.[32] A distinct advantage of φ-value analysis via NMR relaxation dispersion measurements is that accurate kinetic constants, k_{mn}, can be obtained, from which the free-energy landscape is constructed (Figure 5a), even for three-site folding reactions that proceed through low-populated intermediate states. Indeed, values of $\Delta\Delta G$ as small as 0.25 kcal/mol could be measured precisely.[19,22]

Figure 5b shows results obtained from φ-value analysis at two positions in the A39V/N53P/V55L Fyn SH3 domain. Thr47 is surface-exposed and located at the N-terminus of $\beta4$; φ_A values for the T47S replacement are close to 1, indicating that native-like contacts are formed early during folding and that such interactions are maintained throughout the folding process. By contrast, the φ profile for the hydrophobic core mutation, F20L, is more complex. The low φ_{TS1} value indicates "unfolded-like" structure at position 20 early on, while $\varphi_I = 0.73$ can be interpreted, based on a large body of data not described here,[19,22] as reflecting a compact intermediate with formation of at least some non-native interactions. These

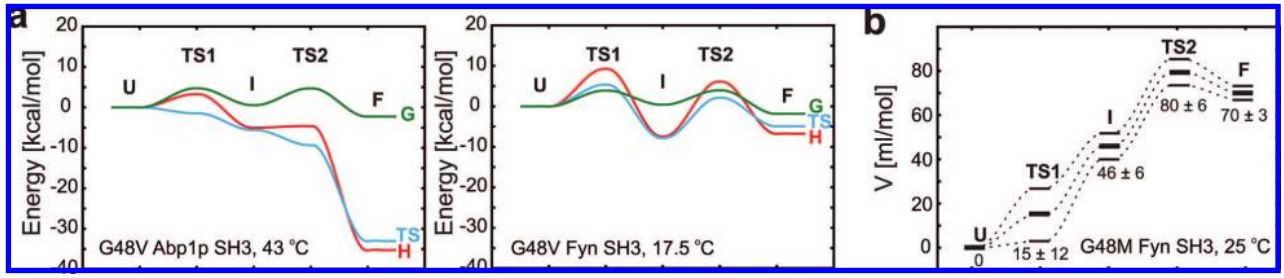


FIGURE 6. (a) One-dimensional free-energy profiles (green), along with the enthalpic (red) and entropic (blue) contributions along the folding pathway of G48V Abp1p SH3[18] (left panel) and G48V Fyn SH3[2] (right panel) domains and (b) partial molar volume, *V*, along the folding pathway of G48M Fyn SH3. Details of data analysis and parameters used are given in the original references.[20] Modified from Korzhnev et al.[18] and Bezsonova et al.[20]

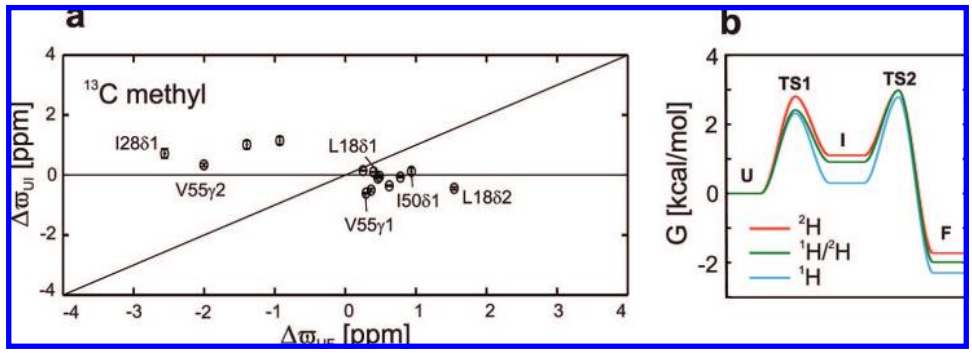


FIGURE 7. (a) Isoleucine (δ1), leucine, and valine methyl chemical shift differences, $\Delta\tilde{\omega}_{UI}$, as a function of $\Delta\tilde{\omega}_{UF}$ obtained from analysis of methyl relaxation dispersion data recorded on a sample of a highly deuterated, isoleucine, leucine, and valine methyl protonated G48M Fyn SH3 domain, 25 °C (the sample was also protonated in the side-chain positions of phenylalanine and tryptophan).[23] (b) Free-energy profiles along the folding pathway of a fully deuterated (^{2}H),[27] partially deuterated (^{1}H/^{2}H; labeling as described in panel a),[23] and fully protonated (^{1}H) G48M Fyn SH3 domain,[2] 25 °C. Modified from Mittermaier et al.[23]

"incorrect" interactions must subsequently be broken prior to formation of the F state, giving rise to a lower value of φ_{TS2}.

Thermodynamic Properties of Intermediate and Transition States. The change in free energy between states *m* and *n*, ΔG_{mn} ($G_m - G_n$), can be decomposed into contributions from enthalpy, ΔH_{mn}, and entropy, ΔS_{mn}, through temperature-dependent studies of folding/unfolding kinetics.[31] Figure 6a shows the one-dimensional folding energy landscape obtained from an analysis of temperature-dependent ^{15}N CPMG relaxation dispersion data recorded on G48V mutants of the Abp1p and Fyn SH3 domains.[2,18] Not surprisingly, the relatively small free-energy difference between U and F states of both proteins ($\Delta G_{UF} \approx 2$ kcal/mol) is the result of larger mutually compensating enthalpic, ΔH_{UF}, and entropic, $T\Delta S_{UF}$, terms. This delicate balance is an example of classical entropy/enthalpy compensation caused by a significant loss in configurational entropy of the polypeptide chain upon folding and a decrease in enthalpy upon folding due to the formation of favorable contacts between protein groups.[31] Complicating the interpretation of the macroscopic thermodynamic parameters is the contribution from solvent, since depending on the temperature, for example, the loss in entropy associated with chain compaction can be well compensated by a gain due to release of water. The compensating enthalpic and entropic terms are expected to increase with temperature due to a large positive heat capacity difference between U and F states that originates primarily from the increased level of hydration in the unfolded state.[31] Such an effect has been observed in a comparison of ΔH_{UF} and $T\Delta S_{UF}$ measured for the Abp1p and Fyn SH3 domains, where values for the Abp1p SH3 domain at 43 °C are significantly larger than those obtained for the Fyn SH3 domain at 17.5 °C (Figure 6a).

Pressure is an important variable that can be used to obtain volumetric properties of various states along the protein folding pathway, providing insight into chain hydration and packing. Figure 6b shows the partial molar volume profile for the G48M Fyn SH3 domain obtained in a pressure-dependent ^{15}N CPMG relaxation dispersion study of this protein.[20] In general, the application of increasing pressure destabilizes proteins since the protein–solvent system occupies a smaller volume when the protein is in the unfolded state. This is attributed to (i) a decrease in the *intrinsic* protein volume upon unfolding due to the elimination of microcavities and packing defects that may exist in the folded state and (ii) a decrease in *hydration* volume upon unfolding due to an

increase of the solvent shell around the protein where the water molecules are more tightly packed than in bulk water.[33] As expected, the volume of the protein–solvent system increases gradually along the folding trajectory of the G48M Fyn SH3 domain, with the intermediate state occupying a volume larger than that of U but smaller than that of F. Since the intermediate state is thought to be more loosely packed than the folded conformer (leading potentially to a larger *intrinsic* protein volume for I relative to F) the fact that $V_I < V_F$ argues that I is hydrated, with the compensation between protein and solvent volumes discussed above producing the observed *net* volume profile of the protein–solvent system, Figure 6b.

Side-Chain Interactions along Protein Folding Pathways. As described above, dispersion studies of protein folding are not restricted to backbone probes that are sensitive to secondary structure formation during folding. Robust methods have also been developed for methyl groups that generate complementary information by providing insight into formation of tertiary interactions.[15,16] Figure 7a shows a plot of the ^{13}C methyl chemical shift differences between states I and U ($\Delta\tilde{\omega}_{UI}$) correlated with differences between states F and U ($\Delta\tilde{\omega}_{UF}$) for the G48M Fyn SH3 domain. These shifts were obtained from a relaxation dispersion study involving a sample in which the isoleucine (δ1), leucine, and valine methyl groups were "$^{13}CH_3$"-labeled in an otherwise highly deuterated protein environment.[23] Shown also is a line corresponding to $\Delta\tilde{\omega}_{UI} = \Delta\tilde{\omega}_{UF}$. Clearly the methyl ^{13}C shifts of the I state are "U-like" suggesting that the side chains in this state are highly dynamic and that stable, long-lived tertiary interactions are not present and hence are only formed during the final stages of the folding trajectory. Dynamic side chains can still make contributions to stability as illustrated in Figure 7b where the folding energy landscape for the G48M Fyn SH3 domain is shown as a function of protein deuteration level. It is well-known that deuteration decreases the strength of van der Waals (VDW) interactions; hence, by following the reaction kinetics vs deuteration, it is possible to quantify formation of VDW interactions during folding.[23] For example, the decrease in stability of TS1 with deuteration suggests that there are some side-chain interactions in the early stages of folding, but because the side chains are most likely very dynamic (Figure 7a), these interactions are of a very diffuse nature. Interestingly, the rate of transition from I to F increases with deuteration. This is consistent with the picture that emerges from φ-value analysis (see above) in which non-native structure formed in the I state must be broken prior to the correct assembly. Since deuteration weakens VDW interactions, the transition from non-native to native-like structure is most easily accomplished in a deuterated molecule, accounting for the acceleration in rate observed experimentally. In summary, the kinetic, thermodynamic, and structural data described above argue for a "molten-globule-like" folding intermediate in SH3 domains with a partially structured backbone and a collapsed yet highly mobile hydrophobic core.

Concluding Remarks

We have described NMR relaxation dispersion methodology for studies of millisecond time scale multistate exchange processes in proteins and presented one application involving the structural and thermodynamic characterization of the SH3 domain folding pathway. Folding of SH3 domains proceeds through the formation of a low-populated on-pathway intermediate that cannot be observed directly in even the most sensitive of NMR experiments. Nevertheless, by manipulations involving temperature, pressure, amino-acid substitution, and different labeling strategies, including variable levels of deuteration, it is possible to "visualize" these invisible excited states and to characterize them in a detailed manner. It is anticipated that the strategies presented here will be applicable to the study of other protein folding trajectories and beyond folding to a large number of important biomolecular conformational processes.

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K. and through postdoctoral funding from the CIHR Training Program in Protein Folding and Disease (D.K.). L.E.K. is the recipient of a Canada Research Chair in Biochemistry.

BIOGRAPHICAL INFORMATION

Dmitry M. Korzhnev attended Moscow Institute of Physics and Technology obtaining a M.S. in Applied Physics and Mathematics (1995) and a Ph.D. in biophysics (1999, with A.S. Arseniev). He then moved to the Swedish NMR Centre at Göteborg University as a postdoctoral fellow (2000–2001, with M. Billeter), and subsequently to the University of Toronto as a postdoctoral fellow (2001 to the present, with L. E. Kay).

Lewis E. Kay attended the University of Alberta, obtaining a B.S. in Biochemistry (1983), and Yale University, obtaining a Ph.D. in Molecular Biophysics (1988, with J.H. Prestegard). He joined the National Institutes of Health as a postdoctoral fellow (1988 to 1992, with A. Bax), the Toronto County Medium Security Correction Facility as Scholar in residence (1991–1992), and the University of Toronto first as Assistant Professor (1992–1995) and then as Professor of Biochemistry, Chemistry, and Medical Genetics (1995 to the present).

REFERENCES

1 Pawson, T. Protein modules and signaling networks. *Nature* **1995**, *373*, 573–580.

2 Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **2004**, *430*, 586–590.

3 Hill, R. B.; Bracken, C.; DeGrado, W. F.; Palmer, A. G. Molecular motions and protein folding: Characterization of the backbone dynamics and folding equilibrium of α_2D using ^{13}C NMR spin relaxation. *J. Am. Chem. Soc.* **2000**, *122*, 11610–11619.

4 Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005**, *438*, 117–121.

5 Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **2006**, *313*, 1638–1642.

6 Popovych, N.; Sun, S. J.; Ebright, R. H.; Kalodimos, C. G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831–838.

7 Palmer, A. G.; Kroenke, C. D.; Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **2001**, *339*, 204–238.

8 Grey, M. J.; Wang, C. Y.; Palmer, A. G. Disulfide bond isomerization in basic pancreatic trypsin inhibitor: Multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J. Am. Chem. Soc.* **2003**, *125*, 14324–14335.

9 Tolkatchev, D.; Xu, P.; Ni, F. Probing the kinetic landscape of transient peptide-protein interactions by use of peptide ^{15}N NMR relaxation dispersion spectroscopy: Binding of an antithrombin peptide to human prothrombin. *J. Am. Chem. Soc.* **2003**, *125*, 12432–12442.

10 Loria, J. P.; Rance, M.; Palmer, A. G. A relaxation-compensated Carr−Purcell−Meiboom−Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **1999**, *121*, 2331–2332.

11 Ishima, R.; Torchia, D. A. Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J. Biomol. NMR* **2003**, *25*, 243–248.

12 Orekhov, V. Y.; Korzhnev, D. M.; Kay, L. E. Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J. Am. Chem. Soc.* **2004**, *126*, 1886–1891.

13 Korzhnev, D. M.; Kloiber, K.; Kay, L. E. Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: Theory and application. *J. Am. Chem. Soc.* **2004**, *126*, 7320–7329.

14 Ishima, R.; Baber, J.; Louis, J. M.; Torchia, D. A. Carbonyl carbon transverse relaxation dispersion measurements and ms-μs timescale motion in a protein hydrogen bond network. *J. Biomol. NMR* **2004**, *29*, 187–198.

15 Skrynnikov, N. R.; Mulder, F. A. A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion NMR measurements: Application to methionine residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **2001**, *123*, 4556–4566.

16 Korzhnev, D. M.; Kloiber, K.; Kanelis, V.; Tugarinov, V.; Kay, L. E. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: Application to a 723-residue enzyme. *J. Am. Chem. Soc.* **2004**, *126*, 3964–3973.

17 Mulder, F. A. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Measurement of slow (μs-ms) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* **2001**, *123*, 967–975.

18 Korzhnev, D. M.; Neudecker, P.; Zarrine-Afsar, A.; Davidson, A. R.; Kay, L. E. Abp1p and Fyn SH3 domains fold through similar low-populated intermediate states. *Biochemistry* **2006**, *45*, 10175–10183.

19 Neudecker, P.; Zarrine-Afsar, A.; Choy, W. Y.; Muhandiram, D. R.; Davidson, A. R.; Kay, L. E. 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.* **2006**, *363*, 958–976.

20 Bezsonova, I.; Korzhnev, D. M.; Prosser, R. S.; Forman-Kay, J. D.; Kay, L. E. Hydration and packing along the folding pathway of a pair of SH3 domains by pressure dependent NMR. *Biochemistry* **2006**, *45*, 4711–4719.

21 Di Nardo, A. A.; Korzhnev, D. M.; Stogios, P. J.; Zarrine-Afsar, A.; Kay, L. E.; Davidson, A. R. Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7954–7959.

22 Neudecker, P.; Zarrine-Afsar, A.; Davidson, A. R.; Kay, L. E. Φ-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15717–15722.

23 Mittermaier, A.; Korzhnev, D. M.; Kay, L. E. Side-chain interactions in the folding pathway of a Fyn SH3 domain mutant studied by relaxation dispersion NMR spectroscopy. *Biochemistry* **2005**, *44*, 15430–15436.

24 Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **1996**, *263*, 627–636.

25 Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile δ1 methyl-protonated ^{15}N, ^{13}C, ^{2}H labeled proteins. *J. Biomol. NMR* **1999**, *13*, 369–374.

26 Lundström, P.; Teilum, K.; Carstensen, T.; Bezsonova, I.; Wiesner, S.; Hansen, F.; Religa, T. L.; Akke, M.; Kay, L. E. Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]- or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone C$^{\alpha}$ and side-chain methyl positions in proteins. *J. Biomol. NMR* **2007**, *38*, 199–212.

27 Korzhnev, D. M.; Neudecker, P.; Mittermaier, A.; Orekhov, V. Y.; Kay, L. E. 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.* **2005**, *127*, 15602–15611.

28 Capaldi, A. P.; Radford, S. E. Kinetic studies of beta-sheet protein folding. *Curr. Opin. Struct. Biol.* **1998**, *8*, 86–92.

29 Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.

30 Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* **2002**, *324*, 151–163.

31 Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman: New York, 1999.

32 Fersht, A. R.; Sato, S. Φ-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7976–7981.

33 Chalikian, T. V. Volumetric properties of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 207–235.