Side-Chain Interactions in the Folding Pathway of a Fyn SH3 Domain Mutant Studied by Relaxation Dispersion NMR Spectroscopy†

Anthony Mittermaier,‡ Dmitry M. Korzhnev,‡,§ and Lewis E. Kay*,‡,§

Departments of Biochemistry, Chemistry, and Medical Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada M5S 1A8

Received September 1, 2005; Revised Manuscript Received October 14, 2005

ABSTRACT: A major challenge to the study of protein folding is the fact that intermediate states along the reaction pathway are generally unstable and thus difficult to observe. Recently developed NMR relaxation dispersion experiments present an avenue to accessing such states, providing kinetic, thermodynamic, and structural information for intermediates with small (greater than or equal to ∼1%) populations at equilibrium. We have employed these techniques to study the three-state folding reaction of the G48M Fyn SH3 domain. Using ^{13}C-, ^{1}H-, and ^{15}N-based methods, we have characterized backbone and side-chain interactions in the folded, unfolded, intermediate, and transition states, thereby mapping the energy landscape of the protein. We find that the intermediate, populated to approximately 1%, contains nativelike structure in a central β-sheet, and is disordered at the amino and carboxy termini. The intermediate is stabilized by side-chain van der Waals contacts, yet ^{13}C chemical shifts indicate that methyl-containing residues remain disordered. This state has a partially structured backbone and a collapsed yet mobile hydrophobic core and thus closely resembles a molten globule. Nonpolar side-chain contacts are formed in the unfolded-intermediate transition state; these interactions are disrupted in the intermediate-folded transition state, possibly allowing side chains to rearrange as they adopt the native packing configuration. This work illustrates the power of novel relaxation dispersion experiments in characterizing excited states that are "invisible" in even the most sensitive of NMR experiments.

The mechanism by which proteins spontaneously fold into their functional native forms has been the subject of great interest since it was noted by Levinthal that the process cannot be an unbiased search (*1*). Many studies have focused on the structures of transition states and partly folded intermediates, as these provide insight into how the free energy landscape drives folding. Because the populations of most intermediates and transition states are very small in a native environment, it is often necessary to probe the corresponding structural ensembles indirectly, by investigating the effects of mutations and buffer conditions on folding/unfolding kinetics. Rate constants are typically measured using a stopped-flow instrument, in which reactions are initiated by rapid changes in denaturant concentration and are monitored spectroscopically in real time. For proteins that fold with rates of a few hundred to a few thousand per second (or in some cases slower) and equilibrium constants ranging from approximately 10^{-2} to 10^{2}, nuclear magnetic resonance (NMR)[1] spectroscopy can also be used to characterize folding. Exchange between conformations with different chemical shifts leads to broadening of peaks in NMR spectra, and folding parameters may be extracted by line shape analysis (*2*). Carr−Purcell−Meiboom−Gill (CPMG) or relaxation dispersion experiments provide an even richer source of protein folding data, yielding exchange rates, populations, and structural information in the form of chemical shift differences between the exchanging species (*3*).

Backbone ^{15}N NMR relaxation dispersion experiments were recently used to investigate single-site mutants of the SH3 domain from the Fyn tyrosine kinase (*4*, *5*). SH3 domains contain ∼60 amino acids in a five-stranded β-sandwich fold and function within larger proteins as modules that recognize proline-rich peptide targets (*6*). They are widely studied as a model system for protein folding, and divergent SH3 domains appear to obey a common mechanism that is two-state, according to calorimetric, equilibrium, and kinetic unfolding/refolding experiments (*7*, *8*). The variants of the Fyn SH3 domain studied by NMR provided evidence of a more complex SH3 domain folding pathway. Strikingly, when ^{15}N backbone dispersion data were fit on a per-residue basis to a two-site model of exchange, both folding and unfolding rate constants were not uniform, with values differing by a factor of as much as 10 for different residues (*5*). This result is incompatible with a global two-site folding process, since all backbone probes should experience identical folding/unfolding kinetics. For the two mutants examined previously, G48M and G48V, NMR relaxation dispersion

† This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K., D.M.K., and A.M. acknowledge postdoctoral support from the CIHR and the Natural Sciences and Engineering Research Council of Canada, respectively. L.E.K. holds a Canada Research Chair in Biochemistry.

* To whom correspondence should be addressed in the Department of Medical Genetics. Phone: (416) 978-0741. Fax: (416) 978-6885. E-mail: kay@pound.med.utoronto.ca.

‡ Department of Biochemistry.

§ Departments of Chemistry and Medical Genetics.

[1] Abbreviations: SH3 domain, Src homology 3 domain; NMR, nuclear magnetic resonance.

10.1021/bi051771o CCC: $30.25
Published on Web 11/02/2005

data could, however, be well fit by a global folding/unfolding model, provided that a three-site folding equilibrium among folded, unfolded, and intermediate states was assumed. Backbone ^{15}N chemical shifts extracted from the NMR data allowed structural models of the intermediate state to be calculated, using as restraints the number of native contacts formed in this state. These contacts, in turn, were estimated on the basis of the positions of intermediate state ^{15}N chemical shifts relative to those for the unfolded and folded forms of the protein. Ensembles of intermediate structures generated by this method were ordered in the central β-sheet and disordered in the terminal regions (*5*). Studies of Fyn SH3 domain folding kinetics have similarly established that the central β-sheet is formed and that the termini are disordered in the rate-limiting step of folding (*9*).

Our initial studies used backbone amide probes of the SH3 domain folding intermediate, and did not provide any information about the hydrophobic core. With this in mind, we have performed a more detailed characterization of the G48M folding pathway, placing particular emphasis on the role of side-chain interactions, using a highly deuterated, ^{15}N-enriched sample with selective ^{13}C and ^{1}H labeling at key hydrophobic sites (see below). We present here for the first time an application involving a suite of six backbone $^{1}H-^{15}N$ relaxation dispersion experiments to probe exchange in proteins and confirm the existence of an intermediate with some native backbone structure. In addition, methyl-based NMR relaxation dispersion experiments were performed, to probe the local environment of hydrophobic residues in the folded, unfolded, and intermediate states. We compare the exchange parameters obtained for this selectively protonated molecule with those generated for a perdeuterated sample. $^{1}H-^{2}H$ substitution perturbs van der Waals (VDW) interactions (*10*), and the effects of deuteration on populations and exchange rates allow an assessment of the relative contributions of VDW energies to the stabilities of the three major forms of G48M and the intervening transition states.

MATERIALS AND METHODS

Sample Preparation. Two samples of the Fyn tyrosine kinase SH3 domain (chicken isoform) with a Gly 48 to Met mutation were employed in this study. The first was uniformly labeled with ^{15}N and contained ^{1}H at all exchangeable hydrogen positions. ^{13}C and ^{1}H were incorporated into the C^{δ} methyl groups of Leu and Ile, and the C^{γ} methyls of Val residues; the side chains of Phe and Trp were fully protonated. All other positions contained ^{2}H and ^{12}C. The second sample was uniformly labeled with ^{15}N and ^{12}C, and contained ^{2}H at nonexchangeable and ^{1}H at exchangeable positions. Protein expression and purification are described elsewhere (*4*, *11*−*13*).

NMR Experiments. ^{15}N single-quantum (*14*, *15*), ^{1}H single-quantum (*16*), $^{15}N-^{1}H$ double-quantum and zero-quantum (*17*), and $^{15}N-^{1}H$ multiple-quantum (*18*) relaxation dispersion experiments were performed on the perdeuterated sample at 25 °C and on the selectively protonated sample at 20, 25, and 30 °C, using Varian INOVA spectrometers operating at proton frequencies of 500, 600, and 800 MHz. ^{13}C single-quantum pulse schemes (*19*) were employed for the selectively protonated sample at 20, 25, and 30 °C using 500 and 800 MHz spectrometers. Relaxation delays for all dispersion experiments were between 20 and 40 ms.

Data Fitting. Effective transverse relaxation rates, $R_{2,\mathrm{eff}}$, were calculated from the ratio of peak intensities in spectra with (I) and without (I_0) relaxation delays of length T, according to the equation $R_{2,\mathrm{eff}} = -\ln(I/I_0)/T$. Backbone data were fit together using in-house software, employing both two-site (unfolded ⇔ folded) and three-site (unfolded ⇔ intermediate ⇔ folded) models of exchange, as described previously (*5*, *17*, *18*) and in the Supporting Information. Fits were performed by adjusting exchange parameters to minimize $\chi^2 = \sum(R_{2,\mathrm{eff}}^{\mathrm{observed}} - R_{2,\mathrm{eff}}^{\mathrm{calculated}})^2/\sigma_{R_{2,\mathrm{eff}}}^2$, where $\sigma_{R_{2,\mathrm{eff}}}$ is the experimental uncertainty of each $R_{2,\mathrm{eff}}$ value, and the sum extends over all measurements included in the fit. Reduced χ^2 values were calculated as χ^2/(number of degrees of freedom) and have an expectation value of 1. In the case of three-site fits, a large number of similar χ^2 minima were found, resulting from either a rugged χ^2 surface or incomplete convergence of the minimizations. Reported values (uncertainties) are the means (standard deviations) of the best six solutions for the fully deuterated molecule and the best 10 solutions for the selectively protonated molecule. Methyl ^{13}C dispersions were fit by holding exchange rates fixed to the backbone results and optimizing chemical shift parameters for each methyl group. This was repeated for each of the 10 backbone solutions, and reported values and uncertainties are the means and standard deviations of the solutions, as for the backbone.

Theoretical ^{13}C chemical shift calculations were generously performed by X. Salvatella (Cambridge University, Cambridge, U.K.) using PROSHIFT (*20*). Chemical shifts were calculated for 25 separate structures previously computed for the G48M I state (*5*), and the values for each methyl position were averaged. The calculated chemical shifts of all Val $^{13}C^{\gamma 1,2}$ nuclei were very similar, as were those of Leu $^{13}C^{\delta 1,2}$ and those of Ile $^{13}C^{\delta}$ nuclei. These values showed a uniform offset of approximately −1 ppm relative to tabulated random-coil methyl chemical shifts; 1 ppm was thus added to each calculated value prior to comparison with experimental data. Tabulated values for *pro-R* and *pro-S* methyl groups based on experiment differ by ∼1 ppm in both Leu and Val residues. The values calculated using PROSHIFT for *pro-R* and *pro-S* methyls were virtually identical, suggesting that the uncertainties in the predicted values are at least 1 ppm.

RESULTS

Two- versus Three-Site Exchange Models. ^{15}N and ^{1}H NMR relaxation dispersion data were collected for both perdeuterated and selectively protonated forms of the G48M Fyn SH3 domain using a suite of recently developed NMR techniques (*16*−*18*) that focus on different ^{1}H and ^{15}N coherences. As shown in Figure 1, the dynamics of a single backbone amide group lead to qualitatively quite different dispersion profiles among the set of experiments. The large amount of data obtained for each residue has allowed us to very precisely characterize the exchange dynamics for every well-resolved amide peak. Moreover, analysis of the six dispersion profiles (see Figure 1) does not require a concomitant increase in the number of fitting parameters that characterize the global exchange process (populations, rates, and chemical shift differences) over that required in fits of ^{1}H and ^{15}N single-quantum dispersion profiles exclusively, although the amount of data has increased substantially so that robust

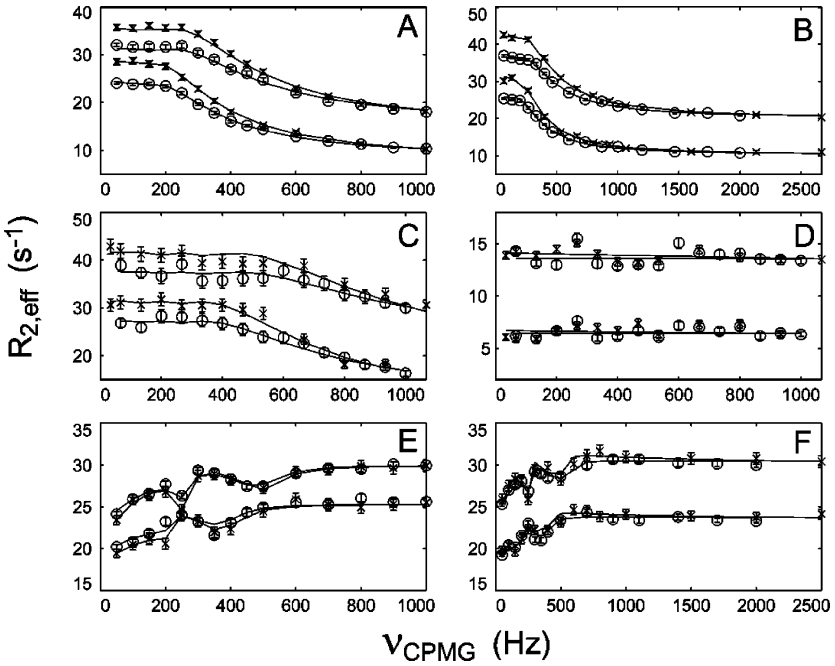

FIGURE 1: Backbone amide dispersion profiles of Glu11 for deuterated (×) and partially protonated (○) G48M samples. Effective transverse relaxation times ($R_{2,eff}$) are plotted as a function CPMG pulse repetition frequency, ν_{CPMG}, for ^{15}N single-quantum (A), ^{1}H single-quantum (B), $^{15}N-^{1}H$ double-quantum (C), $^{15}N-^{1}H$ zero-quantum (D), and $^{15}N-^{1}H$ multiple-quantum coherences employing ^{15}N (E) and ^{1}H (F) pulse trains. Data obtained from 600 and 800 MHz spectrometers are plotted in the bottom and top portions, respectively, of each panel. Vertical offsets [$R_{2,eff}(\infty)$] do not depend on exchange parameters, and have been adjusted to highlight differences between the dispersion profiles.

measures of exchange can be extracted. It is noteworthy that in the absence of large-amplitude millisecond time scale motions, all dispersion profiles in Figure 1 would be flat.

As described above, we have previously shown that ^{15}N relaxation dispersion data for a fully protonated version of the G48M domain are incompatible with a simple two-site exchange process between unfolded and folded states (U ⇔ F) and that a three-site exchange model, involving an additional intermediate state (U ⇔ I ⇔ F), must be invoked (*5*). Not surprisingly, the U ⇔ F model is not compatible with the data for both the perdeuterated and selectively protonated G48M samples. This can be seen very simply by a comparison of folding/unfolding rates obtained on a per-residue basis assuming two-site exchange, which clearly show rates that differ for many residues (data not shown; for a two-site folding mechanism, equal folding rates should be observed for all residues as well as equal unfolding rates).

As expected, the dispersion profiles can, however, be well fit to a three-site folding equilibrium. For example, in the case of the selectively protonated molecule, the χ^2 values from three-site fits of ^{15}N and ^{1}H single-quantum data show 2.2- and 1.7-fold reductions, respectively, compared to corresponding two-site fits. *F* test statistics indicate that the probabilities of the observed reductions occurring purely by chance are vanishingly small. In contrast, the overall agreement between experimental data and those back-calculated from three-site exchange models is good, with reduced χ^2 parameters of 1.0 and 1.4 for perdeuterated and selectively protonated molecules, respectively, when all backbone measurements are included in the fits.

^{13}C NMR single-quantum relaxation dispersion data (*19*) were measured for the selectively protonated G48M sample, providing site-specific measures of conformational exchange at δ-methyl positions of Leu and Ile residues and γ-methyl positions of Val residues. The ^{13}C single-quantum data show better agreement with a two-site exchange model than do those of ^{15}N or ^{1}H nuclei, and switching to a three-site exchange model produces an only 1.1-fold reduction in χ^2, compared with 2.2- and 1.7-fold reductions for ^{15}N and ^{1}H, respectively. As discussed below, this is because methyl ^{13}C chemical shifts of the I state are uniformly close to those of the U state, rendering the process effectively two-site. Indeed, fits of a three-site exchange model to ^{13}C data alone fail to produce a well-defined solution. To obtain detailed three-site ^{13}C chemical shift information, we fit methyl dispersion data while holding the global exchange parameters fixed to values calculated for the backbone. ^{13}C relaxation dispersion data are well reproduced by this model, with a reduced χ^2 value of 1.0, compared to a value of 0.9 for three-site fits of methyl data alone.

Chemical Shift Parameters. NMR chemical shifts depend on the local electronic environment of individual nuclei, and thus provide a sensitive, site-specific probe of protein structure. Only one set of peaks, corresponding to the major conformer of the domain, is observed in NMR spectra of G48M. The peaks are well dispersed, as is typical for a folded protein, indicating that the preferentially populated form of G48M is the F state. Although the U and I states are invisible in standard NMR experiments, chemical shift differences between these states and the F state may be extracted from NMR relaxation dispersion data. Backbone chemical shifts obtained for the corresponding U and I states of the perdeuterated and selectively protonated samples are very similar, with root-mean-square deviations of 0.05 (U state) and 0.17 ppm (I state) for ^{1}H nuclei and 0.47 (U state) and 1.0 ppm (I state) for ^{15}N nuclei. This result gives us confidence in the accuracy of the dispersion-derived chemical shift values and suggests that the structures of the U and I

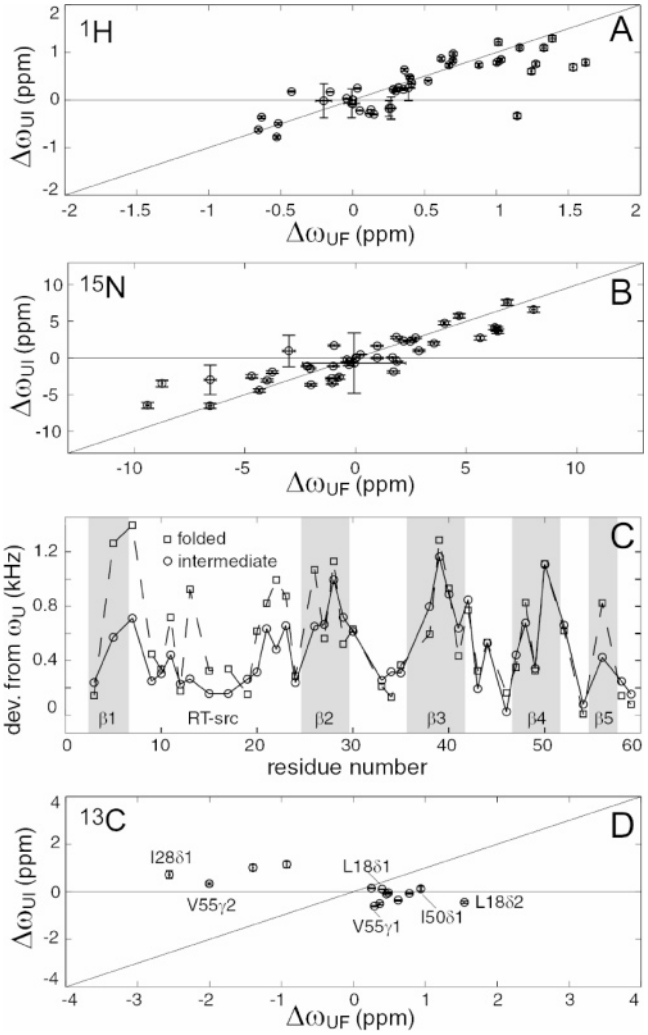

FIGURE 2: Chemical shift differences, $\Delta\omega_{UI}$ ($=\omega_I - \omega_U$), plotted as a function of $\Delta\omega_{UF}$, for ^{1}H (A), ^{15}N (B), and ^{13}C (D) nuclei. Values were extracted from fits of dispersion data as described in Materials and Methods. Horizontal (diagonal) lines indicate $\omega_I = \omega_U$ ($\omega_I = \omega_F$). Deviations from U state backbone chemical shifts, $(\Delta\omega^{15}{}_N{}^2 + \Delta\omega^1{}_H{}^2)^{0.5}$, calculated in absolute frequency units (800 MHz spectrometer), for I (solid) and F (dashed) states, plotted as a function of residue number (C). In panel D, labeled peaks correspond to hydrophobic core residues.

states are not greatly affected by side-chain proton/deuterium substitution.

Amide ^{15}N and ^{1}H as well as methyl ^{13}C chemical shifts obtained for the U state are in good agreement with tabulated random-coil values (*21*), confirming that the U state does, in fact, represent the unfolded form of the protein. The chemical shifts of the unfolded state correspond to those obtained in the absence of local ordering, and deviations from these values are diagnostic for the formation of structure. In what follows, therefore, U state chemical shifts have been used as a reference.

Chemical shift differences between states U and I ($\Delta\omega_{UI} = \omega_I - \omega_U$, in parts per million) are plotted in Figure 2A,B as a function of the differences between U and F ($\Delta\omega_{UF} = \omega_F - \omega_U$, in parts per million) for ^{1}H and ^{15}N nuclei of the selectively protonated G48M protein. The diagonal lines have a slope of 1 and correspond to the case in which $\omega_I = \omega_F$. It is clear that in the intermediate state, backbone chemical shifts deviate significantly from random-coil values and, in

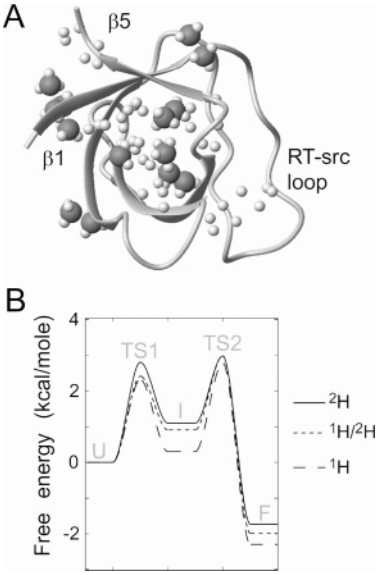

FIGURE 3: (A) Ribbon diagram of the Fyn SH3 domain [Protein Data Bank entry 1SHF (*36*)] with the sites of selective side-chain protonation indicated by light-colored spheres. ^{13}C-labeled methyl groups are represented by dark-colored spheres. (B) Energy level diagram calculated for perdeuterated, selectively protonated, and fully protonated G48M proteins using the kinetic parameters in Table 1. First-order rate constants, k_{XY}, for the conversion of state X to state Y were assumed to obey Arrhenius kinetics, where $k_{XY} = k_o \exp[-(G_{TS} - G_X)/RT]$, where G_{TS} and G_X are the free energies of the transition and X states, respectively, and k_o (10^5 s^{-1}) is a prefactor suitable for protein folding (*37*). For the sake of comparison, the U states of the three proteins have been arbitrarily assigned free energies of 0.

many cases, correspond to those of the folded state. This shows that the backbone is more ordered in the I state than in the U state, and suggests that some regions of the I state have nativelike structure. To identify locally ordered regions of the I state, absolute ^{15}N−^{1}H peak displacements for the I and F states relative to the U state have been plotted as a function of residue number in Figure 2C. Profiles for the I and F states virtually overlap in β-strands 2−4, while the F−U peak displacements are greater than I−U displacements in β-strands 1 and 5, and in the first (RT-src) loop. This suggests that the central β-sheet is largely formed in the I state, while N- and C-terminal regions are more disordered, in good agreement with results for the fully protonated version of G48M (*5*).

Strikingly dissimilar results were obtained for methyl groups in G48M. Plotted in Figure 2D are ^{13}C chemical shift differences between the intermediate and unfolded states, $\Delta\omega_{UI}$, that are close to zero, irrespective of $\Delta\omega_{UF}$ values. The I state is thus uniformly close to the U state in terms of ^{13}C chemical shifts. This is consistent with the observation, noted above, that ^{13}C dispersion data are more compatible with a two-site model of exchange than either ^{1}H or ^{15}N data. Six of the methyl groups characterized in this study are located in the hydrophobic core, as may be seen in Figure 3A, and are therefore well situated to report on the extent of native side-chain packing in the intermediate. Several of these groups have ^{13}C chemical shifts in the F state that differ significantly from random coil values, due to nearby aromatic residues and conformational constraints in the folded protein. Much smaller deviations from the U state are found for these methyls in the I state, and the differences are opposite in

Table 1: Global Three-Site Exchange Parameters for the G48M Fyn SH3 Domain (25 °C)

sample	k_{UI}[a] (s^{-1})	k_{IU}[a] (s^{-1})	k_{IF}[a] (s^{-1})	k_{FI}[a] (s^{-1})	$\Delta G_{U,I}$[b] (kcal/mol)	$\Delta G_{U,F}$[c] (kcal/mol)
^2H[d]	880 ± 30	5600 ± 200	4200 ± 300	35 ± 1	1.10 ± 0.04	−1.73 ± 0.01
$^1H/^2H$[e]	1700 ± 200	8000 ± 200	3000 ± 400	22 ± 1	0.91 ± 0.06	−1.99 ± 0.01
^1H[f]	2000 ± 200	3000 ± 300	1500 ± 100	20 ± 1	0.3 ± 0.1	−2.30 ± 0.02

[a] k_{XY} is the first-order rate constant for the conversion of state X to state Y, where $\Delta G_{A,B} = G_B - G_A$. [b] $\Delta G_{U,I} = -RT \ln(k_{UI}/k_{IU})$, and 1 cal = 4.1868 J. [c] $\Delta G_{U,F} = -RT[\ln(k_{UI}/k_{IU}) + \ln(k_{IF}/k_{FI})]$. [d] Data for the fully deuterated sample, as described in Materials and Methods. [e] Data for the selectively protonated sample, as described in Materials and Methods. [f] Data for a fully protonated sample, taken from ref *5*.

sign with respect to $\Delta\omega_{UF}$. Thus, the intermediate does not possess a nativelike packing arrangement, although further analysis presented below suggests that side-chain interactions persist in this form of the protein. Interestingly, when the ensemble of structures calculated previously for the intermediate on the basis of amide ^{15}N chemical shifts (*5*) is used to predict methyl ^{13}C chemical shifts, values very close to those of random coils (*21*) are obtained, in good agreement with our results here. The overall correspondence between I and F state chemical shifts for amide groups, and I and U state chemical shifts for methyl groups, provides a picture of the intermediate in which the central region of the backbone has adopted a largely nativelike fold while the side chains remain highly disordered.

Global Exchange Parameters. The rates of interconversion among U, I, and F states were extracted from backbone relaxation dispersion data for both perdeuterated and selectively protonated samples of G48M and are listed in Table 1. For the sake of comparison, data for the fully protonated molecule have been included as well (*5*). Small but significant differences in global exchange parameters are observed for the three proteins, reflecting differences in dispersion profiles clearly evident in Figure 1. Notably, the samples are identical in amino acid sequence and solvent conditions, differing in only the isotopic labeling pattern. In particular, the positions represented by light-colored spheres in Figure 3A contained ^2H in the perdeuterated and ^1H in the selectively protonated samples.

Deuteration of side chains destabilizes proteins without greatly perturbing their three-dimensional structures (*22*). The origin of this effect is the higher mass of deuterons relative to protons and the asymmetry of the C−H interatomic potential. C−^2H bonds experience different ground-state vibrations, and are slightly shorter (*23*) and less polarizable (*24*) than C−^1H bonds. This manifests in weaker VDW interaction energies and different physical properties for deuterium isotopomers. For example, deuteriomethane has a higher vapor pressure than proteomethane, reflecting weaker attractive forces in the liquid state (*25*). In a result that is highly relevant to protein folding, a quantitative HPLC study showed that deuterated solutes had shorter retention times on reverse-phase chromatographic media than their protonated analogues, due to less favorable interactions with the stationary hydrophobic phase (*10*). Although it is likely that C−^2H moieties also experience less favorable direct interactions with water than C−^1H groups do, the HPLC results clearly demonstrate that the dominant effect of deuteration is the disruption of nonpolar clustering. In protein folding, therefore, side-chain deuteration is likely destabilizing because it leads to less favorable interactions in the hydrophobic core of the native state. It must be noted that H^α atoms can participate in hydrogen bonds and isotope effects have not been well characterized for this interaction. However, care was taken to ensure that both protein samples used in this study were deuterated at all hydrogen α-positions and differences in stability between the samples can be unambiguously interpreted in terms of side-chain VDW interactions.

Global exchange parameters were used to construct energy-level diagrams, shown in Figure 3B, where we have arbitrarily set the U state energy at zero for each of the ^2H, $^1H/^2H$, and ^1H proteins. Thus, it is important to keep in mind that any comparison of free energy differences upon deuteration reflects changes in free energies relative to the appropriate reference U states. Increasing levels of ^2H incorporation progressively destabilize the folded state relative to the unfolded state. Most of the residues with $^1H/^2H$ substitutions reside in the hydrophobic core of G48M and make extensive side-chain interactions in the native protein, and it is not surprising that deuteration destabilizes the F state. Notably, since the intermediate is similarly destabilized by deuteration, these residues likely form VDW interactions in this form of the protein, as well. Thus, the I state is collapsed to some degree about a core of hydrophobic residues such that intramolecular contacts are made by Ile, Leu, Val, Phe, and Trp side chains. The ^2H dependence of the global exchange parameters (Table 1) can be also be used as a valuable probe of VDW contacts along the folding reaction coordinate. For example, the transition from the unfolded to the intermediate state, governed by k_{UI}, is slowest in the fully deuterated sample. Therefore, this energetic barrier is increased by ^2H substitution, consistent with the formation of side-chain VDW interactions in the first transition state, TS1. In contrast, deuteration increases the rate of exchange between the intermediate and folded states; i.e., both k_{IF} and k_{FI} become larger. As illustrated in Figure 3B, this is because both I and F states are destabilized to a greater extent by ^2H substitution than TS2, indicating that they contain a greater number of VDW contacts than this transition state.

DISCUSSION

Structure of the Intermediate. We have found that the I state retains nativelike structure in the central portion of the backbone while the amino- and carboxy-terminal regions are less ordered. The protein forms a loose hydrophobic core that is stabilized by VDW interactions, while ^{13}C chemical shifts indicate that side chains remain highly disordered. The picture bears a striking resemblance to the molten globule state, which is populated by many proteins under mildly destabilizing conditions or as a kinetic intermediate during folding (*26*). A variety of experimental techniques have shown that molten globules have nativelike topologies and are slightly expanded relative to the folded state. They are

stabilized more by nonspecific hydrophobic collapse than by precise packing arrangements, and are highly dynamic on a millisecond to microsecond time scale (*26*−*28*). In particular, the ability of molten globules to bind the hydrophobic dye 1-anilinonaphthalene-8-sulfonic acid (ANS) indicates that they contain solvent-exposed clusters of nonpolar atoms. Also, far-UV CD spectra show that a significant amount of secondary structure is retained, while near-UV CD spectra point to a loss of specific packing around aromatic side chains (*26*). The experimental techniques mentioned above are very different from the NMR relaxation dispersion experiments performed in this study. Therefore, the similarity between molten globules and the emerging picture of the G48M intermediate is remarkable, with both states showing significant backbone ordering together with collapsed yet highly mobile hydrophobic cores. Notably, the small population of the G48M I state would make it invisible to the population-weighted spectroscopic measurements typically used to identify molten globules, underscoring the importance of these new NMR techniques to the study of protein folding intermediates. It is worth emphasizing that in this study information about the molten globule-like intermediate state has been obtained indirectly, through the well-resolved resonances of the F state. In general, NMR studies of equilibrium molten globules typically produce spectra with extremely broad lines due to extensive millisecond to microsecond time scale motions so that direct studies of such systems are difficult; not surprisingly, there is limited chemical shift information for these molecules. In one study, where Trp and Phe residues of α-lactalbumin were selectively labeled with ^{19}F, one-dimensional spectra of folded, unfolded, and molten globule states showed that molten globule chemical shifts were significantly closer to those of the unfolded state than to those of the folded state (*28*), as we have found for ^{13}C chemical shifts in the G48M intermediate here.

Structures of the Transition States. Computer simulations can generate atomic-resolution models of protein folding intermediates and transition states at a level of detail inaccessible to experiment. A number of calculations have been performed for the src-SH3 domain, generating results in good agreement with mutational studies (*29*−*31*). The folding mechanisms of the src and Fyn SH3 domains are similar (*8*), and these data are therefore relevant to G48M.

Computational studies have mapped the free energy landscape along the folding pathway and found two separate barriers, implying the existence of an intermediate between the folded and unfolded states. The central β-sheet, comprising strands 2−4, is formed in the first transition state, while the first and last β-strands and the first (RT-src) loop are disordered. The structure is very similar to that of the rate-limiting transition states of both src and Fyn SH3 domains, as characterized by mutational analyses with the assumption of a two-state folding reaction (*9*, *32*). The second transition state, which connects the I and F states, involves two preformed hydrophobic surfaces, one comprising strands β2−β4 and the other β1, β5, and the RT-src loop, which desolvate in order to adopt the native side-chain packing configuration (*29*−*31*). More than 10 water molecules are expelled from between the two β-sheets, in a step that has been proposed to constitute a universal feature of hydrophobic core formation (*31*).

Our results for the G48M intermediate are in close agreement with the predicted structure of the first transition state, which contains a structured central β-sheet and disordered termini. This supports the idea that the simulations and the NMR relaxation dispersion experiments are monitoring the same process. Also, we find that the rate constant k_{UI} is diminished with deuteration (Table 1); i.e., VDW interactions are formed in TS1, in good agreement with the theoretical calculations. Just prior to the second transition state, the simulations predict that hydrophobic core side chains reside in two separate clusters (β2−β4 and β1, β5, and the RT-src loop), an arrangement that is presumably disrupted for formation of the native core. We find that TS2 is less destabilized by deuteration than either the I or F state, indicating that VDW interactions are broken and then re-form as the intermediate converts to the folded state. Taken together, these findings suggest that a portion of the TS2 energy barrier results from breaking apart non-native hydrophobic clusters so that side-chain configurations may be remodeled. Several proteins have been found to fold via on-pathway intermediates containing non-native side-chain packing (*33*−*35*). Our results indicate that disrupting such interactions may represent a common energetic barrier in protein folding.

As a final note, it is worth re-emphasizing that the NMR analysis has been based on the folding model, U ⇔ I ⇔ F. What evidence do we have to support such an on-pathway mechanism? First, as discussed above, the features of the I state deduced by NMR measurements coincide with those of intermediates obtained from molecular dynamics simulations of the folding reaction. Second, although our data are fit to a linear reaction scheme of the form A ⇔ B ⇔ C, there are no a priori assumptions that would bias fitting toward on- or off-pathway reactions (i.e., F ⇔ I ⇔ U vs F ⇔ U ⇔ I). The lowest χ^2 solutions are those where the end points of the reaction (i.e., A and C) correspond to the F and U states, as established on the basis of the extracted chemical shifts. Third, there is a good correspondence between those regions of the I state that show nativelike structure and those regions of the transition state that are "nativelike", as measured by ϕ-value analysis, under the assumption of a two-state folding mechanism (*9*).

In summary, we have used NMR relaxation dispersion experiments to study backbone and side-chain interactions in a folding intermediate and two transition states of the G48M mutant of the Fyn SH3 domain. The results confirm that the intermediate has a structured central β-sheet and disordered termini, at the level of backbone contacts, and show that (1) both the first transition state and intermediate are stabilized by VDW interactions, (2) side chains are highly disordered in the intermediate, and (3) VDW interactions are disrupted in the second transition state, only to re-form in the native protein. Considering the amount of information obtained for the intermediate, it is noteworthy that this state is populated to only approximately 1% at equilibrium. In fact, since it is slightly less stable than the unfolded state, it is nearly invisible in conventional folding kinetics and equilibrium hydrogen exchange experiments. NMR relaxation dispersion experiments greatly extend our ability to detect such intermediates, and allow the extraction of structural, kinetic, and thermodynamic information, leading to a better understanding of protein folding pathways.

ACKNOWLEDGMENT

We thank Dr. A. Davidson (University of Toronto) for supplying the G48M expression plasmid, Dr. X. Salvatella (Cambridge University) for chemical shift calculations, and Dr. C. Dobson (Cambridge University) for useful discussions.

SUPPORTING INFORMATION AVAILABLE

Plot of ^{13}C relaxation dispersion profiles, table of temperature-dependent exchange constants obtained for selectively protonated G48M, and detailed description of the NMR data fitting protocol. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Levinthal, C. (1969) in *Mossbauer spectroscopy in biological systems* (de Brunner, J. T. P., and Munck, E., Eds.) pp 22−24, University of Illinois Press, Allerton House, Monticello, IL.
2. Burton, R. E., Busby, R. S., and Oas, T. G. (1998) ALASKA: A Mathematica package for two-state kinetic analysis of protein folding reactions, *J. Biomol. NMR 11*, 355−359.
3. Palmer, A. G., Kroenke, C. D., and Loria, J. P. (2001) NMR methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods Enzymol. 339*, 204−238.
4. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E., and Davidson, A. R. (2004) Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7954−7959.
5. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M., and Kay, L. E. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR, *Nature 430*, 586−590.
6. Musacchio, A., Wilmanns, M., and Saraste, M. (1994) Structure and function of the SH3 domain, *Prog. Biophys. Mol. Biol. 61*, 283−297.
7. Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999) A thermodynamic analysis of a family of small globular proteins: SH3 domains, *Biophys. Chem. 77*, 195−208.
8. Capaldi, A. P., and Radford, S. E. (1998) Kinetic studies of β-sheet protein folding, *Curr. Opin. Struct. Biol. 8*, 86−92.
9. Northey, J. G. B., Maxwell, K. L., and Davidson, A. R. (2002) Protein folding kinetics beyond the ϕ value: Using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state, *J. Mol. Biol. 320*, 389−402.
10. Turowski, M., Yamakawa, N., Meller, J., Kimata, K., Ikegami, T., Hosoya, K., Tanaka, N., and Thornton, E. R. (2003) Deuterium isotope effects on hydrophobic interactions: The importance of dispersion interactions in the hydrophobic phase, *J. Am. Chem. Soc. 125*, 13836−13849.
11. Maxwell, K. L., and Davidson, A. R. (1998) Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides the best prediction of stability effects, *Biochemistry 37*, 16172−16182.
12. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., and Kay, L. E. (1999) 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 13*, 369−374.
13. Crowhurst, K. A., and Forman-Kay, J. D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain, *Biochemistry 42*, 8687−8695.
14. Loria, P. J., Rance, M., and Palmer, A. G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy, *J. Am. Chem. Soc. 121*, 2331−2332.
15. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain, *J. Am. Chem. Soc 123*, 11341−11352.
16. Ishima, R., and Torchia, D. A. (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach, *J. Biomol. NMR 25*, 243−248.
17. Orekhov, V. Y., Korzhnev, D. M., and Kay, L. E. (2004) Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins, *J. Am. Chem. Soc. 126*, 1886−1891.
18. Korzhnev, D. M., Kloiber, K., and Kay, L. E. (2004) Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: Theory and application, *J. Am. Chem. Soc. 126*, 7320−7329.
19. Skrynnikov, N. R., Mulder, F. A. A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001) 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. 123*, 4556−4566.
20. Meiler, J. (2003) PROSHIFT: Protein chemical shift prediction using artificial neural networks, *J. Biol. NMR 26*, 25−37.
21. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR 5*, 67−81.
22. Brockwell, D., Yu, L., Cooper, S., McCleland, S., Cooper, A., Attwood, D., Gaskell, S. J., and Barber, J. (2001) Physicochemical consequences of the perdeuteriation of glutathione S-transferase from *S. japonicum*, *Protein Sci. 10*, 572−580.
23. Bartell, L. S., and Roskos, R. R. (1966) Isotope Effects on Molar Volume and Surface Tension: Simple Theoretical Model and Experimental Data for Hydrocarbons, *J. Chem. Phys. 44*, 457.
24. Bates, F. S., Fetters, L. J., and Wignall, G. D. (1988) Thermodynamics of Isotopic Polymer Mixtures: Poly(Vinylethylene) and Poly(Ethylethylene), *Macromolecules 21*, 1086−1094.
25. Armstrong, G. T., Brickwedde, F. G., and Scott, R. B. (1955) Vapor Pressures of the Methanes, *J. Res. Natl. Bur. Stand. (U.S.) 55*, 39−52.
26. Ptitsyn, O. B. (1995) Molten globule and protein folding, *Adv. Protein Chem. 47*, 83−229.
27. Redfield, C. (2004) Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins, *Methods 34*, 121−132.
28. Bai, P., Luo, L., and Peng, Z. Y. (2000) Side chain accessibility and dynamics in the molten globule state of α-lactalbumin: A F-19-NMR study, *Biochemistry 39*, 372−380.
29. Guo, W. H., Lampoudi, S., and Shea, J. E. (2004) Temperature dependence of the free energy landscape of the src-SH3 protein domain, *Proteins: Struct., Funct., Bioinf. 55*, 395−406.
30. Guo, W. H., Lampoudi, S., and Shea, J. E. (2003) Posttransition state desolvation of the hydrophobic core of the src-SH3 protein domain, *Biophys. J. 85*, 61−69.
31. Shea, J. E., Onuchic, J. N., and Brooks, C. L. (2002) Probing the folding free energy landscape of the src-SH3 protein domain, *Proc. Natl. Acad. Sci. U.S.A. 99*, 16064−16068.
32. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding, *Nat. Struct. Biol. 6*, 1016−1024.
33. Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002) Im7 folding mechanism: Misfolding on a path to the native state, *Nat. Struct. Biol. 9*, 209−216.
34. Feng, H. Q., Vu, N. D., and Bai, Y. W. (2004) Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b_{562}: Native fold with non-native hydrophobic interactions, *J. Mol. Biol. 343*, 1477−1485.
35. Canet, D., Lyon, C. E., Scheek, R. M., Robillard, G. T., Dobson, C. M., Hore, P. J., and van Nuland, N. A. J. (2003) Rapid formation of non-native contacts during the folding of HPr revealed by real-time photo-CIDNP NMR and stopped-flow fluorescence experiments, *J. Mol. Biol. 330*, 397−407.
36. Noble, M. E., Mussachio, A., Saraste, M., Courtneidge, S. A., and Wierenga, R. K. (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.
37. Alm, E., Morozov, A. V., Kortemme, T., and Baker, D. (2002) Simple physical models connect theory and experiment in protein folding kinetics, *J. Mol. Biol. 322*, 463−476.

BI051771O