Published in final edited form as:

J Mol Biol. 2009 October 2; 392(4): 1055–1073. doi:10.1016/j.jmb.2009.07.073.

Quantitative Determination of Site-Specific Conformational Distributions in an Unfolded Protein by Solid State Nuclear Magnetic Resonance

Kan-Nian Hu, Robert H. Havlin, Wai-Ming Yau, and Robert Tycko^{*}

Laboratory of Chemical Physics National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health Bethesda, Maryland 20892-0520

Summary

Solid state nuclear magnetic resonance (NMR) techniques are used to investigate the structure of the 35-residue villin headpiece subdomain (HP35) in folded, partially denatured, and fully denatured states. Experiments are carried out in frozen glycerol/water solutions, with chemical denaturation by guanidine hydrochloride (GdnHCl). Without GdnHCl, two-dimensional solid state ¹³C NMR spectra of samples prepared with uniform ¹³C labeling of selected residues show relatively sharp crosspeaks at chemical shifts that are consistent with the known three-helix bundle structure of folded HP35. At high GdnHCl concentrations, most crosspeaks broaden and shift, qualitatively indicating disruption of the folded structure and development of static conformational disorder in the frozen denatured state. Conformational distributions at one residue in each helical segment are probed quantitatively with three solid state NMR techniques that provide independent constraints on backbone ϕ and ψ torsion angles in samples with sequential pairs of carbonyl ¹³C labels. Without GdnHCl, the combined data are well fit by α -helical conformations. At [GdnHCl] = 4.5 M, corresponding to the approximate denaturation midpoint, the combined data are well fit by a combination of α -helical and partially extended conformations at each site, but with a site-dependent population ratio. At [GdnHCl] = 7.0 M, corresponding to the fully denatured state, the combined data are well fit by a combination of partially extended and polyproline II conformations, again with a site-dependent population ratio. Two entirely different models for conformational distributions lead to nearly the same best-fit distributions, demonstrating the robustness of these conclusions. This work represents the first quantitative investigation of site-specific conformational distributions in partially folded and unfolded states of a protein by solid state NMR.

Keywords

solid state NMR; protein folding; villin; HP35; polyproline II

^{*}corresponding author: Dr. Robert Tycko, National Institutes of Health, Building 5, Room 112, Bethesda, MD 20892-0520. phone: 301-402-8272; fax: 301-496-0825; e-mail: robertty@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Supplementary Material Experimental 2DEXMAS, CTDQFD, and DQCSA data for HP35-CO-AV, HP35-CO-SA, and HP35-CO-LK at [GdnHCl] = 0.0 M, 4.5 M, and 7.0 M (Figures S1, S2, and S3); direct comparisons of experimental 2DEXMAS, CTDQFD, and DQCSA data for HP35-CO-AV, HP35-CO-SA, and HP35-CO-LK at [GdnHCl] = 7.0 with best-fit simulations (Figure S4); results of MCMC simulations for HP35-CO-SA and HP35-CO-LK, using Gaussian models and multiple-point models (Figures S5 and S6); results of simulations to investigate effects of orientation-dependent spin polarizations and variations in CSA asymmetry parameters on fits to the solid state NMR data (Tables S1 and S2).

Introduction

Properly folded proteins are crucial to normal biological functions, while their misfolding can lead to disease. The phenomenon of protein folding, defined as the kinetic or thermodynamic progression from a structurally disordered to a structurally ordered state of a protein, is therefore centrally important to many biological and medical questions. Since protein folding involves both folded and unfolded states, the properties of both of these states can influence the thermodynamic, kinetic, and mechanistic aspects of protein folding¹. The folded states of many proteins are well characterized structurally, primarily through x-ray crystallography and multidimensional nuclear magnetic resonance (NMR) of proteins in fluid solution. Structural properties of unfolded states are more difficult to characterize, because unfolded states adopt a distribution of conformations and because experimental techniques for determining conformational distributions in chemically complex macromolecules are not well established. Partial information about unfolded protein structures is available from solution NMR techniques²⁻⁵, including measurements of nuclear Overhauser effects⁶, chemical shifts⁷, scalar couplings, transverse spin relaxation rates⁸, paramagnetic relaxation enhancements⁹⁻¹¹, and residual dipole-dipole couplings¹²⁻¹⁴. Among other findings, these techniques show that unfolded states may contain residual secondary structure¹⁵, sequence-specific preferential population of turn structures¹⁶, and greater long-range structural order than expected for a simple random coil⁹⁻¹¹. Solution NMR techniques provide a mixture of site-specific and long-range conformational constraints that are averaged over multiple, rapidly exchanging conformations in an unfolded state. In certain cases, solution NMR data are well described by models for unfolded states in which local backbone conformational distributions are similar to distributions derived from folded protein structure databases¹⁷⁻¹⁹, perhaps with additional population of extended and polyproline II (PPII) conformations¹². Additional long-range constraints on unfolded state structure are available from scattering techniques^{20,21}, which serve as measurements of the overall dimensions of a polypeptide chain, but not the conformations at specific sites. Infrared spectroscopy with isotopic labeling can also provide site-specific structural information for unfolded states, in addition to kinetic information regarding protein folding processes²².

This paper demonstrates an alternative experimental approach to investigations of unfolded protein structure that yields quantitative constraints on site-specific conformational distributions. In this approach, solid state NMR techniques are applied to partially or fully unfolded proteins in frozen solutions, with the folding state controlled by solvent conditions such as denaturant concentration. By carrying out measurements on frozen solutions, dynamic averaging of NMR parameters is avoided and the full structural heterogeneity of an unfolded state can be explored. In addition, full (rather than residual) dipole-dipole couplings and anisotropic chemical shift tensors are available as structurally dependent nuclear spin interactions in the solid state. Multiple independent solid state NMR measurements can be carried out on a single protein sample over an extended period of time, provided that the sample remains frozen, so that the same approach would apply to freeze-trapped transient species. In principle, each data point in each measurement represents a constraint on the conformational distribution at a given site in the protein, in particular a site where isotopic labels such as ¹⁵N or ¹³C have been introduced. If enough data points are acquired, the full conformational distribution can in principle be determined, using appropriate models for the conformational distribution and appropriate simulations of the solid state NMR measurements. Several related studies in which solid state NMR data were used to extract conformational distributions in disordered biopolymers²³⁻²⁵ and synthetic polymers²⁶⁻²⁸ have been reported.

Experiments described below were carried out on the 35-residue villin headpiece subdomain (HP35), a protein that has been studied extensively as a simple model system for protein folding by a variety of experimental^{22,29-40} and computational⁴¹⁻⁵³ methods. Earlier studies have

shown that HP35 (or the 36-residue version, HP36) has the primarily helical folded state shown in Fig. 1a^{31,36,37,40,54}, with an unfolding temperature near 70° C in the absence of chemical denaturant³⁵, and that the time scale for equilibration of folded and unfolded states is approximately 5 μ s near room temperature in aqueous solution^{22,34,39}. In our experiments described below, partially and fully unfolded states of HP35 were prepared by addition of the denaturant guanidine hydrochloride (GdnHCl) to glass-forming glycerol/water solutions, and solid state NMR measurements were carried out at -120° C. As previously reported³² and confirmed below, two-dimensional (2D) solid state ¹³C NMR spectra of samples prepared with uniform ¹³C labeling of selected residues indicate that the three-helix folded state of HP35 is preserved in these frozen solutions in the absence of denaturant, but that addition of GdnHCl disrupts the folded state, with a denaturation midpoint at [GdnHCl]_{mid} $\approx$ 4.5 M. Interestingly, this low-temperature denaturation midpoint is not significantly different from the denaturant midpoint at higher temperatures^{32,35}, suggesting that our experiments on frozen solutions are not without relevance to properties of HP35 in fluid solutions at higher temperatures. The temperature-independence of the denaturation midpoint is supported by low-temperature circular dichroism (CD) data described below.

We report results of quantitative measurements of backbone conformational distributions at one site in each helical segment (V50, A57, and K70), using three independent solid state NMR techniques⁵⁵⁻⁵⁸. These techniques are all applicable to polypeptides that are ¹³C-labeled at two sequential backbone carbonyl sites (A49/V50, S56/A57, and L69/K70 in the experiments reported below), and provide constraints on the ϕ and ψ torsion angles between the labeled sites, as shown in Fig. 1b. Analyses of the quantitative conformational data verify the α -helical conformations at [GdnHCl] = 0.0 M, indicate mixtures of α -helical and extended, β -strand-like conformations in the partially denatured state at [GdnHCl] = 4.5 M, and reveal conformational distributions that include both PPII-like and extended conformations in the fully denatured state at [GdnHCl] = 7.0 M. Site-specific differences in the conformational distributions in the partially and fully denatured states are discussed below.

Results

Qualitative picture of chemical denaturation in frozen solutions from 2D solid state NMR

Chemical denaturation of HP35 in frozen solutions was first monitored at a qualitative level by 2D solid state ¹³C NMR spectroscopy of samples in which the protein was uniformly ¹⁵N and ¹³C labeled at selected residues. For these experiments, HP35 (residues 42-76 of the 76-residue chicken villin headpiece domain³⁵, amino acid sequence L₄₂SDEDFKAVFGMTRSAFANLPLWKQNLKKEKGLF76, with the underlined segments forming α -helices in the folded state) was labeled at A49, G52, M53, F58, P62, and L69 (HP35-U-AGMFPL) or V50, A57 and L69 (HP35-U-VAL).

Figure 2 shows 2D spectra of HP35-U-AGMFPL and HP35-U-VAL in frozen glycerol/water solutions with [GdnHCl] = 0.0 M, 4.0 M, 5.0 M, and 7.0 M. These spectra were obtained with magic-angle spinning (MAS), proton decoupling in the two spectral dimensions, and radio-frequency-driven recoupling⁵⁹ (RFDR) during the 1.71-1.79 ms mixing periods. Sample temperatures were -120° C. At [GdnHCl] = 0.0 M, crosspeaks with 2.0-2.5 ppm linewidths (full width at half maximum, FWHM) were observed that connect NMR lines of all directly-bonded ¹³C pairs. The relatively narrow crosspeaks, similar in width to crosspeaks in 2D solid state NMR spectra of antibody-bound peptides in frozen solution^{60,61} and of transmembrane peptides in phospholipid bilayers at low temperatures⁶² under similar measurement conditions, indicate a well-ordered, rigid protein conformation in a non-crystalline environment. ¹³C NMR chemical shifts extracted from the 2D spectrum at [GdnHCl] = 0.0 M are given in Table 1. These chemical shifts are consistent with the expected α -helical folded structure of HP35. In particular, secondary shifts (*i.e.*, differences between observed ¹³C chemical shifts in HP35

and “random coil” shifts for unstructured peptides in fluid solutions) for CO, C $_{\alpha}$, and C $_{\beta}$ sites are positive, positive, and negative, respectively, for A49, V50, A57, F58, and L69, as expected for residues in helical segments⁶³⁻⁶⁷. Secondary shifts for G52, M53, and P62 are non-helical, also as expected based on the folded HP35 structure. ¹³C $_{\alpha}$ and ¹³C $_{\beta}$ chemical shifts in Table 1 agree to within a 0.5 ppm root-mean-squared deviation with values extrapolated to low temperature from solution NMR measurements on HP35 in aqueous solution in the 10-30°C range (unpublished data).

With addition of GdnHCl, crosspeaks in the 2D spectra shift and broaden, but the extent of broadening is site-dependent. Signals from M53, F58, and L69 shift and broaden significantly, so that their C $_{\alpha}$ /C $_{\beta}$ crosspeaks are unresolved at [GdnHCl] = 7.0 M. These spectral changes reflect the disruption of the helical HP35 folded structure and the development of distributions of conformations (and hence distributions of ¹³C chemical shifts for individual carbon sites) in the denatured state. Signals from A49 remain relatively sharp at [GdnHCl] = 7.0 M (2.5 ppm linewidths), although the ¹³C $_{\alpha}$ chemical shift for A49 is reduced by 2.0 ppm and the ¹³C $_{\beta}$ chemical shift is increased by 0.6 ppm (Figure 2a). As previously reported³², signals from V50 in HP35-U-VAL have a similar behavior (Figure 2b). The CO/C $_{\alpha}$ crosspeak for G52 moves approximately 2 ppm towards a lower ¹³C $_{\alpha}$ chemical shift value at [GdnHCl] = 7.0 M, but is relatively narrow in the denatured state. The C $_{\alpha}$ /C $_{\beta}$ crosspeak of A57 is broader at [GdnHCl] = 7.0 M, shifting towards lower ¹³C $_{\alpha}$ chemical shift values but retaining measurable intensity at the position of the folded state. Variations in crosspeak widths in the denatured state suggest variations in conformational distributions. However, quantitative analysis of crosspeak shapes is precluded by the uncertain dependence of individual ¹³C $_{\alpha}$ and ¹³C $_{\beta}$ chemical shifts on sidechain torsion angles and non-local interactions.

Crosspeaks for P62 show comparatively minor changes in chemical shifts and linewidths, reflecting the fact that proline residues have limited conformational freedom even in “random coil” peptides. The 3.6 ppm and 4.2 ppm differences between C $_{\beta}$ and C $_{\gamma}$ chemical shifts in the folded and unfolded states, respectively, indicate *trans* peptide bonds between L61 and P62 in both states⁶⁸.

The evolution of crosspeak positions and shapes with increasing [GdnHCl] indicates that the denaturation midpoint lies between 4 M and 5 M. This is especially clear for the C $_{\alpha}$ /C $_{\beta}$ crosspeaks of A49, V50, and L69 and the CO/C $_{\alpha}$ crosspeaks of V50 and G52. Quantitative conformational measurements described below also indicate a denaturation midpoint near [GdnHCl] = 4.5 M, consistent with CD data at -40°C reported previously³². CD measurements at -55°C discussed below also indicate [GdnHCl]_{mid} = 4.3 ± 0.1 M.

Quantitative constraints on backbone conformational distributions

For quantitative conformational measurements, three samples were prepared with ¹³C labels at sequential backbone carbonyl sites: A49 and V50 (HP35-CO-AV), S56 and A57 (HP35-CO-SA), and L69 and K70 (HP35-CO-LK). As shown in Figure 1b, labeling of these carbonyl sites allows constraints on the ϕ and Ψ torsion angles of V50, A57, and K70 to be obtained from solid state NMR. Three techniques were used to obtain independent constraints, namely 2D MAS exchange spectroscopy^{57,58,69} (2DEXMAS), constant-time double-quantum-filtered dipolar dephasing⁵⁶ (CTDQFD), and double-quantum chemical shift anisotropy spectroscopy⁵⁵ (DQCSA). Radio-frequency (rf) pulse sequences for 2DEXMAS, CTDQFD, and DQCSA measurements are shown in Figure 3. The accuracy of these techniques in conformational studies has been demonstrated on model peptides^{23,55,58}, and the same techniques have been applied in previous studies of peptide/antibody complexes⁶⁰ and helix-forming peptides^{23,56} in frozen solutions, amyloid fibrils^{70,71}, and other systems⁷². Measurements were carried out in frozen glycerol/water with [GdnHCl] = 0.0 M, 4.5 M, and 7.0 M, corresponding to the fully folded state, the approximate denaturation midpoint, and the

fully denatured state. Experimental data for HP35-CO-AV at [GdnHCl] = 0.0 M are shown in Figure 4.

The 2DEXMAS spectrum (Figure 4a), obtained with the rotor-synchronized pulse sequence in Figure 3a with a 500 ms mixing period and relatively slow MAS, contains crosspeaks that connect carbonyl MAS sideband lines. Crosspeak volumes depend on the relative orientation of the two labeled carbonyl chemical shift anisotropy (CSA) tensors (Figure 1c), which in the case of HP35-CO-AV depends on the ϕ and ψ torsion angles of V50 (Figure 1b). Thus, the crosspeak volumes in the 2DEXMAS spectrum of HP35-CO-AV place constraints on the conformational distribution of HP35 at V50. Crosspeaks connecting MAS sidebands of order 0, ± 1 , and ± 2 had measurable intensities (signal-to-noise ≈ 10 for the strongest crosspeaks), yielding 20 data points (*i.e.*, crosspeak volumes) from each 2DEXMAS spectrum.

CTDQFD data (Figure 4b) show the build-up and decay of double-quantum-filtered ^{13}C NMR signal amplitudes from the carbonyl labels. Excitation and evolution of double-quantum (DQ) coherences occurs under radio-frequency-driven dipolar recoupling⁵⁹ (RFDR) sequences, making CTDQFD data sensitive to the ^{13}C - ^{13}C dipole-dipole coupling, or the ^{13}C - ^{13}C distance, which in turn depends on ϕ . The RFDR sequences consist of trains of rotor-synchronized ^{13}C π pulses, with one π pulse in every two MAS rotor periods. Under these experimental conditions, the RFDR recoupling mechanism depends on differences in CSA tensor orientations, making the CTDQFD data additionally dependent on ψ . CTDQFD data were obtained in a “constant-time” manner (*i.e.*, by incrementing M - N while keeping M + N constant in Figure 3b), so that these data were insensitive to transverse ^{13}C spin relaxation rates and could be analyzed quantitatively for ^{13}C - ^{13}C distances up to at least 4.2 Å, covering the entire range of possible ϕ values.

DQCSA data (Figure 4c) show the dephasing of DQ coherences, excited by an RFDR sequence, under the sum of CSA interactions of the A49 and V50 labels (in the case of HP35-CO-AV). The sum of the CSA interactions for a dipole-coupled pair of carbonyl ^{13}C labels depends on the intervening ϕ and ψ angles, making the DQCSA data sensitive to both ϕ and ψ . CSA interactions were recoupled in a constant-time manner by the three ^{13}C π pulses in the $2\tau_R$ period between DQ excitation and mixing periods in Figure 3c, making the DQCSA data insensitive to ^{13}C transverse relaxation, isotropic chemical shifts, and inhomogeneous broadening.

Analyses assuming a single conformation

2DEXMAS, CTDQFD, and DQCSA data can be simulated accurately, allowing quantitative analyses in terms of the distribution of ϕ and ψ torsion angles. The data were first analyzed with the assumption of a single conformation at V50, A57, or K70, *i.e.*, single populated values of ϕ and ψ . Figure 5 shows contour plots of the χ^2 deviation between experimental data for HP35-CO-AV and simulated data as a function of the ϕ and ψ angles assumed in the simulations (see Materials and Methods for details of simulations and data analyses and for the definition of χ^2). At [GdnHCl] = 0.0 M, χ^2 plots for the individual measurements restrict the possible values of ϕ and ψ , but do not determine the best-fit values uniquely. However, when the constraints from all three measurements are combined, a single best-fit point is identified at ϕ , $\psi = -70^\circ \pm 5^\circ, -45^\circ \pm 5^\circ$, where the total χ^2 has the value $\chi^2_{\min} = 69$. This conformation for V50 in the folded state of HP35, determined from the solid state NMR data, agrees well with the values ϕ , $\psi = -68^\circ, -46^\circ$ in the crystal structure of the N68H HP35 mutant at pH 6.7 (PDB file 1yrf), determined by Chiu *et al.*³¹ Similar analyses of 2DEXMAS, CTDQFD, and DQCSA data for HP35-CO-SA and HP35-CO-LK at [GdnHCl] = 0.0 M yield best-fit values ϕ , $\psi = -65^\circ \pm 5^\circ, -45^\circ \pm 5^\circ$ for A57 ($\chi^2_{\min} = 38$) and ϕ , $\psi = -70^\circ \pm 5^\circ, -45^\circ \pm 5^\circ$ for K70 ($\chi^2_{\min} = 53$). Values from the crystal structure are ϕ , $\psi = -62^\circ, 48^\circ$ for A57 and ϕ , $\psi = -65^\circ, -47^\circ$ for K70.

For each sample, the total number of data points is $N_{\text{data}} = 48$. For A57 and K70, $|\chi^2_{\text{min}} - (N_{\text{data}} - 5)|$ is less than or approximately equal to $\sqrt{2(N_{\text{data}} - 5)}$ as expected for a good fit with five adjustable parameters (two torsion angles and scaling factors for the three data sets). For V50, we attribute the excess in χ^2_{min} to contributions from systematic errors. Systematic errors with magnitudes of approximately 70% of the typical random error due to noise in the solid state NMR data would explain the observed value of χ^2_{min} . Given the high signal-to-noise ratios of the data, especially in CTDQFD and DQCSA measurements on HP35-CO-AV, systematic errors of this size are not surprising. Possible sources of systematic errors are discussed below (see Materials and Methods).

In contrast, analyses of the data at $[\text{GdnHCl}] = 4.5 \text{ M}$ with the assumption of a single conformation leads to $\chi^2_{\text{min}} = 85, 102$, and 112 for HP35-CO-AV, HP35-CO-SA, and HP35-CO-LK, respectively. At $[\text{GdnHCl}] = 7.0 \text{ M}$, fitting the data to a single conformation leads to $\chi^2_{\text{min}} = 128, 109$, and 66 , respectively. These larger values of χ^2_{min} (with the same value of N_{data}) suggest broader ranges of populated conformations.

Analyses assuming Gaussian conformational distributions

The 2DEXMAS, CTDQFD, and DQCSA data were analyzed with models in which the probability of adopting specific ϕ, ψ values at V50, A57, or K70 has the form

$$P(\phi, \Psi) = \sum_{k=1}^{N_G} h_k e^{-[(\phi - \phi_k)^2 + (\Psi - \Psi_k)^2] / w_k^2}$$
, where N_G is the number of Gaussian components in the probability distribution. For a given choice of the parameters h_k, w_k, ϕ_k , and ψ_k , simulated data were constructed from tables of 2DEXMAS, CTDQFD, and DQCSA simulations over a grid of ϕ and ψ values. Best-fit values of the parameters were determined by simulated annealing, using the Metropolis algorithm⁷³ to vary these parameters (see Materials and Methods).

Table 2 summarizes the results of the simulated annealing calculations. Data at $[\text{GdnHCl}] = 0.0 \text{ M}$ were fit with a single Gaussian component ($N_G = 1$), again resulting in best-fit ϕ and ψ values very close to the values extracted from the HP35 crystal structure, with a small Gaussian width. Attempts to fit the data at $[\text{GdnHCl}] = 4.5 \text{ M}$ and $[\text{GdnHCl}] = 7.0 \text{ M}$ with a single Gaussian component resulted in only minor reductions in χ^2_{min} compared to the fits with a single ϕ, ψ point discussed above.

At $[\text{GdnHCl}] = 4.5 \text{ M}$, good fits were obtained with two Gaussian components ($N_G = 2$). For all three sites, one component was centered at $\phi, \psi = -58^\circ \pm 12^\circ, -52^\circ \pm 3^\circ$, corresponding to α -helical conformations. The other component was centered at $\phi, \psi = -125^\circ \pm 24^\circ, 92^\circ \pm 16^\circ$, corresponding to extended, β -strand-like conformations. The best-fit population ratios of helical to extended conformations were approximately 1:1 for V50, 2:1 for A57, and 3:2 for K70. These population ratios are roughly consistent with the denaturation midpoint at $[\text{GdnHCl}]_{\text{mid}} \approx 4.3 \text{ M}$ indicated by the 2D ^{13}C NMR spectra in Figure 2 and by CD data discussed below. The larger population of helical conformations for A57 is consistent with our earlier report that this site appears to lag behind other sites under GdnHCl denaturation.³²

At $[\text{GdnHCl}] = 7.0 \text{ M}$, good fits were also obtained with two Gaussian components, with one component centered at $\phi, \psi = -129^\circ \pm 13^\circ, 82^\circ \pm 8^\circ$ and the other component centered at $\phi, \psi = -80^\circ \pm 5^\circ, 160^\circ \pm 5^\circ$. The first component represents extended conformations similar to those observed at $[\text{GdnHCl}] = 4.5 \text{ M}$. The second component represents PPII conformations. Best-fit population ratios of extended to PPII conformations were approximately 3:2 for V50, 3:2 for A57, and 2:3 for K70.

These results suggest a common denaturation “pathway” for all three sites, in which the native helical conformation is first converted to a mixture of helical and extended conformations, and

then to a mixture of extended and PPII conformations. Populations of PPII conformations at the denaturation midpoints are apparently low, as indicated by the fact that fits with three Gaussian components ($N_G = 3$) did not improve the fits significantly (reduction in $\chi^2_{\min}$ less than 0.3, population in third Gaussian component less than 5%, at both $[\text{GdnHCl}] = 4.5 \text{ M}$ and $[\text{GdnHCl}] = 7.0 \text{ M}$).

Site-specificity of conformational distributions

The best-fit two-Gaussian results suggest certain site-specific differences in the details of the conformational differences at the three sites, including differences in the relative populations of preferred conformations and differences in the precise ϕ, ψ values. However, differences in the best-fit parameters can not be interpreted definitively without an assessment of the uncertainties in these parameters. Markov Chain Monte Carlo (MCMC) simulations (see Materials and Methods) were used to determine the ranges of fitting parameters that correspond to acceptable fits to the combined 2DEXMAS, CTDQFD, and DQCSA data. Figure 6 shows the normalized sums of conformational distributions visited during MCMC simulations for V50, using a one-Gaussian model to fit data at $[\text{GdnHCl}] = 0.0 \text{ M}$ and two-Gaussian models to fit data at $[\text{GdnHCl}] = 4.5 \text{ M}$ and 7.0 M . The fact that the functions plotted in Figure 6 have clearly localized peaks demonstrates that only conformational distributions that are similar to the best-fit distributions have low χ^2 values. Similar results were obtained in MCMC simulations for A57 and K70. Uncertainties given in Table 2 are root-mean-squared deviations of individual parameters from best-fit values during the MCMC simulations.

To assess whether the apparent site-specific differences are statistically significant, we calculated the χ^2 values for all three sites while performing the MCMC simulations for each individual site. Results are shown in Figure 7. For example, Figure 7a is a plot of $\Delta\chi^2 = \chi^2 - \chi^2_{\min}$ values for V50, A57, and K70 at $[\text{GdnHCl}] = 4.5 \text{ M}$ during MCMC simulations that sampled conformational distributions that are good fits to the V50 data. As expected, $\Delta\chi$ for V50 is small in Figure 7a, but the observation that $\Delta\chi^2$ is large for both A57 and K70 means that conformational distributions that fit the V50 data at $[\text{GdnHCl}] = 4.5 \text{ M}$ do not fit the A57 or K70 data at $[\text{GdnHCl}] = 4.5 \text{ M}$. Therefore, the conformational distribution at V50 is significantly different from conformational distributions at A57 and K70.

At $[\text{GdnHCl}] = 4.5 \text{ M}$, results in Figures 7a, 7b, and 7c show that conformational distributions at all three sites are significantly different. According to Table 2, the relative populations of helical and extended conformations are different (greatest population of extended conformations at K70), the widths of the helical component are different (greatest width at K70), and the location of the extended conformations in the ϕ, ψ plane are different (especially for V50).

At $[\text{GdnHCl}] = 7.0 \text{ M}$, results in Figures 7d and 7e show that the conformational distributions at V50 and A57 are not significantly different based on existing data, because all distributions with small $\Delta\chi^2$ for V50 also have small $\Delta\chi^2$ for A57 (although not all distributions with small $\Delta\chi^2$ for A57 also have small $\Delta\chi$ with small $\Delta\chi^2$ for V50). Results in Figure 7f show that the conformational distribution at K70 is significantly different. According to Table 2, the major difference is in the relative populations of extended and PPII conformations, with greater population of PPII conformations at K70.

Analyses assuming a multiple-point conformational distribution at V50

To test the robustness of our conclusions from one-Gaussian and two-Gaussian models, an alternative model was considered in which the conformational distributions were represented by a set of equally populated points in the ϕ, ψ plane. The number of points and their locations were varied in a simulated annealing procedure (see Materials and Methods). The virtue of the

multiple-point model is that it does not restrict the number or shape of the populated regions in the ϕ, ψ plane. A possible difficulty is that this model involves a large number of variable parameters, so that many distinct conformational distributions with similarly low χ^2 values might in principle be found. In practice, this difficulty was not encountered, indicating that the solid state NMR data are sufficient to determine the main features of the actual conformational distributions.

Figure 8 shows the results of multiple-point fits to the combined 2DEXMAS, CTDQFD, and DQCSA data for V50. At [GdnHCl] = 0.0 M, the simulated annealing calculation converged to a single cluster of 50 points near $\phi, \psi = -70^\circ, -50^\circ$, with $\chi^2_{\min} = 62$. At [GdnHCl] = 4.5 M, the calculation converged to three clusters, near $\phi, \psi = -70^\circ, -50^\circ$ (27 points), $\phi, \psi = -105^\circ, 80^\circ$ (15 points), and $\phi, \psi = -140^\circ, 160^\circ$ (two points), with $\chi^2_{\min} = 66$. At [GdnHCl] = 7.0 M, the calculation converged to four clusters, near $\phi, \psi = -120^\circ, 80^\circ$ (27 points), $\phi, \psi = -75^\circ, 160^\circ$ (17 points), $\phi, \psi = -75^\circ, 125^\circ$ (two points), and $\phi, \psi = -170^\circ, 175^\circ$ (two points), with $\chi^2_{\min} = 63$. Note that the $\chi^2_{\min}$ values for the multiple-point model are nearly the same as the $\chi^2_{\min}$ values in the Gaussian models. The locations of the two largest clusters at [GdnHCl] = 4.5 M and [GdnHCl] = 7.0 M are also in good agreement with the locations of the best-fit Gaussian components in two-Gaussian models. Thus, the multiple-point and Gaussian models lead to similar conclusions regarding the conformational distributions at V50. Similar agreement between multiple-point and Gaussian models was also obtained for all A57 and K70 data sets.

For A57 at [GdnHCl] = 7.0 M, MCMC simulations show that multiple-point distributions with small $\Delta\chi^2$ contain 12% population of α -helical conformations on the average, although the best-fit distribution has only 5% α -helical population. Similarly, MCMC simulations with a three-Gaussian model show an average α -helical population of 15%, although $\chi^2_{\min}$ for the three-Gaussian model is not significantly lower than $\chi^2_{\min}$ for the two-Gaussian model. This means that our data for HP35-CO-SA are compatible with partial population of α -helical conformations at A57 with [GdnHCl] = 7.0 M, as suggested by the 2D ^{13}C NMR spectra of HP35-U-VAL in Figure 2 (see above), but do not require α -helical conformations. In contrast, MCMC simulations for HP35-CO-LK at [GdnHCl] = 7.0 M show an average α -helical population of 2% in the multiple-point model and 0% in the three-Gaussian model, ruling out significant population of α -helical conformations at K70.

Robustness of χ^2 minimization by simulated annealing

With one exception, repeated simulated annealing runs (5-10 runs for each data set) converged to a unique best-fit conformational distribution for each data set. For V50 at [GdnHCl] = 7.0 M, repeated simulated annealing runs with the two-Gaussian model revealed that the combined 2DEXMAS, CTDQFD, and DQCSA data could also be fit with a sum of α -helical and extended conformations (*i.e.*, Gaussian functions centered at $\phi, \psi = -70^\circ, -46^\circ$ and $-111^\circ, -107^\circ$, with a 3:7 population ratio, $\chi^2_{\min} = 57$). This alternative conformational distribution was discarded because it is inconsistent with the observation of a single $\text{C}_\alpha/\text{C}_\beta$ crosspeak at a non-helical position for V50 at [GdnHCl] = 7.0 M in Fig. 2b. Population of α -helical conformations in multiple-point simulations for V50 at [GdnHCl] = 7.0 M (Figure 8f) was prevented by excluding points with both $\phi > -120^\circ$ and $\psi < -10^\circ$. No such restrictions were required in fits to all A57 and K70 data sets and to all other V50 data sets.

4. Discussion

Our results are summarized as follows: (1) 2D solid state NMR spectra of HP35-UAGMFPL and HP35-U-VAL (Figure 2) indicate a well folded structure in frozen solution without denaturant. ^{13}C NMR chemical shifts are consistent with the three-helix structure determined previously by solution NMR³⁶ and x-ray crystallography³¹, *i.e.* we find no evidence for “cold denaturation” of HP35 under our experimental conditions. The transition from folded to

denatured crosspeak positions and widths occurs between 4.0 M and 5.0 M, consistent with the value $[\text{GdnHCl}]_{\text{mid}} \approx 4.3$ M inferred from CD data (see below); (2) Quantitative conformational measurements on HP35-CO-AV, HP35-CO-SA, and HP35-CO-LK with the 2DEXMAS, CTDQFD, and DQCSA techniques (Figures 3, 4 and 5) yield ϕ, ψ values for V50, A57, and K70 at $[\text{GdnHCl}] = 0.0$ M that are in excellent agreement with the crystal structure³¹; (3) 2DEXMAS, CTDQFD, and DQCSA data at $[\text{GdnHCl}] = 4.5$ M and 7.0 M can be fit with a model in which only two regions of the ϕ, ψ plane are significantly populated, with relatively narrow Gaussian distributions in each region (Table 2). At $[\text{GdnHCl}] = 4.5$ M, the data are best fit by a sum of helical and partially extended conformations. At $[\text{GdnHCl}] = 7.0$ M, the data are fit by a combination of partially extended and PPII conformations. Thus, we identify a “pathway” for denaturation (Figure 6), in which helices are partially converted to extended segments near the denaturation midpoint, and then partially converted to PPII segments at higher denaturant concentrations; (4) Although the three sites examined in this work (one in each helix of the folded HP35 structure) follow the same pathway, our data indicate variations in the precise ratios of populated conformations and the precise ϕ, ψ values. In particular, relative populations of extended and α -helical conformations are site-dependent at $[\text{GdnHCl}] = 4.5$ M, relative populations of extended and PPII conformations are site-dependent at $[\text{GdnHCl}] = 7.0$ M, and the exact locations and widths of the populated regions in the ϕ, ψ plane are site-dependent; (5) Conclusions from the two-Gaussian fits to the 2DEXMAS, CTDQFD, and DQCSA data are corroborated by fits with a multiple-point model (Figure 8). The fact that two distinct models lead to similar conformational distributions supports the robustness of the data analysis.

Evidence for preferential population of PPII conformations in unfolded states of proteins and peptides has been obtained previously, dating back to the work of Tiffany and Krimm⁷⁴, who used CD spectra as evidence that urea- and GdnHCl-denatured peptides and proteins adopt PPII conformations that are stabilized by hydrogen bonding of denaturant molecules to backbone carbonyl groups. More recent CD measurements by Whittington *et al.*⁷⁵ on other peptides and proteins in the presence of urea support the earlier data from Tiffany and Krimm. In addition, extensive solution NMR studies of urea-denatured ubiquitin by Meier *et al.*¹² suggest that the denatured state of a protein has site-specific conformational distributions that are similar to distributions predicted from ϕ, ψ distributions extracted from folded protein structures, but with enhanced population of both β -strand and PPII conformations. Preferential population of PPII conformations in unfolded states without chemical denaturant is supported by vibrational CD studies of poly-L-proline and poly-L-glutamate by Dukor and Keiderling⁷⁶, solution NMR studies of various peptides by Kallenbach and coworkers^{77,78}, and ultraviolet resonance Raman studies of polyalanine peptides by Asher *et al.*⁷⁹ Computational studies of finite-length polyalanine peptides by Kentsis *et al.*⁸⁰ indicate that formation of segments with β -strand and PPII secondary structure is more likely than formation of α -helical segments. Computational studies by Mezei *et al.*⁸¹ suggest that PPII conformations are preferred because PPII segments have lower solvation energies in purely aqueous solution than β -strand or α -helical segments.

The two preferred conformations indicated by our solid state NMR at $[\text{GdnHCl}] = 7.0$ M are similar to the enhanced β -strand and PPII conformations suggested by Meier *et al.* for ureadenatured ubiquitin¹². As discussed above, our data additionally indicate that the PPII conformations develop primarily above the denaturation midpoint, “after” the development of extended conformations, and that the precise conformational distributions have site-specific differences. Although previous investigations of PPII conformations in unfolded and denatured proteins have been based on considerations that derive from the specific structural and chemical properties of polypeptides and their interactions with solvent and denaturant molecules, in our studies of HP35, PPII conformations emerge entirely from fitting the quantitative solid state NMR data with numerical simulations of nuclear spin dynamics. Our data analysis does not

involve molecular modeling, empirical interpretations of spectroscopic signatures, structural data bases, or *a priori* assumptions regarding allowed or preferred conformations.

The experiments described above require low temperatures, to immobilize the protein and thereby allow solid state NMR measurements, and glycerol/water solutions, to prevent solvent crystallization that would perturb the protein structure. Sample cooling rates were approximately 10 deg/s (estimated from the time required for a 240 μ l volume of glycerol/water in a transparent vial to solidify after immersion in liquid nitrogen). As the sample cools, the protein conformational distribution remains at thermal equilibrium until the sample temperature approaches the glass transition of the glycerol/water mixture. The effective temperature of our conformational measurements is then approximately -75° C. For HP35 in glycerol/water, the denaturation midpoint and thermodynamic parameters extracted from CD data were previously shown to be independent of temperature in the range from -40° C to 20° C.³² CD data for denaturation by GdnHCl in pure water and in glycerol/water are also nearly identical at 4° C.^{32,35} Figure 9 shows new low-temperature CD measurements that confirm the earlier results and extend these results to -55° C. Data in Figure 9a can be fit with a standard model³², yielding $\Delta G_0 = 3.07 \pm 0.08$ kcal/mole and $m = 0.75 \pm 0.02$ kcal/mole-M at 22° C and $\Delta G_0 = 3.38 \pm 0.39$ kcal/mole and $m = 0.79 \pm 0.09$ kcal/mole-M at -55° C (where ΔG_0 is the free energy of unfolding at [GdnHCl] = 0.0 M and m is the slope of a linear dependence of unfolding free energy on [GdnHCl]). The denaturation midpoint, given by $\Delta G_0/m$, is 4.1 ± 0.1 M at 22° C and 4.3 ± 0.2 M at -55° C. The denaturation midpoint indicated by our solid state NMR measurements (effective temperature $\approx$ -75° C) is also approximately 4.5 M. The fact that $[\text{GdnHCl}]_{\text{mid}}$ is nearly independent of temperature from -75° C to 22° C is surprising, but suggests that our main conclusions regarding conformational distributions and denaturation pathways from solid state NMR measurements in frozen solutions may be relevant to the behavior of HP35 under more conventional conditions, although the precise populations of extended and PPII conformations may vary with solvent conditions and temperature. The increasingly positive signal at 218 nm in the CD spectrum of denatured HP35 at lower temperatures (Figure 9b) suggests an increase in PPII population^{74,75}. The increasingly negative CD signal at 222 nm in the folded state at lower temperatures suggests a stabilization of the helical secondary structure. Additional experiments will be required to assess definitively the relevance of measurements in glycerol/water at low temperatures to conformational preferences under more conventional conditions.

Standard treatments of single-domain protein folding thermodynamics lead to a non-monotonic temperature dependence of ΔG_0 , with negative values both above the unfolding temperature (approximately 70° C for HP35) and below a “cold denaturation” temperature, attributable to a positive change in heat capacity ΔC_p upon unfolding. Calorimetric data for HP35 indicate $\Delta C_p \approx 0.6$ kcal/mole-K near the unfolding temperature⁸². If ΔG_0 were decreasing towards zero with decreasing temperature, one would also expect $[\text{GdnHCl}]_{\text{mid}}$ to decrease. At this point, the temperature-independence of $[\text{GdnHCl}]_{\text{mid}}$ at low temperatures (to within $\pm 8\%$ from -75° C to 22° C) is mysterious, but may be related to the non-two-state nature of HP35 denaturation, temperature-dependent stabilization of the folded state by glycerol, temperature-dependent denaturing strength of GdnHCl, or a fortuitous combination of these factors.

Raleigh and coworkers have examined the unfolded state of HP36 through solution NMR and CD studies of shorter peptides^{38,83}. They report that a 21-residue peptide that includes the N-terminal and central α -helices of HP36 has significant population of helical secondary structure and tertiary interactions similar to those in the folded state of the full-length protein. From these data, they suggest that the conformational distribution of unfolded HP36 (and presumably also HP35) may have significant population of native-like structure. Related conclusions were reached by Pande and coworkers from their analysis of molecular dynamics simulations on HP36⁵³. Molecular dynamics simulations by Lei and Duan led to the identification a partially

folded state of HP35 in which the N-terminal and central α -helices were present as an off-pathway folding intermediate, and a state in which the central and C-terminal α -helices were present as an on-pathway intermediate⁴⁴. These experimental and computational studies do not involve chemical denaturation, and do not reveal preferential population of PPII and extended conformations.

Finally, we emphasize that this work represents the first quantitative determination of site-specific conformational distributions in partially and fully unfolded states of a protein by solid state NMR. The experimental techniques and approaches to data analysis described below are likely to have applications in future studies of other biopolymer systems, including studies of folding intermediates that are kinetically trapped by rapid freeze-quenching^{84,85}.

Materials and Methods

Protein synthesis and sample preparation

HP35 samples were prepared by automated solid-phase synthesis on an Applied Biosystems 433A peptide synthesizer using fluoroenylmethoxycarbonyl (Fmoc) chemistry with O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate/1-hydroxy-7-azabenzotriazole (HBTU/HOAt) activation and a Fmoc-Phe Wang resin (0.27 mmol/g substitution, Novabiochem), cleaved with trifluoroacetic acid/water (20:1 v/v) containing phenol, thioanisole, and 1,2-ethanedithiol as precipitated in cold t-butyl methyl ether, and purified by high-performance liquid chromatography (water/acetonitrile gradient, C18 reverse-phase column). Syntheses were carried out at the 0.1 mmol scale, using a ten-fold excess of unlabeled amino acids and a three-fold excess of isotopically labeled amino acids. Capping with acetic anhydride was performed after each coupling step. Protein purity was verified to be at least 95% by electrospray mass spectrometry. Samples for solid state NMR were prepared by dissolving approximately 5-7 mg of HP35 in 250 μ l of glycerol/water (1:1 v/v) with 1 mM sodium acetate, 10 mM CuSO₄, and pH 5. CuSO₄ was included to reduce proton NMR spin-lattice relaxation times to approximately 1 s in the frozen solutions, as described previously.^{23,32,61} The folding state was controlled by addition of GdnHCl to concentrations between 0.0 M and 7.0 M. Values of [GdnHCl] for solid state NMR samples have a precision of ± 0.1 M, while those in CD measurements have a precision of ± 0.02 M. Approximately 240 μ l of the HP35 solution was transferred into a 6 mm Varian MAS rotor and frozen by immersion in liquid nitrogen before solid state NMR measurements.

NMR experiments

NMR data were obtained at a ¹³C NMR frequency of 100.8 MHz (9.4 T magnetic field), using a Varian Infinity spectrometer with a Varian 6 mm MAS probe operating in two-channel (¹H, ¹³C) mode. The probe was precooled to a sample temperature of -120° C and the MAS axis was adjusted by observation of ⁷⁹Br NMR signals from KBr powder at low temperature prior to loading the frozen HP35 sample into the probe.

2D ¹³C-¹³C NMR spectra of HP35-U-AGMFPL and HP35-U-VAL (Figure 2) were recorded at MAS frequencies of 6.7 kHz or 7.0 kHz, using an rf pulse sequence that included ramped-amplitude ¹H-¹³C cross-polarization, two-pulse phase modulated (TPPM) ¹H decoupling⁸⁶ during t₁, t₂, and mixing periods, and radio-frequency-driven recoupling (RFDR) during the 1.71-1.79 ms mixing period^{59,87}. ¹H decoupling field amplitudes were 70-80 kHz. ¹³C π pulse lengths during RFDR were 10-30 μ s, with the carrier frequency near 50 ppm. To minimize artifacts due to out-of-phase crosspeaks among MAS sidebands, the mixing period was synchronized with MAS and data were processed as described previously^{32,69}. ¹³C chemical shifts are reported with respect to tetramethylsilane, based on an external adamantane reference at 38.56 ppm.

2DEXMAS measurements^{57,58} were carried out at MAS frequencies of 2.50 kHz and 2.75 kHz, with the latter frequency chosen to prevent overlap of MAS sideband signals of GdnHCl (natural abundance ^{13}C) and carbonyl ^{13}C labels. The mixing period τ_{mix} was 0.5 s, with minor variations for rotor synchronization. 2D data sets with two different modes of rotor synchronization (SI and SII in Figure 3a) were combined and processed as previously described⁶⁹, yielding a single 2D spectrum in which intersite crosspeaks among MAS sidebands of labeled carbonyl sites arise from spin polarization transfer between the two carbonyl ^{13}C labels. TPPM decoupling was applied in t_1 and t_2 periods with a 93 kHz ^1H rf amplitude. Cross-polarization periods were approximately 1 ms, with a 56 kHz ^1H rf amplitude and a ramped ^{13}C rf amplitude from 50 kHz to 58 kHz. ^{13}C $\pi/2$ pulse lengths were 5.0 μs . The t_1 period ranged from 0.1 μs to 3175.1 μs in 25.0 μs increments. Total acquisition times for 2DEXMAS data were approximately 22 hr, with a recycle delay of 2 s.

CTDQFD measurements⁵⁶ were carried out at an MAS frequency of 4.00 kHz. The RFDR sequences used for excitation and dephasing of DQ coherences were comprised of rotor-synchronized π pulses, with one π pulse centered in each $2\tau_{\text{R}}$ period and with XY-8 phases⁸⁸. The ^{13}C cross-polarization pulse and the first ^{13}C $\pi/2$ pulse were phase-cycled for DQ filtering, selecting ^{13}C NMR signals from labeled carbonyl pairs, suppressing all other signals, and creating a DQ preparation period of $2L\tau_{\text{R}}$ (see Figure 3b). As previously described⁵⁶, the third and fourth ^{13}C $\pi/2$ pulses in the CTDQFD sequence refocus evolution under ^{13}C - ^{13}C magnetic dipole-dipole couplings, producing an effective dipolar evolution period of $2(M-N)\tau_{\text{R}}$. In measurements described above, $L = 32$, $M + N = 56$, and $M - N$ was increased from 0 to 56 in increments of 8. ^{13}C rf amplitudes were 37 kHz for π pulses and 56 kHz for $\pi/2$ pulses. ^1H decoupling amplitudes were 93 kHz. Total acquisition times for CTDQFD data sets were approximately 7 hr.

DQCSA data⁵⁵ were acquired at a 4.00 kHz MAS frequency, with RFDR sequences as described above and DQ preparation and mixing periods of $64\tau_{\text{R}}$. DQ filtering was accomplished by phase cycling of the ^{13}C cross-polarization pulse and the first ^{13}C $\pi/2$ pulse. CSA interactions were recoupled by three π pulses in two rotor periods, as shown in Figure 3c, with 9.0 μs pulse lengths. Total acquisition times for DQCSA data sets were approximately 17 hr.

DQ filtering efficiencies in CTDQFD and DQCSA measurements were 20%. Pulsed spin-lock detection⁸⁹ improved the signal-to-noise of CTDQFD and DQCSA data by reducing the effective carbonyl ^{13}C NMR linewidths from approximately 300 Hz to 25 Hz.

CD measurements

CD spectra were recorded with a Jasco 720 spectropolarimeter, using HP35 concentrations of approximately 60 μM and 2.0 mm path lengths. Low temperatures were achieved by placing the sample cuvette within a home-built insulated box, equipped with quartz windows, inside the sample chamber of the spectropolarimeter. The temperature inside the box was controlled with a continuous flow of cold nitrogen gas and monitored with a standard Pt resistor.

Data analyses

Tables of simulated 2DEXMAS, CTDQFD, and DQCSA data for $-180^\circ \leq \phi \leq 0^\circ$ and $-180^\circ \leq \psi \leq 180^\circ$, with 5° increments in both variables, were generated with Fortran simulation programs written specifically for this purpose (available on request from robertty@mail.nih.gov). CSA tensor anisotropy and asymmetry parameters were set to $\delta = 151$ ppm and $\eta = 0.52$, respectively. These CSA parameters gave a good fit to MAS sideband intensities in a DQ filtered one-dimensional ^{13}C NMR spectrum of HP35-CO-AV at 2.75 kHz MAS frequency and agree well with average parameters determined for helical residues in a

microcrystalline protein by Wylie *et al.*⁹⁰ Variations in δ for backbone carbonyls are small, regardless of secondary structure⁹⁰, and 2DEXMAS, CTDQFD, and DQCSA data are relatively insensitive to variations in η over the range of values found in proteins. Uncertainties in η lead to uncertainties of roughly $\pm 5^\circ$ in best-fit ϕ and ψ values. Standard peptide bond geometries and CSA tensor principal axis directions were assumed^{91,92}. The same simulation programs have been used in previous applications of these techniques to model peptides⁵⁵⁻⁵⁸, amyloid fibrils^{70,71}, peptide/antibody complexes⁶⁰, and other systems^{23,72}. As rf inhomogeneity significantly affects CTDQFD data, simulations of CTDQFD data were averaged over a distribution of rf amplitudes determined from ^{13}C nutation experiments.

Crosspeaks in 2DEXMAS spectra contain both intersite contributions, arising from polarization transfers between ^{13}C -labeled carbonyl sites during τ_{mix} and dependent on the intervening ϕ and ψ angles, and intrasite contributions, arising from ^{14}N spin-lattice relaxation during τ_{mix} and independent of ϕ and ψ ^{57,58}. In the fitting procedures, calculated intrasite contributions were added to simulated 2DEXMAS crosspeak amplitudes with a scaling factor of 0.56 relative to the intersite contributions, based on the known value of τ_{mix} and the value $T_1 \approx 0.6$ s determined previously for peptides in frozen solution²³. Intrasite contributions represented approximately 20% of the total crosspeak amplitudes.

Experimental data were analyzed by comparison with the simulated data tables, using scripts for MATLAB (MathWorks Inc., Natick, MA). Deviations between experimental and simulated

data were represented by the quantities $\chi_q^2 \equiv \sum_{i=1}^{N_q} [E_{q,i} - \lambda_q S_{q,i}]^2 / \sigma_q^2$, where the index q indicates the type of measurement (2DEXMAS, CTDQFD, or DQCSA), N_q is the number of data points in the measurement, $E_{q,i}$ and $S_{q,i}$ are experimental and simulated data, σ_q is the uncertainty in the experimental data (determined from the root-mean-squared noise in the experimental NMR spectra), and λ_q is a scaling factor calculated to minimize χ_q^2 as previously

described⁵⁷. The total deviation was given by $\chi^2 = \sum_{q=1}^3 \chi_q^2$, since the contribution of each measurement to χ^2 was automatically properly weighted through inclusion of the σ_q and λ_q quantities in χ_q^2 . Ideally, a good fit to the experimental data is indicated by the condition

$\chi^2_{\text{min}} \approx (N_{\text{data}} - N_{\text{param}}) \pm \sqrt{2(N_{\text{data}} - N_{\text{param}})}$ where N_{param} is the number of adjustable parameters (corresponding to roughly 90% probability that χ^2_{min} will lie in this range⁹³). Somewhat larger values of χ^2_{min} can occur when systematic errors in experiments or simulations are not negligible compared with σ_q .

Three models for conformational distributions were tested against the experimental 2DEXMAS, CTDQFD, and DQCSA data. In the first and simplest model, the data were fit with a single pair of ϕ and ψ values, representing a narrow conformational distribution at V50 in HP35-CO-AV. In this case, χ_q^2 and $S_{q,i}$ were functions of ϕ and ψ . Values of $S_{q,i}(\phi, \psi)$ were taken directly from the simulated data tables. The results of the analyses are displayed as contour plots of $\chi_q^2(\phi, \psi)$ and $\chi^2(\phi, \psi)$ in Figure 5.

In the second model, conformational distributions were approximated by a sum of symmetric Gaussian components, *i.e.*, a total probability distribution with the form

$P(\phi, \Psi) = \sum_{k=1}^{N_G} h_k e^{-[(\phi - \phi_k)^2 + (\Psi - \Psi_k)^2] / w_k^2}$, where N_G is the number of Gaussian components and h_k , w_k , ϕ_k , and ψ_k are the height, width, central ϕ value and central ψ value of the k^{th} component. With this form, the volume of each component (representing the population of conformations near ϕ_k, ψ_k) was $\pi h_k w_k^2$. In this case, χ_q^2 and $S_{q,i}$ were functions of h_k , w_k , ϕ_k , and ψ_k , with $S_{q,i} \sum_m \sum_n P(\phi_m, \Psi_n) \tilde{S}_{q,i}(\phi_m, \Psi_n)$, where $\tilde{S}_{q,i}(\phi_m, \Psi_n)$ were derived

from the simulated data tables, using bilinear interpolation to permit ϕ_m and ψ_n values in increments of 2° . Gaussian functions in $P(\phi, \psi)$ were truncated at $1/e^4$.

In the third model, conformational distributions at V50 were approximated by a large number N_P of equally-populated points in the ϕ, ψ plane. In this case, χ_q^2 and $S_{q,i}$ were functions of

N_P and the ϕ and ψ values of each point, with $S_{q,i} = \frac{1}{N_P} \sum_{n=1}^{N_P} \tilde{S}_{q,i}(\phi_n, \psi_n)$. Values of ϕ_n and ψ_n were chosen randomly during the fitting procedure, as described below, and $\tilde{S}_{q,i}(\phi_n, \psi_n)$ were determined by bilinear interpolation from the data tables.

Parameters that minimize χ^2 were found by simulated annealing. For the Gaussian model, initial values of h_k , w_k , ϕ_k , and ψ_k for a given N_G were chosen randomly. In each attempted move of the simulated annealing algorithm, h_k , w_k , or both ϕ_k and ψ_k were changed by random quantities with maximum values Δh , Δw , $\Delta \phi$, and $\Delta \psi$. An initial unitless temperature T was chosen to give an approximate 50% acceptance probability with large values of $\Delta \phi$, and $\Delta \psi$. Moves were accepted or rejected according to the usual Metropolis criterion⁷³, using χ^2 as the energy. Simultaneously, Δh , Δw , $\Delta \phi$, and $\Delta \psi$ were adjusted to maintain an average acceptance probability of 50%. T was reduced by sequential multiples of 0.85 after approximately 1000 attempted moves. Simulated annealing was terminated when χ^2 converged to a local minimum. Minimizations were repeated 5-10 times for each experimental data set with different initial parameter values. For the multiple-point model, N_P was chosen initially to be 40 and values of ϕ_n and ψ_n were chosen randomly. In each attempted move, the position of one point was changed by $\Delta \phi$, and $\Delta \psi$ (also adjusted to give 50% acceptance probabilities). N_P was also allowed to vary in a small range after approximately 400 attempted moves. The temperature schedule was similar to that used for the Gaussian model.

Uncertainties in the fitted parameters were defined as the variations in these parameters that increased χ^2 by less than $\sqrt{2N_{\text{data}}}$ from its minimum value χ_{min}^2 , where N_{data} was the total number of data points. These variations were determined by MCMC simulations, in which the parameters were varied and the Metropolis criterion was used as described above, but with T adjusted in an iterative fashion to maintain $\chi^2 \approx \chi_{\text{min}}^2 + \sqrt{2N_{\text{data}}}$. Error limits reported in Table 2 are the root-mean-squared excursions of individual parameters from their best-fit values during MCMC simulations.

For Gaussian models, the value of N_G can be justified by evaluation of the F statistic⁹³: $F_{N_G, N_G'} = \{[\chi_{\text{min}}^2(N_G) - \chi_{\text{min}}^2(N_G')] v(N_G')\} / \{\chi_{\text{min}}^2(N_G') [v(N_G) - v(N_G')]\}$, with $\chi_{\text{min}}^2(N_G)$ and $v(N_G)$ being the best-fit χ^2 value and the number of degrees of freedom for a given N_G . Specifically, $v(N_G) = 42, 38$, or 34 for $N_G = 1, 2$, or 3 , respectively. The χ_{min}^2 values from Table 2 imply $F_{1,2} = 5.7, 7.8$, and 2.8 for HP35-CO-AV, HP35-CO-SA, and HP35-CO-LK at $[\text{GdnHCl}] = 7.0$ M, corresponding to 0.999, 0.9999, and 0.96 confidence levels that $N_G = 2$ is a better fit than $N_G = 1$. $F_{1,2} = 2.0, 7.8$, and 9.5 for HP35-CO-AV, HP35-CO-SA, and HP35-COLK at $[\text{GdnHCl}] = 4.5$ M, corresponding to 0.89, 0.9999, and 0.99998 confidence levels that $N_G = 2$ is a better fit than $N_G = 1$. For all data sets at $[\text{GdnHCl}] = 4.5$ M and 7.0 M, $F_{2,3} < 0.06$, corresponding to confidence levels below 0.01 that $N_G = 3$ is a better fit than $N_G = 2$.

Analysis of systematic errors

As pointed out above (see Results), values of χ_{min}^2 in Table 2 are larger than N_{data} , especially for fits to experimental data for HP35-CO-AV, which had the highest signal-to-noise ratio and were therefore most sensitive to systematic errors. One conceivable source of small systematic errors is the assumption that all molecular orientations in the MAS rotor contribute equally to the experimental NMR signals (*i.e.*, that ^1H - ^{13}C cross-polarization efficiencies and spin

relaxation rates are independent of molecular orientation). Under the conditions of our CTDQFD and DQCSA measurements, we have determined the carbonyl ^{13}C transverse relaxation time T_2 to be approximately 50 ms, implying that spin relaxation is too slow to contribute significantly to $\chi^2_{\min}$ or to affect conformational distributions derived from our data. We have investigated whether orientation-dependent ^1H - ^{13}C cross-polarization efficiencies might affect our results by performing simulations of 2DEXMAS, CTDQFD, and DQCSA experiments in which initial ^{13}C spin polarizations were strongly orientation-dependent, but consistent with spinning sideband patterns in experimental one-dimensional ^{13}C MAS NMR spectra. Fits to the simulated data indicate that orientation-dependent spin polarizations do not affect conformational distributions derived from the data significantly, but may increase $\chi^2_{\min}$ by roughly 3-8 units.

Another conceivable source of systematic errors is the assumption in simulations that all carbonyl CSA tensors have identical anisotropy and asymmetry parameters. The work of Wylie *et al.* shows that anisotropy parameters are nearly constant for carbonyl CSA tensors in proteins, but that asymmetry parameters η vary between 0.45 and 0.85, with the largest values being observed in β -sheets.⁹⁰ To examine effects of variations in η , we analyzed our experimental data at $[\text{GdnHCl}] = 4.5 \text{ M}$ and 7.0 M with the two-Gaussian model, using simulated data tables with either $\eta = 0.52$ (as described above) or $\eta = 0.65$, the latter value being representative of non- α -helical conformations that are not in β -sheets. Best-fit populations varied by less than 5%. Best-fit values of ϕ and ψ varied by less than 9° . $\chi^2_{\min}$ values were 2-10 units higher with $\eta = 0.65$ than with $\eta = 0.52$, except in the case of HP35-CO-LK at $[\text{GdnHCl}] = 4.5 \text{ M}$ where $\chi^2_{\min}$ was reduced from 55 to 47. We conclude that the choice of η may affect $\chi^2_{\min}$, but does not affect conformational distributions derived from our data significantly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health.

References

1. Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* 1996;10:27–34. [PubMed: 8566543]
2. Sadqi M, Fushman D, Munoz V. Atom-by-atom analysis of global downhill protein folding. *Nature* 2006;442:317–321. [PubMed: 16799571]
3. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. *Chem. Rev* 2004;104:3607–3622. [PubMed: 15303830]
4. Meier S, Blackledge M, Grzesiek S. Conformational distributions of unfolded polypeptides from novel NMR techniques. *J. Chem. Phys* 2008;128
5. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol* 2007;17:3–14. [PubMed: 17250999]
6. Neri D, Billeter M, Wider G, Wuthrich K. NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* 1992;257:1559–1563. [PubMed: 1523410]
7. Marsh JA, Singh VK, Jia ZC, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: Implications for fibrillation. *Protein Sci* 2006;15:2795–2804. [PubMed: 17088319]
8. Schwarzsinger S, Wright PE, Dyson HJ. Molecular hinges in protein folding: The urea-denatured state of apomyoglobin. *Biochemistry* 2002;41:12681–12686. [PubMed: 12379110]

9. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured α -synuclein. *Proc. Natl. Acad. Sci. U. S. A* 2005;102:1430–1435. [PubMed: 15671169]
10. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. 1. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol* 1997;268:158–169. [PubMed: 9149149]
11. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. 2. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol* 1997;268:170–184. [PubMed: 9149150]
12. Meier S, Grzesiek S, Blackledge M. Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. *J. Am. Chem. Soc* 2007;129:9799–9807. [PubMed: 17636913]
13. Obolensky OI, Schlepckow K, Schwalbe H, Solov'yov AV. Theoretical framework for NMR residual dipolar couplings in unfolded proteins. *J. Biomol. NMR* 2007;39:1–16. [PubMed: 17619170]
14. Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 2001;293:487–489. [PubMed: 11463915]
15. Fieber W, Kristjansdottir S, Poulsen FM. Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol* 2004;339:1191–1199. [PubMed: 15178258]
16. Mukrasch MD, Markwick P, Biernat J, von Bergen M, Bernado P, Griesinger C, Mandelkow E, Zweckstetter M, Blackledge M. Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc* 2007;129:5235–5243. [PubMed: 17385861]
17. Fiebig KM, Schwalbe H, Buck M, Smith LJ, Dobson CM. Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem* 1996;100:2661–2666.
18. Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* 1997;36:8977–8991. [PubMed: 9220986]
19. Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol* 1996;255:494–506. [PubMed: 8568893]
20. McCarney ER, Kohn JE, Plaxco KW. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit. Rev. Biochem. Mol. Biol* 2005;40:181–189. [PubMed: 16126485]
21. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A* 2004;101:12491–12496. [PubMed: 15314214]
22. Brewer SH, Song BB, Raleigh DP, Dyer RB. Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. *Biochemistry* 2007;46:3279–3285. [PubMed: 17305369]
23. Long HW, Tycko R. Biopolymer conformational distributions from solid state NMR: α -helix and 3_{10} -helix contents of a helical peptide. *J. Am. Chem. Soc* 1998;120:7039–7048.
24. van Beek JD, Meier BH. A DOQSY approach for the elucidation of torsion angle distributions in biopolymers: Application to silk. *J. Magn. Reson* 2006;178:106–120. [PubMed: 16243550]
25. van Beek JD, Beaulieu L, Schafer H, Demura M, Asakura T, Meier BH. Solid state NMR determination of the secondary structure of *Samia cynthia ricini* silk. *Nature* 2000;405:1077–1079. [PubMed: 10890452]
26. Utz M. Measurement of structural distribution functions in disordered systems: A general approach for sensitivity estimation. *J. Chem. Phys* 1998;109:6110–6124.
27. Dunbar MG, Novak BM, Schmidt-Rohr K. Trans content in atactic polystyrene estimated by double-quantum solid state NMR. *Solid State Nucl. Magn. Reson* 1998;12:119–137. [PubMed: 9809785]

28. Schmidt-Rohr K, Hu W, Zumbulyadis N. Elucidation of the chain conformation in a glassy polyester, PET, by two-dimensional NMR. *Science* 1998;280:714–717. [PubMed: 9563942]
29. Brewer SH, Vu DM, Tang YF, Li Y, Franzen S, Raleigh DP, Dyer RB. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. U. S. A* 2005;102:16662–16667. [PubMed: 16269546]
30. Cellmer T, Henry ER, Kubelka J, Hofrichter J, Eaton WA. Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. *J. Am. Chem. Soc* 2007;129:14564–+. [PubMed: 17983235]
31. Chiu TK, Kubelka J, Herbst-Irmer R, Eaton WA, Hofrichter J, Davies DR. High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. U. S. A* 2005;102:7517–7522. [PubMed: 15894611]
32. Havlin RH, Tycko R. Probing site-specific conformational distributions in protein folding with solid state NMR. *Proc. Natl. Acad. Sci. U. S. A* 2005;102:3284–3289. [PubMed: 15718283]
33. Kubelka J, Chiu TK, Davies DR, Eaton WA, Hofrichter J. Sub-microsecond protein folding. *J. Mol. Biol* 2006;359:546–553. [PubMed: 16643946]
34. Kubelka J, Eaton WA, Hofrichter J. Experimental tests of villin subdomain folding simulations. *J. Mol. Biol* 2003;329:625–630. [PubMed: 12787664]
35. McKnight CJ, Doering DS, Matsudaira PT, Kim PS. A thermostable 35-residue subdomain within villin headpiece. *J. Mol. Biol* 1996;260:126–134. [PubMed: 8764395]
36. McKnight CJ, Matsudaira PT, Kim PS. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol* 1997;4:180–184. [PubMed: 9164455]
37. Meng JM, Vardar D, Wang YM, Guo HC, Head JF, McKnight CJ. High-resolution crystal structures of villin headpiece and mutants with reduced F-actin binding activity. *Biochemistry* 2005;44:11963–11973. [PubMed: 16142894]
38. Tang YF, Goger MJ, Raleigh DP. NMR characterization of a peptide model provides evidence for significant structure in the unfolded state of the villin headpiece helical subdomain. *Biochemistry* 2006;45:6940–6946. [PubMed: 16734429]
39. Wang MH, Tang YF, Sato SS, Vugmeyster L, McKnight CJ, Raleigh DP. Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. *J. Am. Chem. Soc* 2003;125:6032–6033. [PubMed: 12785814]
40. Wickstrom L, Bi Y, Hornak V, Raleigh DP, Simmerling C. Reconciling the solution and x-ray structures of the villin headpiece helical subdomain: Molecular dynamics simulations and double mutant cycles reveal a stabilizing cation- π interaction. *Biochemistry* 2007;46:3624–3634. [PubMed: 17338549]
41. Yang JS, Wallin S, Shakhnovich EI. Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc. Natl. Acad. Sci. U. S. A* 2008;105:895–900. [PubMed: 18195374]
42. Chakraborty S, Sinha SK, Bandyopadhyay S. Low-frequency vibrational spectrum of water in the hydration layer of a protein: A molecular dynamics simulation study. *J. Phys. Chem. B* 2007;111:13626–13631. [PubMed: 17994720]
43. Lucent D, Vishal V, Pande VS. Protein folding under confinement: A role for solvent. *Proc. Natl. Acad. Sci. U. S. A* 2007;104:10430–10434. [PubMed: 17563390]
44. Lei HX, Duan Y. Two-stage folding of HP35 from ab initio simulations. *J. Mol. Biol* 2007;370:196–206. [PubMed: 17512537]
45. Lei HX, Wu C, Liu HG, Duan Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A* 2007;104:4925–4930. [PubMed: 17360390]
46. Khandogin J, Raleigh DP, Brooks CL. Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network. *J. Am. Chem. Soc* 2007;129:3056–+. [PubMed: 17311386]
47. Jayachandran G, Vishal V, Pande VS. Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys* 2006;124
48. Herges T, Wenzel W. Free-energy landscape of the villin headpiece in an all-atom force field. *Structure* 2005;13:661–668. [PubMed: 15837204]

49. De Mori GMS, Colombo G, Micheletti C. Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins* 2005;58:459–471. [PubMed: 15521059]
50. Shen MY, Freed KF. All-atom fast protein folding simulations: The villin headpiece. *Proteins-Structure Function and Genetics* 2002;49:439–445.
51. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 1998;282:740–744. [PubMed: 9784131]
52. Zagrovic B, Snow CD, Shirts MR, Pande VS. Simulation of folding of a small α -helical protein in atomistic detail using worldwide distributed computing. *J. Mol. Biol* 2002;323:927–937. [PubMed: 12417204]
53. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS. Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol* 2002;323:153–164. [PubMed: 12368107]
54. Vardar D, Buckley DA, Frank BS, McKnight CJ. NMR structure of an F-actin-binding “headpiece” motif from villin. *J. Mol. Biol* 1999;294:1299–1310. [PubMed: 10600386]
55. Blanco FJ, Tycko R. Determination of polypeptide backbone dihedral angles in solid state NMR by double quantum ^{13}C chemical shift anisotropy measurements. *J. Magn. Reson* 2001;149:131–138.
56. Bennett AE, Weliky DP, Tycko R. Quantitative conformational measurements in solid state NMR by constant-time homonuclear dipolar recoupling. *J. Am. Chem. Soc* 1998;120:4897–4898.
57. Tycko R, Weliky DP, Berger AE. Investigation of molecular structure in solids by two-dimensional NMR exchange spectroscopy with magic-angle spinning. *J. Chem. Phys* 1996;105:7915–7930.
58. Weliky DP, Tycko R. Determination of peptide conformations by two-dimensional magic-angle-spinning NMR exchange spectroscopy with rotor synchronization. *J. Am. Chem. Soc* 1996;118:8487–8488.
59. Bennett AE, Rienstra CM, Griffiths JM, Zhen WG, Lansbury PT, Griffin RG. Homonuclear radio-frequency-driven recoupling in rotating solids. *J. Chem. Phys* 1998;108:9463–9479.
60. Weliky DP, Bennett AE, Zvi A, Anglister J, Steinbach PJ, Tycko R. Solid state NMR evidence for an antibody-dependent conformation of the V3 loop of HIV-1 gp120. *Nat. Struct. Biol* 1999;6:141–145. [PubMed: 10048925]
61. Sharpe S, Kessler N, Anglister JA, Yau WM, Tycko R. Solid state NMR yields structural constraints on the V3 loop from HIV-1 gp120 bound to the 447-52d antibody Fv fragment. *J. Am. Chem. Soc* 2004;126:4979–4990. [PubMed: 15080704]
62. Sharpe S, Yau WM, Tycko R. Structure and dynamics of the HIV-1 Vpu transmembrane domain revealed by solid state NMR with magic-angle spinning. *Biochemistry* 2006;45:918–933. [PubMed: 16411768]
63. Wishart DS, Case DA. Use of chemical shifts in macromolecular structure determination. *Methods Enzym* 2001;338:3–34.
64. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. 1. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 1995;5:67–81. [PubMed: 7881273]
65. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol* 1991;222:311–333. [PubMed: 1960729]
66. Spera S, Bax A. Empirical correlation between protein backbone conformation and C_α and C_β ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc* 1991;113:5490–5492.
67. Saito H. Conformation-dependent ^{13}C chemical shifts: A new means of conformational characterization as obtained by high-resolution solid state ^{13}C NMR. *Magn. Reson. Chem* 1986;24:835–852.
68. Schubert M, Labudde D, Oschkinat H, Schmieder P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. *J. Biomol. NMR* 2002;24:149–154. [PubMed: 12495031]
69. Tycko R, Berger AE. Dual processing of two-dimensional exchange data in magic-angle-spinning NMR of solids. *J. Magn. Reson* 1999;141:141–147. [PubMed: 10527751]
70. Antzutkin ON, Balbach JJ, Tycko R. Site-specific identification of non- β -strand conformations in Alzheimer's β -amyloid fibrils by solid state NMR. *Biophys. J* 2003;84:3326–3335. [PubMed: 12719262]

71. Balbach JJ, Ishii Y, Antzutkin ON, Leapman RD, Rizzo NW, Dyda F, Reed J, Tycko R. Amyloid fibril formation by A β 16-22, a seven-residue fragment of the Alzheimer's β -amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* 2000;39:13748–13759. [PubMed: 11076514]
72. Blanco FJ, Hess S, Pannell LK, Rizzo NW, Tycko R. Solid state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J. Mol. Biol* 2001;313:845–859. [PubMed: 11697908]
73. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J. Chem. Phys* 1953;21:1087–1092.
74. Tiffany ML, Krimm S. Extended conformations of polypeptides and proteins in urea and guanidine hydrochloride. *Biopolymers* 1973;12:575–587.
75. Whittington SJ, Chellgren BW, Hermann VM, Creamer TP. Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry* 2005;44:6269–6275. [PubMed: 15835915]
76. Dukor RK, Keiderling TA. Reassessment of the random coil conformation: Vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers* 1991;31:1747–1761. [PubMed: 1793813]
77. Shi ZS, Chen K, Liu ZG, Ng A, Bracken WC, Kallenbach NR. Polyproline I propensities from GGXGG peptides reveal an anticorrelation with β -sheet scales. *Proc. Natl. Acad. Sci. U. S. A* 2005;102:17964–17968. [PubMed: 16330763]
78. Shi ZS, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U. S. A* 2002;99:9190–9195. [PubMed: 12091708]
79. Asher SA, Mikhonin AV, Bykov S. UV Raman demonstrates that α -helical polyalanine peptides melt to polyproline II conformations. *J. Am. Chem. Soc* 2004;126:8433–8440. [PubMed: 15238000]
80. Kentsis A, Mezei M, Gindin T, Osman R. Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins* 2004;55:493–501. [PubMed: 15103613]
81. Mezei M, Fleming PJ, Srinivasan R, Rose GD. Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins* 2004;55:502–507. [PubMed: 15103614]
82. Godoy-Ruiz R, Henry ER, Kubelka J, Hofrichter J, Munoz V, Sanchez-Ruiz JM, Eaton WA. Estimating free energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J. Phys. Chem. B* 2008;112:5938–5949. [PubMed: 18278894]
83. Tang YF, Rigotti DJ, Fairman R, Raleigh DP. Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: The villin headpiece subdomain. *Biochemistry* 2004;43:3264–3272. [PubMed: 15023077]
84. Cherepanov AV, de Vries S. Microsecond freeze-hyperquenching: Development of a new ultrafast micro-mixing and sampling technology and application to enzyme catalysis. *Biochim. Biophys. Acta-Bioenerg* 2004;1656:1–31.
85. Evans JNS, Appleyard RJ, Shuttleworth WA. Detection of an enzyme intermediate complex by time-resolved solid state NMR spectroscopy. *J. Am. Chem. Soc* 1993;115:1588–1590.
86. Bennett AE, Rienstra CM, Auger M, Lakshmi KV, Griffin RG. Heteronuclear decoupling in rotating solids. *J. Chem. Phys* 1995;103:6951–6958.
87. Ishii Y. ^{13}C - ^{13}C dipolar recoupling under very fast magic angle spinning in solid state nuclear magnetic resonance: Applications to distance measurements, spectral assignments, and high-throughput secondary structure determination. *J. Chem. Phys* 2001;114:8473–8483.
88. Gullion T, Baker DB, Conradi MS. New, compensated Carr-Purcell sequences. *J. Magn. Reson* 1990;89:479–484.
89. Petkova AT, Tycko R. Sensitivity enhancement in structural measurements by solid state NMR through pulsed spin locking. *J. Magn. Reson* 2002;155:293–299. [PubMed: 12036340]
90. Wylie BJ, Sperling LJ, Frericks HL, Shah GJ, Franks WT, Rienstra CM. Chemical shift anisotropy measurements of amide and carbonyl resonances in a microcrystalline protein with slow magic-angle-spinning NMR spectroscopy. *J. Am. Chem. Soc* 2007;129:5318–5319. [PubMed: 17425317]
91. Teng Q, Iqbal M, Cross TA. Determination of the ^{13}C chemical-shift and ^{14}N electric field gradient tensor orientations with respect to the molecular frame in a polypeptide. *J. Am. Chem. Soc* 1992;114:5312–5321.

92. Oas TG, Hartzell CJ, McMahon TJ, Drobny GP, Dahlquist FW. The carbonyl ^{13}C chemical shift tensors of five peptides determined from ^{15}N dipole-coupled chemical shift powder patterns. *J. Am. Chem. Soc* 1987;109:5956–5962.
93. Bevington, PR. Data reduction and error analysis for the physical sciences. McGraw-Hill; New York: 1969.

Figure 1.

(a) Structure of folded HP35³¹, showing residues that are uniformly isotopically labeled for 2D spectra in Figure 2. (b) Atomic representation of the A49-V50 segment within the HP35 sequence, with ^{13}C -labeled carbonyl sites in green. Solid state NMR measurements of the ^{13}C - ^{13}C distance and the relative orientation of labeled carbonyl groups serve as constraints on the ϕ and ψ torsion angles of V50. (c) CSA tensor principal axes for a ^{13}C -labeled carbonyl site in a polypeptide.^{91,92} The δ_{33} axis is perpendicular to the peptide plane.

Figure 2. 2D ^{13}C NMR spectra of HP35-U-AGMFPL (a) and HP35-U-VAL (b) in frozen glycerol/water with the indicated GdnHCl concentrations. Spectra were obtained at 100.8 MHz ^{13}C NMR frequency. Chemical shift assignments at $[\text{GdnHCl}] = 0.0$ M and partial assignments at $[\text{GdnHCl}] = 7.0$ M are shown. Strong signals at 72.5 ppm and 63.1 ppm are from natural-abundance ^{13}C of glycerol.

Figure 3.

Rf pulse sequences used in quantitative conformational measurements. (a) 2DEXMAS sequence. Arrows mark time points that are actively synchronized to the MAS rotor phase, making either τ_{mix} and $t_1 + \tau_{\text{mix}}$ multiples of the rotor period in SI and SII, respectively. [Phase cycle: $\theta_0 = 90^\circ, 270^\circ$; $\theta_1^* = 270^\circ$; $\theta_2^* = 0^\circ, 180^\circ$. Asterisks indicate phases to which overall CYCLOPS cycling is also applied. Phase sensitive 2D data were constructed by repeating an acquisition for the same t_1 period with a 90° phase shift of the $\pi/2$ pulse before the mixing period.] (b) CTDQFD sequence, with $\pi/2$ pulses represented by black bars. RFDR periods contain L, M, and N π pulses, with one π pulse per two MAS rotor periods. The double-quantum excitation period is $2L\tau_R$, where τ_R is the rotor period, and the effective evolution period under dipole-dipole couplings is $2(M-N)\tau_R$. [Phase cycle: $\theta_0 = 90^\circ, 270^\circ$; $\theta_1^* = 0^\circ, 0^\circ, 900^\circ, 900^\circ, 180^\circ, 180^\circ, 270^\circ, 270^\circ$; $\theta_2^* = 90^\circ, 90^\circ, 180^\circ, 180^\circ, 270^\circ, 270^\circ, 0^\circ, 0^\circ$; $\theta_3^* = 0^\circ$; $\theta_4^* = 180^\circ$; $\theta_5^* = 270^\circ$; $\theta_6^* = 0^\circ, 180^\circ, 180^\circ, 0^\circ$] (c) DQCSA sequence, with π pulses represented by thicker black bars. The evolution period for double-quantum coherences under the sum of carbonyl CSA interactions is τ . [Phase cycle: $\theta_0 = 90^\circ, 270^\circ$; $\theta_1 = 90^\circ, 90^\circ, 180^\circ, 180^\circ, 270^\circ, 270^\circ, 0^\circ, 0^\circ$; $\theta_2 = 180^\circ, 180^\circ, 270^\circ, 270^\circ, 0^\circ, 0^\circ, 90^\circ, 90^\circ$; $\theta_3^* = 270^\circ$; $\theta_4^* = 0^\circ, 180^\circ, 180^\circ, 0^\circ$]

Figure 4.

2DEXMAS, CTDQFD, and DQCSA data for HP35-CO-AV in frozen glycerol/water with $[\text{GdnHCl}] = 0.0 \text{ M}$. (a) 2DEXMAS spectrum at 2.50 kHz MAS frequency (*left*), showing crosspeaks among MAS sideband lines of A49 and V50 carbonyl sites. Crosspeak volumes are analyzed as constraints on ϕ and ψ of V50. One-dimensional slices through the $m = 0, \pm 1$, and ± 2 sideband positions (*right*) indicate the signal-to-noise. (b) CTDQFD data obtained at 4.00 kHz MAS frequency with $L=32$ and $M+N=56$. (c) DQCSA data, obtained at 4.00 kHz MAS frequency with $L=32$. The dependences of carbonyl ^{13}C NMR signals on $M-N$ in (b) and on τ in (c) are analyzed as constraints on ϕ and ψ .

Figure 5.

Fits of 2DEXMAS, CTDQFD, DQCSA, and combined data for HP35-CO-AV, assuming a single backbone conformation at V50, shown as contour plots of the χ^2 deviations between experimental and simulated data as functions of the ϕ and ψ torsion angles assumed in the simulations. At $[\text{GdnHCl}] = 0.0 \text{ M}$, the χ^2_{min} value for the combined data is 69, at $\phi = -70^\circ$ and $\psi = -45^\circ$. At $[\text{GdnHCl}] = 4.5 \text{ M}$ and 7.0 M , χ^2_{min} values are significantly higher (85 and 128, respectively), indicating the presence of multiple conformations at V50.

Figure 6.

Results of MCMC simulations for HP35-CO-AV in frozen glycerol/water at $[\text{GdnHCl}] = 0.0$ M, 4.5 M, and 7.0 M (a,b,c). These simulations show the regions of the ϕ, ψ plane for V50 that are populated in two-Gaussian conformational distributions that are acceptable fits to the combined 2DEXMAS, CTDQFD, and DQCSA data. Uncertainties in the fitting parameters in Table 2 are extracted from the MCMC simulations.

Figure 7.

Assessment of site-specific variations in conformational distributions. Each panel shows the values of $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min}$ for V50 (blue), A57 (green) and K70 (red) during MCMC simulations that explored the acceptable two-Gaussian conformational distributions for V50 (a,d), A57 (b,e), or K70 (c,f). Conformational distributions at two sites are significantly different if all low- $\Delta\chi^2$ distributions for one site are high $-\Delta\chi^2$ distributions for the other site.

Figure 8.

Results of multiple-point fits to the combined 2DEXMAS, CTDQFD, and DQCSA data for HP35-CO-AV in frozen glycerol/water at [GdnHCl] = 0.0 M, 4.5 M, and 7.0 M (a,b,c). Panels on the left are the best-fit configurations from simulated annealing. Final configurations consist of 50, 50, and 44 populated ϕ, ψ points, with χ^2 values of 62, 66, and 63 in (a), (b), and (c), respectively. Panels on the right are populated regions in MCMC simulations.

Figure 9. Denaturation of HP35 monitored by CD. (a) Dependence of mean molar ellipticity at 222 nm on [GdnHCl] for HP35 in glycerol/water at 22° C (circles, solid lines) and -55° C (squares, dashed lines). Lines are fits to a standard two-state model³². (b) CD spectra at indicated temperatures and GdnHCl concentrations.

Table 1
¹³C NMR chemical shifts (ppm relative to tetramethylsilane) in the folded state of HP35 in frozen glycerol/H₂O solution. Uncertainties are approximately ±0.3 ppm. Shifts in parentheses are random coil values for peptides in fluid aqueous solution, adapted from Wishart *et al.*⁶⁴ Uncertainties are approximately ±0.3 ppm.

residue	CO	C _α	C _β	C _γ	C _δ
A49	176.4 (176.1)	52.5 (50.8)	16.6 (17.4)		
V50	176.2 (174.6)	63.5 (60.5)	30.2 (31.2)	20.5, 17.4 (19.4, 18.6)	
G52	170.1 (173.2)	44.5 (43.4)			
M53	170.0 (174.6)	51.0 (53.7)	32.2 (31.2)	26.3 (30.3)	
A57	178.9 (176.1)	52.6 (50.8)	17.0 (17.4)		
F58	174.8 (174.1)	58.6 (56.0)	37.7 (37.9)	136.7 (137.2)	
P62	174.8 (175.6)	59.9 (61.6)	30.4 (30.4)	26.8 (25.5)	47.0 (48.1)
L69	177.1 (175.9)	55.8 (53.4)	40.4 (40.7)	25.8 (25.2)	20.9 (23.2, 21.6)

Best fits of combined 2DEXMAS, CTDQFD, and DQCSA data with single-point and Gaussian models for ϕ, ψ distributions at V50, A57, and K70 in HP35. Uncertainties in fitting parameters are root-mean-squared deviations from best fit values, extracted from MCMC calculations.

sample	fitting model	$\chi^2_{\min}$	central ϕ, ψ	width	population	central ϕ, ψ	width	population
HP35-CO-AV, [GdnHCl] = 0.0 M	single ϕ, ψ	69	-70±5°, -45±5°	---	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 4.5 M	single ϕ, ψ	85	-75±5°, -40±5°	---	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 7.0 M	single ϕ, ψ	128	-75±5°, -20±5°	---	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 0.0 M	one Gaussian	68	-70±3°, -44±2°	2±2°	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 4.5 M	one Gaussian	81	-77±3°, -41±2°	6±2°	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 7.0 M	one Gaussian	104	-90±6°, -33±5°	41±5°	100%	---	---	---
HP35-CO-AV, [GdnHCl] = 4.5 M	two Gaussians	67	-70±4°, -55±5°	2±2°	53±5%	-101±6°, 76±10°	11±5°	47±5%
HP35-CO-AV, [GdnHCl] = 7.0 M	two Gaussians	65	-116±3°, 75±3°	6±2°	63±3%	-75±2°, 156±4°	6±2°	37±3%
HP35-CO-SA, [GdnHCl] = 0.0 M	single ϕ, ψ	38	-65±5°, -45±5°	---	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 4.5 M	single ϕ, ψ	102	-70±5°, -35±5°	---	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 7.0 M	single ϕ, ψ	109	-90±5°, -15±5°	---	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 0.0 M	one Gaussian	34	-71±2°, -46±2°	5±2°	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 4.5 M	one Gaussian	100	-74±4°, -36±2°	6±2°	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 7.0 M	one Gaussian	100	-92±5°, -28±4°	20±3°	100%	---	---	---
HP35-CO-SA, [GdnHCl] = 4.5 M	two Gaussians	55	-68±7°, -50±7°	10±4°	65±6%	-131±15°, 107±25°	9±6°	35±6%
HP35-CO-SA, [GdnHCl] = 7.0 M	two Gaussians	55	-126±6°, 82±5°	4±3°	59±5%	-78±5°, 160±6°	4±3°	41±5%
HP35-CO-LK, [GdnHCl] = 0.0 M	single ϕ, ψ	53	-70±5°, -45±5°	---	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 4.5 M	single ϕ, ψ	112	-85±5°, -20±5°	---	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 7.0 M	single ϕ, ψ	66	-110±5°, -110±5°	---	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 0.0 M	one Gaussian	53	-70±2°, -45±2°	2±2°	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 4.5 M	one Gaussian	110	-84±2°, -21±2°	5±2°	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 7.0 M	one Gaussian	62	-109±3°, -112±4°	2±2°	100%	---	---	---
HP35-CO-LK, [GdnHCl] = 4.5 M	two Gaussians	55	-145±13°, 72±16°	14±5°	57±6%	-52±12°, -45±12°	20±4°	43±6%
HP35-CO-LK, [GdnHCl] = 7.0 M	two Gaussians	49	-91±8°, 165±8°	8±5°	61±8%	-138±12°, 90±14°	6±5°	39±8%