

Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein

Wenli Meng,¹ Bing Shan,¹ Yuefeng Tang,¹ and Daniel P. Raleigh^{1,2*}

¹Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794-3400

²Graduate Program in Biochemistry and Structural Biology, and Graduate Program in Biophysics, State University of New York at Stony Brook, Stony Brook, New York 11794

Received 6 March 2009; Revised 19 April 2009; Accepted 20 April 2009

DOI: 10.1002/pro.152

Published online 6 May 2009 proteinscience.org

Abstract: The villin headpiece subdomain, HP36, is the smallest naturally occurring protein that folds cooperatively. Its small size, rapid folding, and simple three-helix topology have made it an extremely popular system for computational studies of protein folding. The role of unfolded state structure in rapid folding is an area of active investigation, but relatively little is known about the properties of unfolded states under native conditions. A peptide fragment, HP21, which contains the first and second helices of HP36 has been shown to be a good model for structure in the unfolded state of the intact domain but a detailed description of the conformational propensities of HP21 is lacking and the balance between native and nonnative interactions is not known. A series of three-dimensional NMR experiments were performed on ¹³C, ¹⁵N-labeled HP21 to investigate in detail its conformational propensities. Analysis of ¹³C^α, ¹³C^β, ¹³CO chemical shifts, $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ secondary shifts, the secondary structure propensity scores, NOEs, ¹⁵N *R*₂ values and comparison of experimental chemical shifts with those of HP36 and with chemical shifts calculated using the SHIFTS and SHIFTX programs all indicate that there is significant native like structure in the HP21 ensemble, and thus by implication in the unfolded state of HP36.

Keywords: unfolded state; rapid folding; NMR; chemical shift analysis; villin headpiece

Introduction

A detailed understanding of the process of protein folding is still lacking. Most of the experimental and theoretical effort has focused on characterizing the

transition state for folding and on the analysis of the properties of any intermediates which may be populated. In contrast, relatively little attention has been directed towards the unfolded state, however knowledge of the conformational propensities of the unfolded state is essential for understanding the protein folding process because it is the starting point for folding and can be the starting state for aggregation.^{1–5} The potential role of preformed structure in the unfolded state has received increasing attention, particularly for rapid folding proteins.^{2,6,7} The free energy balance for folding and the normally cooperative nature of folding mean that only a tiny fraction of molecules normally populate the unfolded state under native conditions, and the vast majority of the reported studies of unfolded proteins are performed under strongly denaturing conditions, such as in the presence of denaturant, acid or high temperature. The unfolded state which is populated under native

Additional Supporting Information may be found in the online version of this article.

Abbreviations: DSS, 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt; HP21, 21-residue peptide starting from the N-terminus of HP36; HP36, the villin headpiece helical subdomain corresponding to residues 791–825 of chicken villin and residues 42–76 of the chicken villin headpiece domain with an additional Met at the N-terminus.

Grant sponsor: NSF; Grant number: MCB-614365 (D.P.R.); D.P.R is a member of the New York Structural Biology Center. NMR resources in NYSBC are supported by NIH P41 GM66354.

*Correspondence to: Daniel P. Raleigh, Department of Chemistry, State University of New York at Stony Brook, Stony Brook, NY 11794-3400. E-mail: draleigh@notes.cc.sunysb.edu

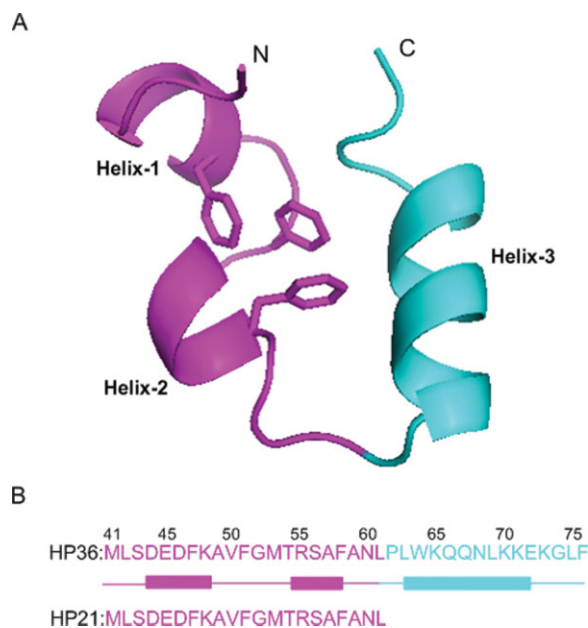


Figure 1. Primary structure of the villin headpiece subdomain HP36. (A) Ribbon diagram of HP36 created using PyMOL (PDB entry 1VII). The fragment corresponding to HP21 is in purple. The three phenylalanines are shown, and the N- and C-termini are labeled. (B) Primary sequence of HP36 and HP21 with α -helices represented by rectangles. The numbering system for the full length villin headpiece subdomain is used, where the helical subdomain starts from residue Leu42. An additional Met at the N-terminus results from the expression of the protein, and this residue is designated as Met41. In the NMR structure (PDB entry 1VII), the three helices run from Asp44 to Lys48, Arg55 to Phe58, and Leu63 to Glu72, and are donated as helix-1, helix-2, and helix-3 respectively. In the X-ray structure,³³ the boundaries of the helices are Asp44 to Phe51, Arg55 to Asn60, and Leu63 to Lys73.

conditions is the unfolded state which is directly relevant for protein folding, and it is clear that the conformational propensities of this unfolded ensemble can be dramatically different from the unfolded state populated under strongly denaturing conditions. However, studies of the unfolded ensemble under native conditions are still rare owing to their difficulty.

The 36-residue villin headpiece helical subdomain HP36 is the smallest naturally occurring protein which folds cooperatively.^{8–11} Its small size and rapid folding have made it an extraordinarily popular system for computational and theoretical studies.^{12–42} The domain is made up of three α -helices and contains a well-packed hydrophobic core made up in part of a cluster of three phenylalanines, Phe47, Phe51, Phe58 (see Fig. 1). The numbering system used here is derived from that used for the intact headpiece. The helical subdomain begins at Leu42. An additional Met results from the expression of the domain and is designated Met41. This notation is widely used in the literature. Peptides which corresponded to each of the individual helices of HP36 have been shown to be largely

unstructured in isolation while a peptide which included the first two helices was found to contain considerable α -helical structure as judged by CD and proton NMR.^{43,44} The peptide, denoted HP21 (HP1/2 in Ref. 43), comprises the first 21 residues of HP36 and includes the three core phenylalanines. HP21 has been shown to be a good model for structure populated in the unfolded state of the intact subdomain.⁴³ Previously reported CD studies of this peptide indicate the presence of some residual structure but CD obviously can not define the location of the helices.⁴³ In addition, the CD signal of short α -helices is much weaker than “normal” α -helical spectra thus it is difficult to directly relate the rotational strength observed in CD spectra of systems such as HP21 to helical content. In contrast, NMR chemical shifts are dependent on local conformation and thus are sensitive reporters of ψ , ϕ preferences. The initial characterization of HP21 provided important information but necessarily could not define the limits of the structured regions and was unable to determine if non-native structure was populated. The purpose of this work is to define the structural propensities of the HP21 ensemble at high resolution, using ^1H , ^{13}C , ^{15}N triple resonance NMR methods.

Results

To obtain a detailed picture of the conformational propensities of HP21 and by implication of the unfolded state of HP36 under native like conditions, we obtained complete $^1\text{H}^\alpha$, $^1\text{H}^\text{N}$, ^{15}N , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and ^{13}CO NMR assignments for HP21. Assignments were also obtained for the native state of HP36 to serve as a basis of comparison since complete ^{13}C assignments of HP36 wild type have not been reported. Figure 2 compares the ^{15}N - ^1H HSQC spectra of HP36 and HP21. The spectrum of HP36 is very well resolved and the resonances are sharp as expected for a well folded small monomeric protein. The spectrum of HP21 is also well dispersed. $^{13}\text{C}^\alpha$ chemical shifts were obtained from HNCA and HN(CO)CA triple resonance experiments, with reference to the previously assigned ^{15}N - ^1H HSQC spectrum.⁴⁴ $^{13}\text{C}^\beta$ chemical shifts were obtained from the CBCA(CO)NH experiment, and ^{13}CO chemical shifts were obtained from a HNCO experiment. The assignments are included in the Supporting Information. Assignments for HP36 were obtained using a similar strategy with reference to the previously HP36 assignments.⁸ They are also included in the Supporting Information. The estimated uncertainties are about 0.02 ppm for ^1H chemical shifts, 0.20 ppm for ^{15}N and ^{13}C chemical shifts.

It is well known that the NMR chemical shifts are sensitive to secondary structure. In particular, the deviations of observed shifts relative to their random coil values ($\Delta\delta = \delta_{\text{observed}} - \delta_{\text{randomcoil}}$), i. e., secondary shifts, are very useful for indentifying secondary structure. Figure 3 displays plots of the $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and ^{13}CO

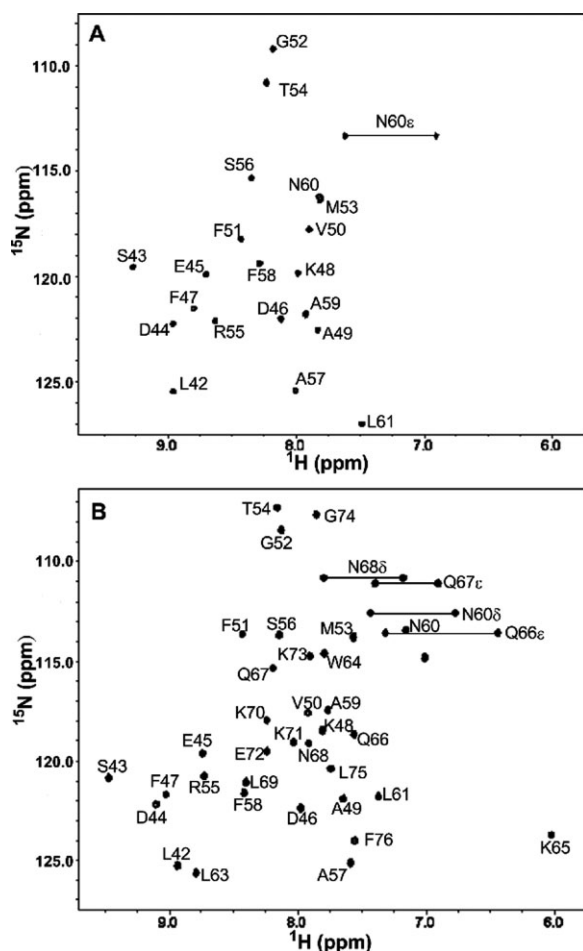


Figure 2. (A) The ^{15}N - ^1H HSQC spectrum of HP21 recorded at pH 5.4 and 12°C . (B) ^{15}N - ^1H HSQC spectrum of HP36 recorded at pH 5.0 and 25°C . The samples contained 150 mM sodium chloride and 10 mM sodium acetate. The resonances are labeled via residue number.

secondary shifts for HP21 and HP36. $^{13}\text{C}^\alpha$ secondary shifts are very sensitive to α -helical structure, and downfield $^{13}\text{C}^\alpha$ shifts often indicate a propensity to populate α -helical conformations. Figure 3(A) shows the $^{13}\text{C}^\alpha$ secondary shifts for HP21. Residues Asp44–Phe51 and Arg55–Ala59 are helical in the X-ray structure of intact HP36.³³ For the regions corresponding to the native helices, residues Asp44–Phe51 and Arg55–Ala59, the $^{13}\text{C}^\alpha$ secondary shifts of HP21 are positive and large with values ranging from 0.97 to 3.71 ppm. The average deviations from random coil values are 2.71 ppm for the region corresponding to helix-1 and 1.88 ppm for helix-2. Figure 3(B) displays the $^{13}\text{C}^\alpha$ secondary shifts for the HP36 native state for comparison. In the native state of HP36, positive $^{13}\text{C}^\alpha$ secondary shifts ranging from 0.14 to 4.26 ppm are detected in the region of Asp44–Phe51 and Arg55–Ala59. The patterns of $^{13}\text{C}^\alpha$ secondary shifts are strikingly similar in HP21 and HP36, suggesting a clear propensity to populate native like α -helical content in the ensemble of HP21.

Like $^{13}\text{C}^\alpha$ secondary shifts, ^{13}CO secondary shifts are expected to exhibit downfield-shifted, (positive), values in α -helices. As shown in Figure 3(C), positive ^{13}CO secondary shifts are observed in the region of Asp44–Phe51 and Arg55–Phe58 for HP21. The values range from 1.01 ppm to 2.40 ppm, with an average secondary shift of 1.67 ppm in helix-1 and 1.45 ppm in helix-2. Ala59 shows a very small negative secondary shift of -0.06 ppm. Figure 3(D) shows the ^{13}CO secondary shifts for HP36. Positive deviations from random coil values are found for residues Asp44–Phe51 and Arg55–Phe58. The deviations range from 1.68 to 3.20 ppm. The ^{13}CO secondary shifts have very similar patterns in HP21 and HP36, including Ala59 which has a negative deviation of -0.09 ppm in HP36.

$^{13}\text{C}^\beta$ secondary shifts also depend on secondary structure, although they are not as sensitive as other secondary shifts to α -helical structure. $^{13}\text{C}^\beta$ chemical shifts relative to random coil values are mostly up-field shifted in α -helices but, they can exhibit down-field shifted values as well.⁴⁶ The $^{13}\text{C}^\beta$ secondary shifts for HP21 and HP36 are displayed in Figure 3(E) and Figure 3(F) respectively. For HP36 residues Asp44–Phe51 and Arg55–Ala59 display $^{13}\text{C}^\beta$ secondary shifts which are generally upfield-shifted with values ranging from -0.62 to -2.42 ppm. A different pattern of deviations is detected in HP21: those residues located in the helical regions in the native state display small downfield-shifted deviations ranging from 0.34 ppm to 1.49 ppm except for Ser56 for which the deviation is -0.39 ppm. However, as noted above, the $^{13}\text{C}^\beta$ secondary shift is not as sensitive as the $^{13}\text{C}^\alpha$, ^{13}CO , or $^1\text{H}^\alpha$ secondary shifts for the identification of helical structure, and previous statistical analysis has shown that a considerable percentage of $^{13}\text{C}^\beta$ secondary shifts in helices show positive deviations from random coil values.

Individual sets of secondary shifts provide useful information about secondary structure preferences, but a more reliable picture can be obtained from analyzing combinations of shifts. A common method is to use the difference in $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ shifts, $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$.⁴⁷ One practical advantage of this approach is that it cancels out any potential uncertainty in chemical shift referencing. Positive $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values are an indication of α -helical structure, and negative values indicate a propensity to populate β -strand structure. $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values for HP21 and HP36 are shown in Figures 4(A,B) respectively. For HP21 in the region of Asp44–Phe51 and Arg55–Ala59, $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values range from -0.07 to 2.86 ppm with an average value of 1.85 ppm for helix-1 and 1.22 ppm for helix-2. The same regions in HP36 display values ranging from 2.31 to 6.04 ppm with an average value of 4.18 ppm for helix-1 and 3.76 ppm for helix-2. The pattern of $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values for HP21 and HP36 are strikingly similar to each other, including negative values for the two residues located N-terminal to helix-1 and helix-2. This provides additional evidence

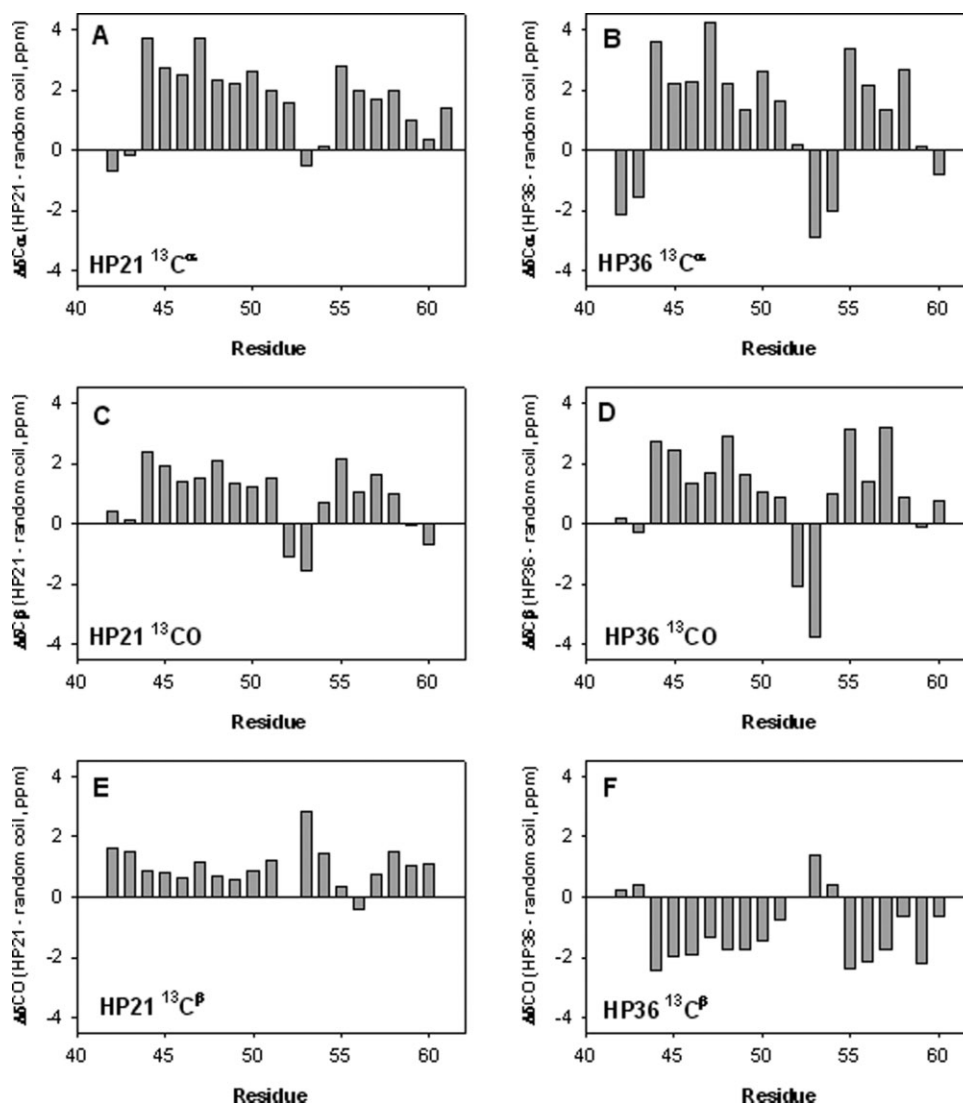


Figure 3. Deviation of the ^{13}C chemical shifts of HP21 and HP36 from random coil values. Data are plotted as observed – random coil. (A) HP21 $^{13}\text{C}^\alpha$, (B) HP36 $^{13}\text{C}^\alpha$, (C) HP21 $^{13}\text{C}^\text{CO}$, (D) HP36 $^{13}\text{C}^\text{CO}$, (E) HP21 $^{13}\text{C}^\beta$, and (F) HP36 $^{13}\text{C}^\beta$. Data were collected at 12°C and pH 5.4 for HP21, and 25°C and pH 5.0 for HP36. Both samples contain 10 mM sodium acetate and 150 mM sodium chloride. Random coil values of Wishart *et al.* were used.⁴⁵

that native like α -helical structure is populated in the ensemble of HP21.

Another very useful method to quantify and compare secondary structure propensities is the secondary structure propensity (SSP) score developed by Forman-Kay and coworkers for application to natively disordered sequences.⁴⁷ This method combines different chemical shifts to generate a single SSP score for each residue. An SSP score of 1 is interpreted to indicate fully formed α -helical structure and a value of -1 is assumed to represent fully formed β -strand structure. SSP scores for HP21 and HP36 calculated from $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$, and $^{13}\text{C}^\beta$ chemical shifts are shown in Figure 5. Most of the residues in the helical regions of HP36 show positive SSP scores near 1. The pattern of SSP scores is very similar for HP21 and HP36, although the values calculated for HP21 are smaller. The SSP scores for the sequence Asp44–Phe51 in HP21 range

from 0.07 to 0.55, with average value of 0.35 for helix-1. In the region corresponds to residues Arg55–Ala59, the SSP scores range from 0.12 to 0.27 with an average value of 0.18 for helix-2, while the SSP scores are much closer to 0 for residues which are not helical in the native state. The SSP analysis is consistent with the analysis of the $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values, and further confirms that the HP21 ensemble contains native like α -helical structure. SSP scores are sometimes calculated using only $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts. SSP scores based on just the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts are very similar to the SSP scores calculated from the combination of $^{13}\text{C}^\alpha$, $^1\text{H}^\alpha$ and $^{13}\text{C}^\beta$ shifts (Supporting Information).

There is a strong correlation between the observed SSP scores of HP21 and HP36 as well as a significant correlation between the respective values of $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ (see Fig. 6). Both plots are linear

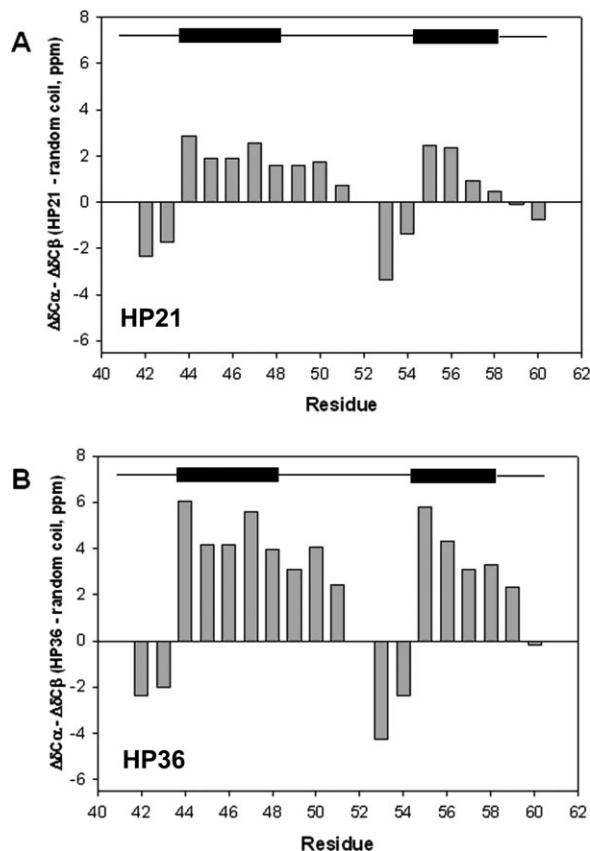


Figure 4. Plots of $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ values for (A) HP21 and (B) HP36. A schematic diagram of the boundaries of the helices in the NMR structure of HP36 (PDB entry 1VII) is shown at the top. Positive values indicate α -helical propensity and negative values indicate β -strand structure propensity.

with $r^2 = 0.96$ for $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ and $r^2 = 0.78$ for the SSP scores. As expected the slopes are less than one consistent with HP21 populating native like secondary structure a fraction of the time. The observed

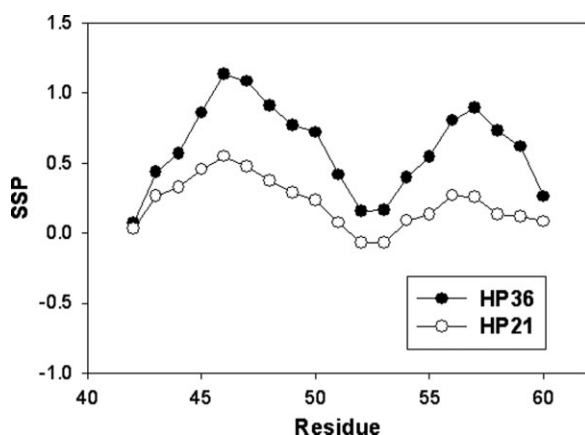


Figure 5. Secondary structure propensity (SSP) analysis of HP21 (○) and first 21 residues of HP36 (●) using the program developed by Forman-Kay and coworkers.⁴⁷ Positive values represent α -helical propensity and negative values indicate β -strand propensity. The calculations were based on $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^1\text{H}^\alpha$ chemical shifts.

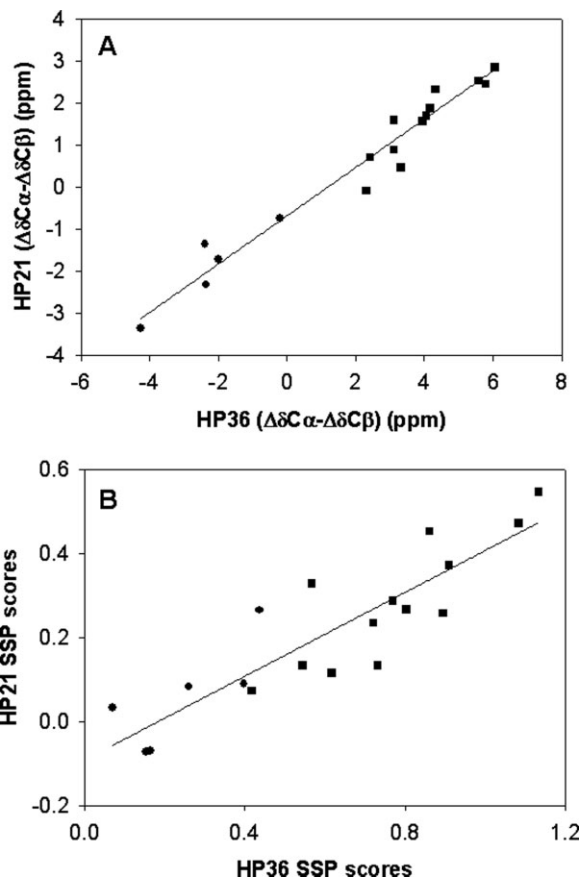


Figure 6. (A) Comparison of $\Delta\delta^{13}\text{C}^\alpha - \Delta\delta^{13}\text{C}^\beta$ for HP36 and HP21. The value of r^2 is 0.96. (B) Comparison of the SSP scores for HP36 and HP21. The value of r^2 is 0.78. Residues which are helical in the native state are indicated as ■, while residues located in nonhelical regions in the native state are indicated as ●.

correlation strongly suggest a tendency for HP21 to adopt native like secondary structure. This is reinforced by comparison of the experimental shifts of HP21 to those calculated from a “native like” model generated by taking the first 21 residues of the high resolution crystal structure of HP36. The observed shifts for HP21 are compatible with a strong tendency for the peptide to adopt native like structure as judged by comparison of the experimental shifts to those expected if HP21 adopts the same structure in isolation as it does in the intact domain (Supporting Information).

We supplemented the chemical shift data by conducting NOE measurements. A 3D ^{15}N - ^1H NOESY-HSQC experiment was performed to detect sequential NOEs in HP21. The data is summarized in Figure 7. Sequential H^{N} to H^{N} NOEs were observed in the helical region from Asp44 to Phe51 and Arg55 to Ala59. In general the NOEs for residues in the central region of helices have larger intensities than those for residues at the boundaries of helices. Interestingly, sequential H^{N} to H^{N} NOEs were also observed in the turn region from Gly52 to Thr54, where the adjacent

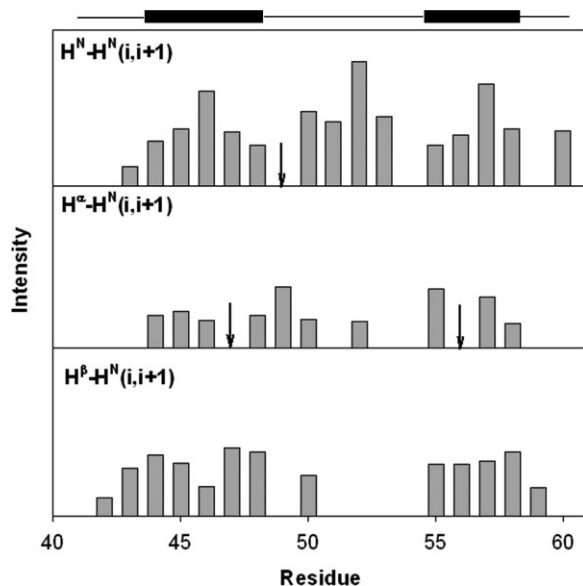


Figure 7. Summary of sequential NOEs observed for HP21 including $H^N-H^N(i,i+1)$, $H^\alpha-H^N(i,i+1)$ and $H^\beta-H^N(i,i+1)$ obtained from a three-dimensional NOESY-HSQC spectrum of HP21. “↓” indicates that the corresponding NOEs were observed but their intensities could not be obtained accurately due to peak overlap. A schematic diagram of the boundaries of the helices in the NMR structure of HP36 (PDB entry 1VII) is shown at the top.

amide proton distances are within 5 Å in the X-ray structure.³³ Those sequential H^N to H^N NOEs were also observed in the native state of HP36.⁹ In addition, a series of sequential H^α to H^N NOEs and H^β to H^N NOEs were observed in the two helical regions of HP21, while few such NOEs were observed for the nonhelical residues. The observed sequential H^N to H^N NOEs in HP21 and their comparison with those in native state HP36 strongly suggest that native like secondary structure is populated in HP21.

^{15}N transverse relaxation rates R_2 are frequently used to study the unfolded state of proteins because they are strongly influenced by local motions. A phenomenological model was developed by Schwalbe and coworkers to describe the R_2 values of random coil polypeptide.^{48,49} In this model, an intrinsic relaxation rate R_{int} of an amide ^{15}N nucleus and an apparent persistence length λ of the peptide chain are used to fit the experimental relaxation data. Completely unstructured peptides are expected to produce relatively uniform R_2 values. Deviations from the predicted values often indicate the presence of hydrophobic clusters. Figure 8 shows the plot of the measured R_2 values for HP21 and the values calculated using the random coil model excluding the two obvious outliers Phe51 and Thr54. The calculated R_2 values are relatively uniform with an average value of 3.98 s⁻¹. Most residues exhibit relatively uniform experimental values near the calculated random coil values, but several residues have R_2 values which are obviously large. These

include Leu42, Ser43, Phe51, Met53, Thr54, and Ala59. The observed deviations are significantly larger than the uncertainty in the experimental values. This suggests that these residues are involved in either local or long range contacts. The same R_2 values were detected for 130 μM and 65 μM protein samples, strongly arguing that the effects are not due to self association, especially as ultracentrifugation studies at 150 μM give an apparent molecular weight which is in excellent agreement with true molecular weight.⁴³ In the native state of HP36, several NOEs were observed involving the residues, which display large R_2 values in HP21. These include NOEs between Phe51 and Leu42, Phe51 and Met53, Met53 and Thr54.⁹ NOEs between the aromatic protons of Phe51 and the δ protons of Leu42 were also observed in the 2D 1H - 1H NOESY spectrum of HP21,⁴⁴ indicating the Leu42 and Phe51 may be involved in the native like hydrophobic clustering in HP21.

Discussion

The extensive chemical shift data reported here together with the calculations are consistent with a strong propensity to preferentially adopt native like backbone dihedral angles ϕ and ψ . It is striking how similar the details of the patterns of $^{13}C^\alpha$, $^{13}C^\beta$ shifts, $\Delta\delta^{13}C^\alpha - \Delta\delta^{13}C^\beta$ and SSP scores are for HP21 and HP36. Values for individual residues in HP21 are smaller in magnitude than those observed for HP36 but display virtually identical patterns. Consider for example, the plots of $\Delta\delta^{13}C^\alpha - \Delta\delta^{13}C^\beta$ displayed in Figure 4. The first two residues before each α -helix display negative values with the residue immediately adjacent to the helix exhibiting smaller values. Exactly the same relative pattern is observed for HP21 and

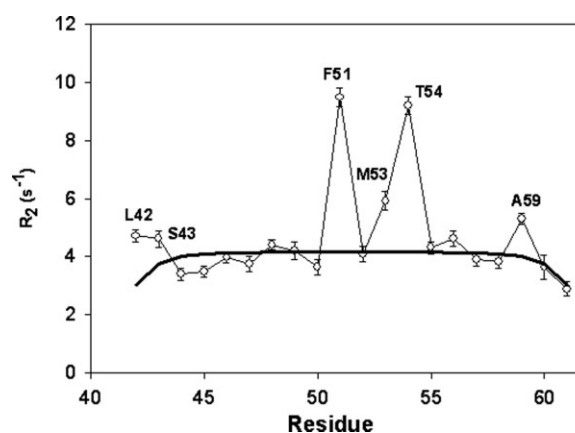


Figure 8. A plot of ^{15}N R_2 values for HP21 (○). The error bars were obtained from the nonlinear fitting of the experimental data with NMRViewJ. The solid line is the best fit to the phenomenological model of Schwalbe and coworkers.⁴⁸ Experiments were conducted at 12°C, pH5.4 in 10 mM sodium acetate, 150 mM sodium chloride. Same results were obtained using 130 μM and 65 μM samples.

HP36. Furthermore the residues which display the largest positive deviations in HP36 do so in HP21 with the same rank order. The agreement in the pattern of SSP scores is also strikingly similar. The excellent agreement between the observed shifts of HP21 and those calculated for the native model reinforces these conclusions. These observations are supported by the NOE experiments which reveal an extensive set of native like backbone NOEs. Solid state NMR studies of the villin headpiece at low temperature in mixed water glycerol frozen glasses in the presence of denaturant suggest a propensity to sample helix structure in helix-2.⁵⁰ However, it is very difficult to compare that study to the present work which was conducted under native conditions at 12°C.

As previously noted the helical structure found in HP21 is not due to an exceptionally high propensity for the individual helices to fold in isolation. The helices need to be stabilized by tertiary interactions and the R_2 relaxation analysis identifies several likely candidates that could contribute, such as Leu42, Phe51, Met53, Thr54, and Ala59. At least some of the side-chains of these residues are close in the folded structure, including Phe51 and Leu42 as well as Phe51 and Met53.

Studies of several other small helical proteins have shown that preorganized secondary structure in the unfolded state plays an important role in the fast folding process.^{51–53} The propensity for native like backbone structure in the ensemble of HP21 observed here gives rise to the possibility that it could play a role in the fast folding of HP36 by restricting the initial conformational search. However if the free energy barrier is dominated by the need to form the third helix and dock it onto the HP21 motif then mutations which alter the denatured state structure in the region corresponding to HP21 need not significantly affect the folding rate. This scenario can arise if the structure is almost as well formed in the unfolded state as in the transition state. In this case mutations which alter the structure captured by our HP21 model would have little effect on the folding rate since the free energy level of the transition state and unfolded state would be perturbed by similar amount. This might offer an explanation for why a double Phe to Leu mutant of HP36 folds at virtually the same rate as wild type even though the mutations significantly reduce the structure in the unfolded state as judged by its effect on HP21.^{14,43}

Finally we note that the presence of significant amount of structure in the unfolded state of HP36 can complicate Φ -value analysis of the transition state for its folding. Φ -values, defined as $\Phi = \Delta\Delta G^{\text{TS}}/\Delta\Delta G^\circ$ where $\Delta\Delta G^{\text{TS}}$ is the change in activation free energy and $\Delta\Delta G^\circ$ is the change in stability, are traditionally viewed as representing the development of native like structure in the transition state. This quasi-structural analysis is valid when mutations do not perturb the

free energy of the unfolded state. In contrast, small Φ -values can arise for mutations which perturb the unfolded state even though they may have significant effects on the transition state.⁵⁴ Thus potential unfolded state effects should be considered when interpreting mutational studies of the folding HP36, particularly as the measured Φ -values for HP36 are all small.

Materials and Methods

Peptide/protein expression and purification

HP21 enriched with ^{15}N , HP21 enriched with ^{15}N and ^{13}C , and HP36 enriched with ^{15}N and ^{13}C were prepared recombinantly as fusion peptides with the N-terminal domain of L9 (NTL9) in M9 minimal medium as described previously.^{44,55} The M9 minimal medium contains 0.8 g/L $^{15}\text{NH}_4\text{Cl}$ as the nitrogen source and 10 g/L ^{13}C -glucose as the carbon source. The proteins were purified as previously described.⁵⁵ 3 mg pure HP21 was obtained from 1 L culture and 10 mg pure HP36 was obtained from 1 L culture.

NMR sample preparation

HP21 and HP36 samples for NMR experiments were prepared in 90% $\text{H}_2\text{O}/10\%$ $^2\text{H}_2\text{O}$ with 10 mM sodium acetate and 150 mM sodium chloride (pH 5.4 for HP21 and pH 5.0 for HP36). The HP21 sample was prepared at a concentration of ~ 100 μM to avoid self-association.⁴³ The concentration of HP36 was ~ 1 mM. 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) was added as an internal reference (0.00 ppm) for both samples.

NMR spectroscopy

All NMR experiments on HP21 were conducted on a 500 MHz Bruker spectrometer with a cryoprobe at the New York Structural Biology Center at 12°C, and all NMR spectra for HP36 were recorded on 800 MHz Bruker spectrometer at 25°C. The ^1H offset frequency was set at the water resonance and the ^{15}N offset frequency was set at 118.0 ppm for all experiments.

The ^{15}N - ^1H HSQC spectrum of ^{13}C , ^{15}N -labeled HP21 was collected with a matrix size of 1024 (^1H) \times 256 (^{15}N), and the spectral widths were 5000 Hz (^1H) \times 1115 Hz (^{15}N). The following triple-resonance experiments were conducted to generate ^{13}C assignments: HNCA, HN(CO)CA, HNC0, and CBCA(CO)NH. All four of the triple-resonance experiments were collected with a matrix size of 1024 (^1H) \times 30 (^{15}N) \times 40 (^{13}C). For the HNCA experiment, the ^{13}C offset frequency was set at 54 ppm and the spectral widths were 5000 Hz (^1H) \times 1115 Hz (^{15}N) \times 2767 Hz (^{13}C). For the HN(CO)CA experiment, the ^{13}C offset frequency was also set at 54 ppm and the spectral widths were 7003 Hz (^1H) \times 1774 Hz (^{15}N) \times 4024 Hz (^{13}C). For the HNC0 experiment, the ^{13}C offset frequency was set at

176 ppm and the spectrum was collected with spectral widths of 5000 Hz (^1H) \times 1115 Hz (^{15}N) \times 1887 Hz (^{13}C). For the CBCA(CO)NH experiment, the ^{13}C offset frequency was 39 ppm and the spectral widths were 5000 Hz (^1H) \times 1115 Hz (^{15}N) \times 8177 Hz (^{13}C). The chemical shift assignments for the $^1\text{H}^\alpha$, $^1\text{H}^\beta$ and backbone ^{15}N resonances have been reported previously.⁴⁴ A 3D ^{15}N - ^1H NOESY-HSQC spectrum of ^{15}N -labeled HP21 was collected with a matrix size of 1024 (^1H) \times 22 (^{15}N) \times 128 (^1H), and the spectral widths were 5000 Hz (^1H) \times 1115 Hz (^{15}N) \times 5001 (^1H); a mixing time of 300 ms was used.

For ^{13}C , ^{15}N -labeled HP36, a ^{15}N - ^1H HSQC spectrum was collected with a matrix size of 1024 (^1H) \times 256 (^{15}N), and the spectral widths were 9615 Hz (^1H) \times 2433 Hz (^{15}N). The following triple-resonance experiments were conducted to generate ^{13}C assignments: HNCO, CBCA(CO)NH, and (H)CC(CO)NH. A HNCO spectrum was collected with a matrix size of 1024 (^1H) \times 40 (^{15}N) \times 64 (^{13}C) and spectral widths of 9615 Hz (^1H) \times 2433 Hz (^{15}N) \times 5031 Hz (^{13}C). CBCA(CO)NH was collected with a matrix size of 1024 (^1H) \times 40 (^{15}N) \times 128 (^{13}C) and spectral widths 9615 Hz (^1H) \times 2433 Hz (^{15}N) \times 13,078 Hz (^{13}C). (H)CC(CO)NH was collected with a matrix size of 1024 (^1H) \times 40 (^{15}N) \times 128 (^{13}C) and spectral widths 9615 Hz (^1H) \times 2433 Hz (^{15}N) \times 15090 Hz (^{13}C). A HBHACONH experiment was conducted to generate $^1\text{H}^\alpha$ and $^1\text{H}^\beta$ chemical shifts. The HBHACONH spectrum was collected with a matrix size of 1024 (^1H) \times 40 (^{15}N) \times 128 (^1H) and spectral widths 9615 Hz (^1H) \times 2433 Hz (^{15}N) \times 5601 Hz (^1H).

The R_2 relaxation experiments conducted with HP21 were carried out as described.⁵⁶ The R_2 relaxation time were set to 16.64, 33.28, 49.92, 66.56, 83.20, 99.84, 116.48, 133.12, and 149.76 ms. The spectra were collected with a matrix size of 1024 (^1H) \times 256 (^{15}N) and spectral widths 5000 Hz (^1H) \times 1267 Hz (^{15}N). A recycle delay of 3 s was used. Experiments were conducted at two protein concentrations, 130 μM and 65 μM to test for any potential self association. The R_2 values were the same for both samples.

Data processing and analysis

Data were processed with the NMRPipe software package developed by Delaglio and coworkers, and the chemical shifts were analyzed with the NMRViewJ software package developed by Johnson and coworkers.^{57,58} ^1H chemical shifts were referenced to DSS directly, ^{15}N , ^{13}C chemical shifts were indirectly referenced using standard methods. The random coil values of Wishart *et al.* were used to calculate the secondary shifts.⁴⁵ The SSP program developed by Forman-Kay and coworkers (<http://pound.med.utoronto.ca/software.html>) was used to calculate SSP scores for HP21 and HP36.⁴⁷ Structure based calculations of chemical shifts for HP21 and HP36 were obtained

using the SHIFTS program of Xu and Case (<http://casegroup.rutgers.edu/qshifts/qshifts.htm>),^{59,60} and the SHIFTX program of Wishart and coworkers (<http://redpoll.pharmacy.ualberta.ca/shiftx/>).⁶¹ The “native model” of HP21 was generated by taking residues 41–61 of the X-ray structure determined by Bi *et al.*³³ This is a structure of a K70M mutant of HP36. The file used for the calculations includes an *in silico* mutation of the Met back to Lys: Met70 was replaced by a Lys with χ_1 , χ_2 set to the same values as those for Met70, and χ_3 , χ_4 set to 180°. ^{15}N transverse relaxation rates R_2 were calculated with NMRViewJ by fitting the peak intensities to two-parameter mono-exponential decay shown in Eq. (1):

$$I(t) = I_0 \exp(-t/T_2) \quad (1)$$

where $I(t)$ is the peak intensity after a delay of time t and I_0 is the intensity at the time point $t = 0$. The R_2 rates were analyzed using a simple model by fitting the experimental R_2 rates to Eq. (2):

$$R_2(i) = R_2(\text{int}) \sum_{j=1}^N \exp\left(-\frac{|i-j|}{\lambda}\right) \quad (2)$$

where $R_2(i)$ is the experimental R_2 value for residue i , $R_2(\text{int})$ is the intrinsic relaxation rate of the chain which depends on the temperature and the viscosity of the solution, N is the chain length, and λ is related to the apparent persistence length of the chain.⁴⁹

Acknowledgments

The authors thank Ms. Lauren Wickstrom for helpful discussions and assistance with the chemical shift calculations and Mr. Vadim Patsalo for assistance with the chemical shift calculations.

References

1. Wright PE, Dyson HJ, Lerner RA (1988) Conformation of peptide fragments of protein in aqueous solution: implications for initiation of protein folding. *Biochemistry* 27: 7167–7175.
2. Ferguson N, Fersht AR (2003) Early events in protein folding. *Curr Opin Struct Biol* 13:75–81.
3. Munoz V, Serrano L (1996) Local versus nonlocal interactions in protein folding and stability: an experimentalist's point of view. *Folding Des* 1:R71–R78.
4. Baldwin RL (2002) A new perspective on unfolded proteins. *Adv Protein Chem* 62:361–367.
5. Dill KA, Shortle D (1991) Denatured states of proteins. *Annu Rev Biochem* 60:795–825.
6. Viguera AR, Villegas V, Aviles FX, Serrano L (1997) Favorable native-like helical local interactions can accelerate protein folding. *Folding Des* 2:23–33.
7. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277:985–994.
8. McKnight CJ, Doering DS, Matsudaira PT, Kim PS (1996) A thermostable 35-residue subdomain within villin headpiece. *J Mol Biol* 260:126–134.

9. McKnight CJ, Matsudaira PT, Kim PS (1997) NMR structure of the 35-residue villin headpiece subdomain. *Nat Struct Biol* 4:180–184.
10. Vardar D, Buckley DA, Frank BS, McKnight CJ (1999) NMR structure of an F-actin-binding “headpiece” motif from villin. *J Mol Biol* 294:1299–1310.
11. Chiu TK, Kubelka J, Herbst-Irmer R, Eaton WA, Hofrichter J, Davies DR (2005) High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc Natl Acad Sci USA* 102:7517–7522.
12. Wang M, Tang Y, Sato S, Vugmeyster L, McKnight CJ, Raleigh DP (2003) Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. *J Am Chem Soc* 125:6032–6033.
13. Kubelka J, Eaton WA, Hofrichter J (2003) Experimental tests of villin subdomain folding simulations. *J Mol Biol* 329:625–630.
14. Brewer SH, Vu DM, Tang Y, Li Y, Franzen S, Raleigh DP, Dyer RB (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc Natl Acad Sci USA* 102:16662–16667.
15. Islam SA, Karplus M, Weaver DL (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins. *J Mol Biol* 318:199–215.
16. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282:740–744.
17. Duan Y, Wang L, Kollman PA (1998) The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proc Natl Acad Sci USA* 95:9897–9902.
18. Sullivan DC, Kuntz ID (2002) Protein folding as biased conformational diffusion. *J Phys Chem B* 106:3255–3262.
19. Shen MY, Freed KF (2002) All-atom fast protein folding simulations: the villin headpiece. *Proteins: Struct Funct Genet* 49:439–445.
20. Srinivas G, Bagchi B (2002) Folding and unfolding of chicken villin headpiece: energy landscape of a single-domain model protein. *Curr Sci* 82:179–185.
21. Fernandez A, Shen MY, Colubri A, Sosnick TR, Berry RS, Freed KF (2003) Large-scale context in protein folding: villin headpiece. *Biochemistry* 42:664–671.
22. Zagrovic B, Snow CD, Shirts MR, Pande VS (2002) Simulation of folding of a small α -helical protein in atomistic detail using worldwide-distributed computing. *J Mol Biol* 323:927–937.
23. Zagrovic B, Snow C, Khaliq S, Shirts M, Pande V (2002) Native-like mean structure in the unfolded ensemble of small proteins. *J Mol Biol* 323:153–164.
24. Lin CY, Hu CK, Hansmann UHE (2003) Parallel tempering simulations of HP-36. *Proteins: Struct Funct Genet* 52:436–445.
25. He JB, Zhang ZY, Shi YY, Liu HY (2003) Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions. *J Chem Phys* 119:4005–4017.
26. Jang SM, Kim E, Shin S, Pak Y (2003) Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J Am Chem Soc* 125:14841–14836.
27. van der Spoel D, Lindahl E (2003) Brute-force molecular dynamics simulations of villin headpiece: comparison with NMR parameters. *J Phy Chem B* 107:11178–11187.
28. Ripoll DR, Vila JA, Scheraga HA (2004) Folding of the villin headpiece subdomain from random structures. Analysis of the charge distribution as a function of Ph. *J Mol Biol* 339:915–925.
29. Kim SY, Lee J, Lee J (2005) Folding simulations of small proteins. *Biophys Chem* 115:195–200.
30. De Mori GMS, Colombo G, Micheletti C (2005) Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins: Struct Funct Bioinf* 58:459–471.
31. De Mori GMS, Micheletti C, Colombo G (2004) All-atom folding simulations of the villin headpiece from stochastically selected coarse-grained structure. *J Phys Chem B* 108:12267–12270.
32. Godoy-Ruiz R, Henry ER, Kubelka J, Hofrichter J, Munoz V, Sanchez-Ruiz JM, Eaton WA (2008) Estimating free energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J Phys Chem B* 112:5938–5949.
33. Bi Y, Cho JH, Kim EY, Shan B, Schindelin H, Raleigh DP (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain. *Biochemistry* 46:7497–7505.
34. Ensign DL, Kasson PM, Pande VS (2007) Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J Mol Biol* 374:806–816.
35. Hansmann UHE (2002) Protein-folding simulations in generalized ensembles. *Int J Quant Chem* 90:1515–1521.
36. Jayachandran G, Vishal V, Garcia AE, Pande VS (2007) Local structure formation in simulations of two small proteins. *J Struct Biol* 157:491–499.
37. Lei HX, Duan Y (2007) Two-stage folding of HP-35 from ab initio simulations. *J Mol Biol* 370:196–206.
38. Micheletti C, Colombo G (2008) Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics (vol 58, pg 459, 2005). *Proteins: Struct Funct Bioinf* 70:309.
39. Mukherjee A, Bagchi B (2003) Correlation between rate of folding, energy landscape, and topology in the folding of a model protein HP-36. *J Chem Phys* 118:4733–4747.
40. Piana S, Laio A, Marinelli F, Van Troys M, Bourry D, Ampe C, Martins JC (2008) Predicting the effect of a point mutation on a protein fold: the villin and advillin headpieces and their Pro62Ala mutants. *J Mol Biol* 375:460–470.
41. Wei Y, Nadler W, Hansmann UHE (2008) Backbone and side-chain ordering in a small protein. *J Chem Phys* 128:025105.
42. Yang JS, Wallin S, Shakhnovich EI (2008) Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc Natl Acad Sci USA* 105:895–900.
43. Tang YF, Rigotti DJ, Fairman R, Raleigh DP (2004) Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: the villin headpiece subdomain. *Biochemistry* 43:3264–3272.
44. Tang YF, Goger MJ, Raleigh DP (2006) NMR characterization of a peptide model provides evidence for significant structure in the unfolded state of the villin headpiece helical subdomain. *Biochemistry* 45:6940–6946.
45. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biol NMR* 5:67–81.
46. Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and $\text{C}\alpha$ and $\text{C}\beta$ ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492.
47. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α - and γ -synuclein: implications for fibrillation. *Protein Sci* 15:2795–2804.

48. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. *Science* 295:1719–1722.
49. Schwalbe HF, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea *Biochemistry* 36: 8977–8991.
50. Havlin RH, Tycko R (2005) Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc Natl Acad Sci USA* 102:3284–3289.
51. Myers JK, Oas TG (2001) Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Biol* 8:552–558.
52. Ferguson N, Shape TD, Schartau PJ, Sato S, Allen MD, Johnson CM, Rutherford TJ, Fersht AR (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J Mol Biol* 353:427–446.
53. Chowdhury S, Lei HX, Duan Y (2005) Denatured-state ensemble and the early -stage folding of the G29A mutant of the B-domain of protein A. *J Phys Chem B* 109:9073–9081.
54. Cho JH, Raleigh DP (2006) Denatured state effects and the origin of nonclassical phi values in protein folding. *J Am Chem Soc* 128:16492–16493.
55. Bi Y, Tang Y, Raleigh DP, Cho J (2006) Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: application to the villin headpiece helical subdomain. *Protein Expression Purif* 47:234–240.
56. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33:5984–6003.
57. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe—a multi-dimensional spectral processing system based on Unix Pipes. *J Biomol NMR* 6: 277–293.
58. Johnson BA (2004) Using NMRView to visualize and analyze NMR spectra of macromolecules. *Methods Mol Biol* 278:313–352.
59. Xu XP, Case DA (2001) Automated prediction of ¹⁵N, ¹³C α , ¹³C β and ¹³C' chemical shifts in proteins using a density functional database. *J Biomol NMR* 21: 321–333.
60. Xu XP, Case DA (2002) Probing multiple effects on ¹⁵N, ¹³C α , ¹³C β and ¹³C' chemical shifts in peptides using density functional theory. *Biopolymers* 65:408–423.
61. Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ¹H, ¹³C, and ¹⁵N chemical shifts. *J Biomol NMR* 26:215–240.