NMR Characterization of a Peptide Model Provides Evidence for Significant Structure in the Unfolded State of the Villin Headpiece Helical Subdomain†

Yuefeng Tang,‡ Michael J. Goger,§ and Daniel P. Raleigh*,‡,||

Department of Chemistry, State University of New York, Stony Brook, New York 11790-3400, New York Structural Biology Center, 89 Convent Avenue, New York, New York 10027-7556, and Graduate Program in Biochemistry and Structure Biology, State University of New York, Stony Brook, New York 11794

Received December 5, 2005; Revised Manuscript Received April 11, 2006

ABSTRACT: The villin headpiece subdomain (HP36) is the smallest naturally occurring protein that folds cooperatively. The protein folds on a microsecond time scale. Its small size and very rapid folding have made it a popular target for biophysical studies of protein folding. Temperature-dependent one-dimensional (1D) NMR studies of the full-length protein together with CD and 1D NMR studies of the 21-residue peptide fragment (HP21) derived from HP36 have shown that there is significant structure in the unfolded state of HP36 and have demonstrated that HP21 is a good model of these interactions. Here, we characterized the model peptide HP21 in detail by two-dimensional NMR. Strongly upfield shifted C_α protons, the magnitude of the $^3J_{NH,\alpha}$ coupling constants, and the pattern of backbone−backbone and backbone−side chain NOEs indicate that the ensemble of structures populated by HP21 contains α-helical structure and native as well as non-native hydrophobic contacts. The hydrogen-bonded secondary structure inferred from the NOEs is, however, not sufficient to confer significant protection against amide H−D exchange. These studies indicate that there is significant secondary structure and hydrophobic clustering in the unfolded state of HP36. The implications for the folding of HP36 are discussed.

A full understanding of the protein folding process requires a structural description of the unfolded state which is the starting point of the folding reaction. Characterization of the unfolded states for rapidly folding proteins is extremely important since it has been postulated that preformed structure in the unfolded state could enhance rapid folding (*1*−*4*). The unfolded state that is most relevant for folding studies is the state that is in the equilibrium with the folded state under physiological conditions. Unfortunately, this state is very difficult to access experimentally since the free energy balance normally strongly favors the folded state under native conditions and the dominant species are fully folded proteins. Heat and denaturants are often used to unfold proteins, but the structure of the denatured state under strongly denaturing conditions can be very different from the unfolded state under physiological conditions. Characterizing the unfolded state under native conditions is a major challenge in protein folding, and indirect methods normally have to be used. One approach is to use peptide fragments corresponding to different elements of the protein (*5*−*13*). Some potential tertiary interactions will be missing in these fragments which may play a role in the unfolded state of the full-length protein, but locally stabilized interactions should be maintained. If structure is observed, the normal inference is that it should be present in the unfolded state of the full-length protein.

Here, a 21-residue peptide fragment is used to study the unfolded state of the villin headpiece helical subdomain (HP36).[1] HP36 is the smallest naturally occurring protein domain that folds cooperatively (*14*−*17*). It is one of the fastest folding proteins identified so far, folding on the microsecond time scale (*18*−*20*). The simple topology, small size, and rapid folding of HP36 have made it a very popular model system for computational and experimental folding studies (*21*−*37*). HP36 corresponds to the 35 C-terminal residues of the villin headpiece with an additional Met at the N-terminus. This residue results from expression and was included to allow comparison with other studies. The domain is made up of three α-helices (Figure 1) and contains a well-packed hydrophobic core which includes three phenylalanines, Phe47, Phe51, and Phe58 (*15*−*17*). The numbering system used here corresponds to that used for the full-length villin headpiece. The N-terminal residue is thus denoted Met41.

† This research was supported by NIH Grant GM54233 to D.P.R. The New York Structural Biology Center is supported by NIH Grant GM66354.

* To whom correspondence should be addressed. Phone: (631) 632-9547. Fax: (631) 632-7960. E-mail: draleigh@notes.cc.sunysb.edu.

‡ Department of Chemistry, State University of New York.

§ New York Structural Biology Center.

|| Graduate Program in Biochemistry and Structure Biology, State University of New York.

[1] Abbreviations: DQF-COSY, double-quantum-filtered correlated spectroscopy; EDTA, ethylenediaminetetraacetic acid; Fmoc, 9-fluorenylmethyloxycarbonyl; HP35, helical subdomain of the villin headpiece domain corresponding to residues 791−825 of chicken villin and residues 42−76 of the chicken villin headpiece domain; HP36, HP35 with an additional Met at the N-terminus; HP21, 21-residue fragment starting from the N-terminus of HP36; HPLC, high-pressure liquid chromatography; HSQC, heteronuclear single-quantum coherence; IPTG, isopropyl β-D-thiogalactoside; NOESY, nuclear Overhauser effect spectroscopy; PAL-PEG-PS, poly(ethylene glycol) polystyrene; PMSF, phenylmethanesulfonyl fluoride; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy.

10.1021/bi052484n CCC: $33.50
Published on Web 05/13/2006

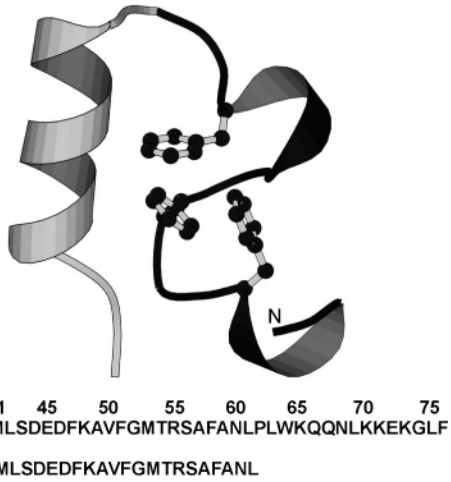

FIGURE 1: Ribbon diagram of HP36 created with Molscript (*48*) (PDB entry 1VII). The region corresponding to HP21 is shaded. The three phenylalanines are indicated, and the N-terminus is labeled. The sequences of HP36 and HP21 are included. HP36 corresponds to the 35 C-terminal residues of the headpiece (Leu42−Phe76), with an additional Met at the N-terminus. Residues are numbered according to their position in the intact headpiece. The N-terminal Met is designated residue 41. The C-terminus of synthetic HP21 is amidated.

Our previous studies have provided evidence that there is significant structure in the unfolded ensemble of HP36 (*38*). The conclusion was based on analysis of the intact protein and low-resolution structural studies of a 21-residue peptide fragment which includes the two N-terminal α-helices. These studies were limited to CD and one-dimensional (1D) NMR because of problems with self-association at concentrations of >150 μM. The 21-residue fragment, denoted here HP21, exhibited significant secondary structure as judged by CD. Ring current-shifted resonances in the 1D ^{1}H NMR spectrum suggested that there is significant clustering of hydrophobic side chains that may be nativelike. Similar structure appears to be populated in the unfolded state of the intact domain as judged by NMR (*38*). HP36 is in rapid exchange on the NMR time scale between the folded and unfolded conformations at elevated temperatures. This allows unfolded state chemical shifts to be estimated by NMR line shape analysis (*18*). The important conclusion that resulted from those studies was that the same protons which experienced significant ring current shifts in HP21 also did so in the thermally unfolded state of HP36. In fact, the shifts were even more pronounced in the thermally unfolded state of the intact protein (*38*). These studies show that the peptide fragment adopts conformations similar to those sampled in the thermally unfolded state of HP36 and strongly suggest that HP21 provides an excellent model of at least some of the interactions in the unfolded state of HP36. Mutation of the two phenylalanines (Phe47 and Phe51) in HP21 was shown to significantly reduce the level of structure, suggesting that specific packing interactions are necessary for stabilizing the unfolded state; however, more detailed information could not be obtained (*38*). Detailed structural characterization of HP21 should improve the understanding of the early stages of the folding process of HP36. Unfortunately, self-association limited our ability to fully characterize the peptide by NMR with conventional NMR instrumentation. Here we characterize this peptide in considerably more detail by taking advantage of NMR cryoprobe technology to study the system at low peptide concentrations where it is known to be monomeric.

MATERIALS AND METHODS

Peptide Synthesis, Expression, and Purification. Unlabeled HP21 was synthesized by standard solid-phase Fmoc chemistry using an Applied Biosystems 433A automated peptide synthesizer as previously described (*38*). PAL-PEG-PS resin was used, generating an amidated C-terminus upon cleavage with a 91% trifluoroacetic acid (TFA)/3% anisole/3% thioanisole/3% ethanedithiol mixture. Purification of synthesized HP21 was achieved by reverse-phase HPLC with 10% water/90% acetonitrile gradients containing 0.067% TFA (v/v) on a C18 column (Vydac).

^{15}N-labeled HP21 was prepared recombinantly as a fusion peptide with the N-terminal domain of L9 (NTL9) (*39*). Cultures for expression of ^{15}N-labeled protein were grown in M9 minimal medium supplemented with 1 g/L $(^{15}NH_4)_2SO_4$ as the main nitrogen source. A single colony of transformed *Escherichia coli* BL21(DE3) strain cells was inoculated into 50 mL of M9 minimal medium with 100 μg/mL ampicillin at 37 °C. When the absorbance at 600 nm reached 0.6, the cells were harvested by centrifugation at 2500*g* for 10 min. The harvested cells were resuspended in 10 mL of M9 minimal medium and then transferred into 1 L of M9 minimal medium with 100 μg/mL ampicillin. The cells were grown at 37 °C to an absorbance at 600 nm of 1.2 and then induced with 1 mM IPTG for 4 h before being harvested. The cells were lysed by sonication in 20 mM Tris buffer with 1 mM PMSF and 1 mM EDTA (pH 7.5). The fusion protein was purified with an ion exchange column followed by reverse-phase HPLC. Cleavage of the fusion protein by factor Xa was performed at 4 °C with 0.5 unit of factor Xa per milligram for 24 h. HP21 was purified by reverse-phase HPLC using the same conditions employed for synthesized HP21. Expressed HP21 has a free C-terminus. The identities of both synthesized and expressed HP21 were confirmed by mass spectrometry.

Nuclear Magnetic Resonance (NMR) Spectroscopy. TOCSY (total correlation spectroscopy), NOESY (nuclear Overhauser effect spectroscopy), and DQF-COSY (double-quantum-filtered correlated spectroscopy) experiments were conducted on a Bruker DRX600 MHz spectrometer with a cryoprobe at the New York Structure Biology Center. ROESY (rotating frame nuclear Overhauser spectroscopy) spectra could not be collected using this instrument. Samples for NMR experiments were dissolved in a 90% H_2O/10% ^2H_2O mixture with 10 mM sodium acetate and 150 mM sodium chloride (pH 5.5) with an HP21 concentration of ~100 μM. Sodium 3-trimethylsilyl-2,2,3,3-d_4-propionate (TSP) was used as an internal reference (0.0 ppm). All the ^{1}H two-dimensional (2D) experiments have a spectral width of 8000 Hz in both dimensions. An additional NOESY spectrum was collected in 99.9% ^2H_2O.

Chemical shift assignments were made using standard procedures. The TOCSY spectrum was used for spin system assignments. Sequential assignment of the peptide was achieved using the NOESY spectrum. Both TOCSY and NOESY spectra were collected at 12 °C with a matrix size of 4096 × 128. The mixing time was 80 ms for TOCSY and 300 ms for NOESY. *J* coupling constants were measured

using a DQF-COSY spectrum at 12 °C with a data set of 8192 × 256 real points. The matrix was zero filled to 16 384 points in t_2. *J* coupling constants were calculated by measuring the separation of the absorptive and dispersive cross-peaks in the DQF-COSY spectrum using the method of Kim and Prestegard (*40*). Ring current contributions to the native state $C_\alpha H$ chemical shifts of HP36 were calculated using the algorithm of Case and co-workers (http://www.scripps.edu/mb/case/qshifts/qshifts.htm). The coordinates of the HP21 region of the full helical subdomain structures were input. In other words, the calculation predicts the ring current contributions that would arise if HP21 populated a fully nativelike fold. The random coil values of Wishart were used to calculate $C_\alpha H$ chemical shift deviations (*41*).

A hydrogen−deuterium exchange experiment for ^{15}N-labeled HP21 was carried out on a Varian Inova 500 MHz spectrometer at 5 °C. A lyophilized dry protein sample which was previously adjusted to pH 5.0 was dissolved in 99.9% ^2H_2O with 10 mM sodium acetate and 150 mM sodium chloride (pD 5.0). A series of $^1H-^{15}N$ HSQC experiments were acquired. The first spectrum was recorded approximately 6 min after the dissolution of the protein. Each spectrum took ∼5 min to record. A matrix of 2048 × 16 complex points was collected with spectral widths of 8000 Hz in the ^1H dimension and 2200 Hz in the ^{15}N dimension. Intrinsic exchange rates were calculated using the SPHERE program of Roder and colleagues (http://www.fccc.edu/research/labs/roder/sphere/). Assignments of the $^1H-^{15}N$ HSQC spectra of expressed HP21 were obtained by comparison of the amide proton chemical shifts to the assignments for the synthesized construct and were confirmed with a three-dimensional (3D) HSQC-TOCSY experiment. The three-dimensional $^1H-^{15}N$ HSQC-TOCSY experiment was performed on a Bruker DMX700 MHz spectrometer with 100 μM ^{15}N-labeled HP21 in a 90% H_2O/10% ^2H_2O mixture with 10 mM sodium acetate and 150 mM sodium chloride at pH 5.5 and 5 °C. The spectral width was 2200 Hz in the ^{15}N dimension and 8000 Hz in the ^1H dimension. The matrix size was 128 × 1024 × 128.

RESULTS

To analyze any structural preferences of HP21, 2D ^1H TOCSY and NOESY spectra were recorded at 12 °C for a 100 μM sample. The peptide is predominately (>90%) monomeric under these conditions as demonstrated by analytical ultracentrifugation experiments (*38*). The TOCSY spectrum was used to identify spin systems for each residue, and sequential assignments were made using the NOESY spectrum. All of the cross-peaks in the 2D spectra could be easily assigned (Figure 2). The chemical shift dispersion in the amide proton region was ∼1.7 ppm, and the spectra were very well resolved, suggesting some well-developed structure in the ensemble of conformations adopted by HP21. The deviation of the C_α proton chemical shifts from random coil values is commonly used to probe the formation of secondary structure (*41*, *42*). Upfield-shifted C_α proton resonances relative to random coil values are indicative of a propensity to adopt α-helical conformations. HP21 contains two of the helices found in HP36. Residues Ser43−Gly52 and Thr54−Leu61 are helical in the crystal structure of the intact protein. C_α proton chemical shifts of 16 residues (20 residues in total could be identified in the spectra) are shifted by more than 0.2 ppm relative to random coil values (Figure 3a). In the region of Asp44−Phe47, Ala49−Phe51, and Ser56−Leu61, the C_α protons are all upfield-shifted and the secondary shifts for eight of these are more negative than 0.2 ppm, suggesting a strong tendency for α-helix formation. Moreover, the C_α proton secondary shifts for all of the residues in HP21 are considerably larger than the values measured for the corresponding residues in two small peptide fragments which contain either of the two helices of HP21 (*38*). Ring current effects can also contribute to the deviation of C_α proton chemical shifts from random coil values and can cloud their interpretation in terms of secondary structure propensities. A number of residues are predicted to have significant (>0.21 ppm) ring current effects in the native state. These include Val50, Phe51, Thr54, Arg55, and Ala59 (Supporting Information). Of course, the ensemble of HP21 conformations does not adopt the unique well-defined structure found in this region of HP36, but the analysis suggests that some caution may be necessary in interpreting the C_α proton deviations for these residues in HP21 in terms of secondary structure since ring current effects could be significant. Nevertheless, comparison of C_α proton shifts between HP21 and HP36 is still a useful metric of structure formation. It is also interesting to compare the C_α proton secondary shifts of HP21 with those observed for HP36 (*14*). There are significant deviations for residues Arg55, Phe47, and Phe58 ranging from 1.36 to 0.27 ppm; however, the differences in

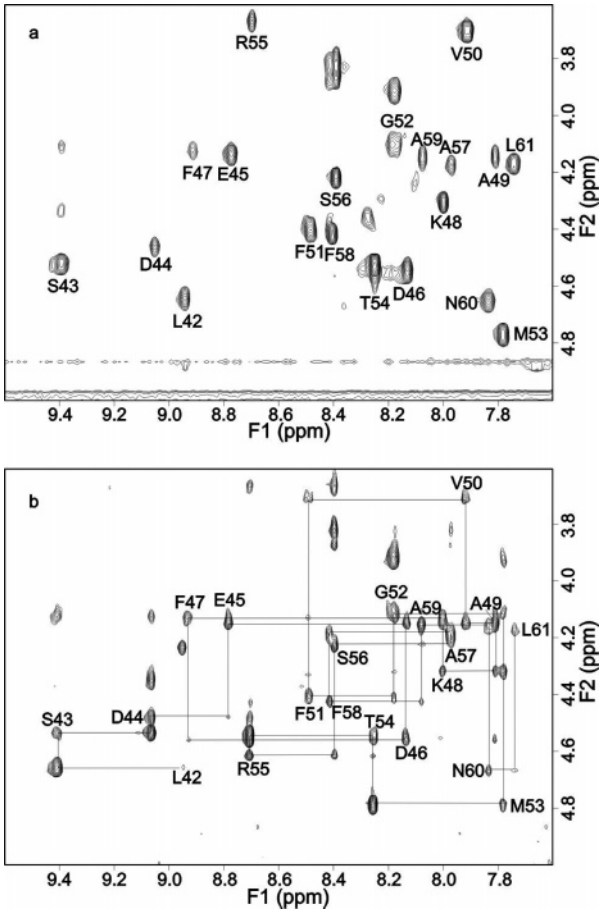

FIGURE 2: Fingerprint region of the TOCSY spectrum (a) and the NOESY spectrum (b) of HP21 recorded at pH 5.5 and 12 °C in 150 mM sodium chloride and 10 mM sodium acetate. Assignments are listed, and the sequential NOE connectivities are shown.

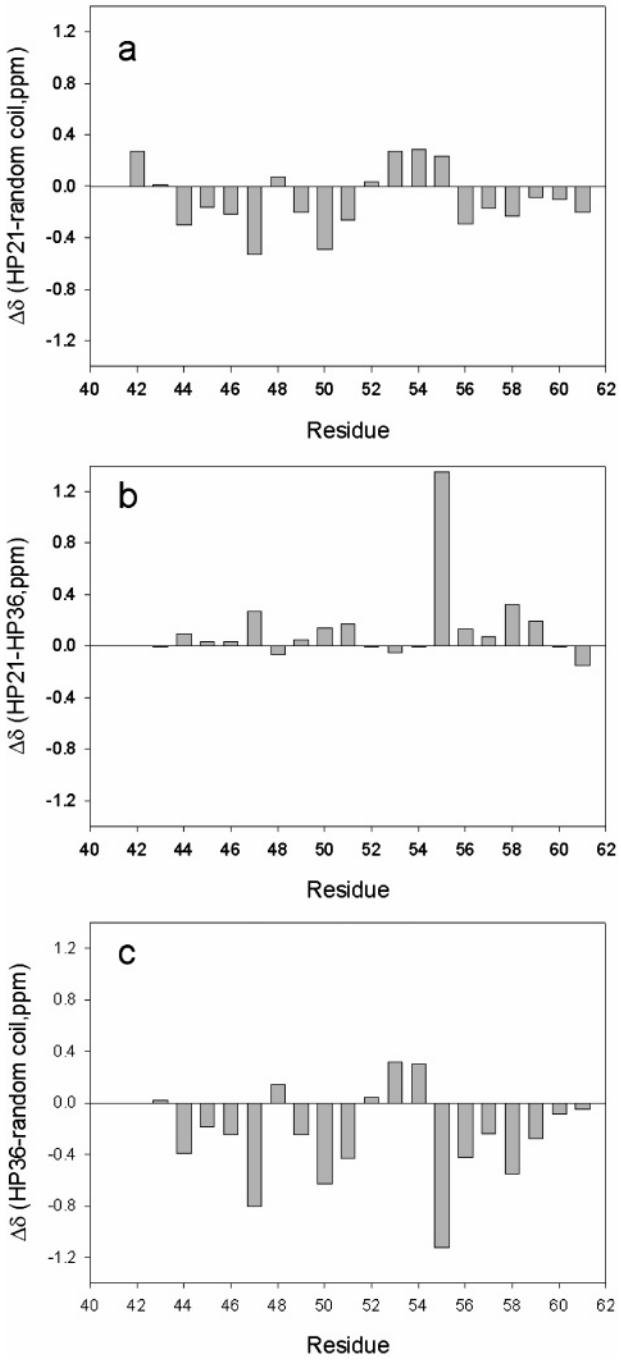

FIGURE 3: Deviation of HP21 $C_\alpha H$ chemical shifts from (a) random coil values and (b) those of wild-type HP36. (c) Deviation of HP36 $C_\alpha H$ chemical shifts from random coil values. Data were collected at 12 °C and pH 5.5 in 150 mM sodium chloride and 10 mM sodium acetate for HP21 and at pH 5.0 and 30 °C for HP36 (*14*).

C_α proton shifts for the rest of the residues are less than 0.2 ppm, and 11 residues exhibit deviations that are less than 0.1 ppm (Figure 3b). This provides additional evidence that these regions have a tendency to sample the native fold in the HP21 ensemble.

$^3J_{NH,\alpha}$ coupling constants are sensitive to the φ angle and are commonly used to probe backbone conformation. $^3J_{NH,\alpha}$ coupling constants were measured at 12 °C and pH 5.5 for HP21. Lys48, Ala49, Met53, Arg55, and Ala59 have noticeably reduced $^3J_{NH,\alpha}$ coupling constants (<5 Hz) relative to the values expected for a disordered peptide, suggesting a strong α-helical preference. The coupling constants were

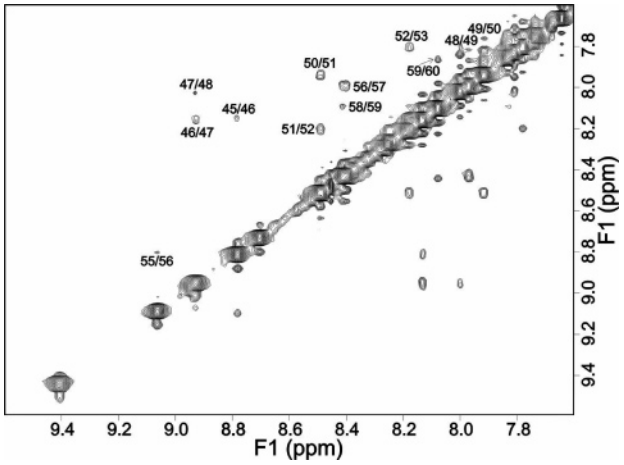

FIGURE 4: Amide−amide region of the NOESY spectrum of HP21 recorded at 12 °C and pH 5.5 in 150 mM sodium chloride and 10 mM sodium acetate. Sequential NH_i−NH_{i+1} connectivities are labeled.

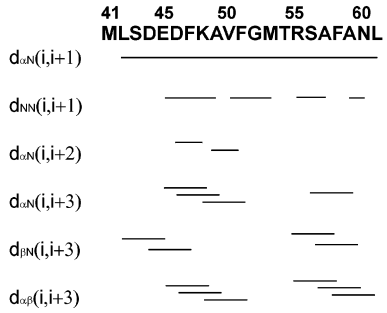

FIGURE 5: Summary of inter-residue backbone−backbone and backbone−side chain NOEs observed for HP21.

previously measured for two small peptide fragments corresponding to the two helices in HP21 and range from 5.3 to 7.4 Hz (*38*). The $^3J_{NH,\alpha}$ coupling constants measured for HP21 are smaller or similar in magnitude to those measured in the small peptide fragments. The one exception is Asn60. The chemical shift assignments as well as the *J* coupling constants for each residue are listed in the Supporting Information.

A NOESY spectrum acquired at 12 °C provides additional evidence for helical structure in the HP21 ensemble. The amide region of the NOESY spectra is shown in Figure 4. A total of 12 amide (i, i + 1) NOEs were observed. In contrast, only a few sequential amide−amide cross-peaks could be detected in ROESY spectra of the two small peptide fragments. Medium-range α- or β-proton−amide and α-proton−β-proton NOEs are additional indicators of helical structure. An extensive set of α- or β-proton−amide and α-proton−β-proton NOEs are observed for HP21. The pattern of connectivities suggests that helical structure is populated in some members of the ensemble from Glu45 to Phe51 and from Arg55 to Leu61 (Figure 5). In the crystal structure, helices are located from Ser43 to Gly52 and from The54 to Leu61. The NMR parameters strongly suggest that although the peptide is sampling a range of conformations, there is a significant tendency to adopt native α-helical structure. The pattern of NOEs is summarized in Figure 5.

Side chain−side chain NOEs are important parameters for defining tertiary packing. A number of methyl−phenyl peaks could be observed in the NOESY spectra recorded in 99.9%

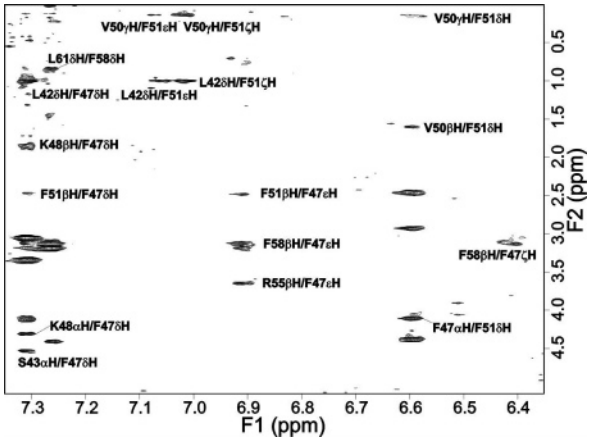

FIGURE 6: Aromatic−methyl region of the NOESY spectrum of HP21 recorded at 12 °C and pD 5.5 in D_2O with 150 mM sodium chloride and 10 mM sodium acetate. NOE cross-peaks between different residues are labeled.

^2H_2O that are broadly consistent with a tendency toward nativelike hydrophobic clustering in the HP21 ensemble (Figure 6). For example, cross-peaks between the γ- and β-protons of Val50 and the phenyl ring protons of Phe51, the δ-proton of Leu42 and the Phe47/Phe51 aromatic ring protons, and the Leu61 δ-proton and the Phe58 ring protons provide evidence of nativelike hydrophobic packing. A cross-peak between the ζ-proton of Phe47 and the β-proton of Phe58 is observed, indicating that the two phenyl rings are close, rationalizing the strong downfield shift of the Phe47 ζ-resonance. Most of these interproton distances are less than 4.5 Å in the native structure of HP36, and some are less than 3.5 Å. Thus, the observation of these NOEs is consistent with a tendency to adopt nativelike side chain clustering in the ensemble of conformations adopted by HP21. Interestingly, there were also several non-native NOEs in the NOESY spectrum of HP21 demonstrating a tendency to form some non-native contacts in the unfolded state ensemble. For example, NOEs were observed between the α-proton of Ser43 and the δ-proton of Phe47 and between the β-proton of Met53 and the γ-proton of Val50.

A ^{15}N-labeled sample of HP21 was prepared recombinantly to study the amide exchange behavior of HP21. This peptide has a free C-terminus in contrast to the amidated C-terminus of the synthetic peptide. However, this does not affect the structure. CD experiments with this peptide indicate that the amount of α-helical structure is very similar to that found in the synthetic peptide (data not shown). NMR experiments show a very similar pattern of peaks for the expressed and synthesized peptide (data not shown). Therefore, the conformational propensities of expressed HP21 are likely to be very close to those of the synthetic variant. The HSQC spectrum of this peptide in H_2O displays a wide chemical shift range in both the ^1H and ^{15}N dimensions, consistent with a significant tendency to populate ordered structure (Figure 7a). An amide exchange experiment shows that all of the amide protons in HP21 exchange rapidly even at pD 5.0 and 5 °C. Only the amide protons of Phe47, Lys48, Ala49, Val50, Phe58, and Leu61 could be observed for HP21 in the first spectrum recorded after 6 min (Figure 7b). The peaks disappear too rapidly to allow accurate measurement of protection factors, but it is worth noting that these residues are among the amino acids that form the hydrophobic core

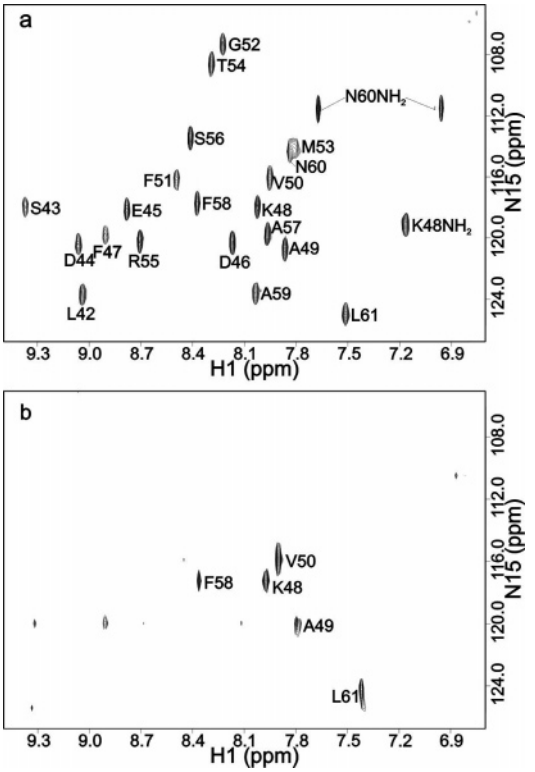

FIGURE 7: HSQC spectra of (a) HP21 in H_2O at pH 5.0 and (b) HP21 recorded at pD 5.0 (uncorrected meter reading) after the proteins had been dissolved in D_2O. The spectrum required 5 min, and data collection started 6 min after the sample was dissolved. Both spectra were recorded at 5 °C. The buffer consisted of 150 mM sodium chloride and 10 mM sodium acetate.

of HP36 and they all have relatively large protection factors in the intact domain. One possibility is that this subset of resonances is observed simply because they have the lowest intrinsic exchange rates. Several of the residues do have slow intrinsic rates, particularly Val50 and Leu61 (Supporting Information) and to a lesser extent Phe47; however, Lys48, Ala49, and Phe58 have faster intrinsic rates than several residues which do not exhibit cross-peaks. This suggests that the structure detected by CD and by NOE measurements may confer some low-level protection against exchange, although the effects are not large. It is important to emphasize that the absence of significant protection factors does not indicate a lack of hydrogen-bonded structure since partially formed, moderate levels of helical structure can lead to low protection factors (*43*). The exchange data further suggest that HP21 and HP36 share some features in common since the only resonances present in the first HSQC spectrum of HP21 in ^2H_2O arise from residues that are strongly protected in HP36.

DISCUSSION

In this study, a 21-residue peptide fragment derived from chicken villin headpiece HP36 was used as a model of the unfolded state of HP36. Strongly upfield shifted C_α protons, the magnitude of the $^3J_{NH,\alpha}$ coupling coefficients, and the pattern of backbone−backbone and backbone−side chain NOEs show that α-helical secondary structure is well-developed in the HP21 ensemble. Analysis of the 2D NOE data clearly indicates a tendency to form nativelike interactions between the three phenylalanines and several hydrophobic residues. The NMR studies clearly demonstrate that

the ensemble of conformations populated by HP21 experiences a bias toward both helical structure and native side chain clustering. This in turn strongly suggests that there is significant structure in the unfolded ensemble of HP36. Interestingly, there were also several non-native NOEs in the NOESY spectrum of HP21 demonstrating a tendency to form some non-native contacts in the unfolded state ensemble. It is natural to inquire if it is possible to calculate a structure representing the conformation of HP21. This is dangerous for several reasons. First and most importantly, HP21 is sampling a range of conformations; thus, even if it were possible, the meaning of the resulting structure would be unclear. Second, the limited number of restraints would lead to very limited permission.

The role of unfolded state structure in several other rapid folding small helical proteins has been considered. In some cases, preorganized structure is proposed to play an important role in rapid folding, while in others, rapid folding appears to occur from a less structured unfolded state (*12*, *44*−*47*). The studies presented here provide detailed information about structural tendencies in the unfolded ensemble of this protein. On the basis of these studies, the unfolded state ensemble of HP36 contains considerable structure in the region of the first two α-helices, including a tendency toward nativelike clustering of side chains. It is tempting to speculate that the extremely rapid folding rate found for HP36 is a consequence of preformed structure in the unfolded state. While this may be the case, it is very important, we believe, to avoid making kinetic arguments based upon equilibrium studies, especially since there are well-documented examples where preformed structure does not contribute to rapid folding. What the data presented here do allow is the design of experiments which examine the effect of modulating unfolded state structure upon the rapid folding of HP36. Preliminary investigations along these lines have shown that disruption of the hydrophobic/aromatic clusters by Phe to Leu substitutions does not alter the folding rate, suggesting that the unfolded state structure or at least the Phe interactions may not be a strict prerequisite for rapid folding (*20*). The analysis presented here provides the needed information required to further test this preliminary conclusion by targeting other unfolded state interactions.

ACKNOWLEDGMENT

We thank Yuan Bi and Jae-Hyun Cho for help with peptide expression and Lauren Wickstrom for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

A table of ^{1}H chemical shifts for HP21; a table listing C_α proton deviations from random coil values, the calculated contribution of ring current effects to native state C_α proton chemical shifts, and the measured $^3J_{NH,\alpha}$ coupling constants; and a table of the calculated intrinsic rates of amide proton exchange. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Viguera, A. R., Villegas, V., Aviles, F. X., and Serrano, L. (1997) Favourable native-like helical local interactions can accelerate protein folding, *Folding Des. 2*, 23−33.
2. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol. 277*, 985−94.
3. Munoz, V., and Serrano, L. (1996) Local versus nonlocal interactions in protein folding and stability: An experimentalists's point of view, *Folding Des. 1*, R71−8.
4. Ferguson, N., and Fersht, A. R. (2003) Early events in protein folding, *Curr. Opin. Struct. Biol. 13*, 75−81.
5. Wright, P. E., Dyson, H. J., and Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding, *Biochemistry 27*, 7167−75.
6. Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., and Wright, P. E. (1992) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin, *J. Mol. Biol. 226*, 795−817.
7. Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H. C., Lerner, R. A., and Wright, P. E. (1992) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin, *J. Mol. Biol. 226*, 819−35.
8. Viguera, A. R., Jimenez, M. A., Rico, M., and Serrano, L. (1996) Conformational analysis of peptides corresponding to β-hairpins and a β-sheet that represent the entire sequence of the α-spectrin SH3 domain, *J. Mol. Biol. 255*, 507−21.
9. Ladurner, A. G., Itzhaki, L. S., de Prat Gay, G., and Fersht, A. R. (1997) Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2, *J. Mol. Biol. 273*, 317−29.
10. Kemmink, J., and Creighton, T. E. (1993) Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding, *J. Mol. Biol. 234*, 861−78.
11. Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat Gay, G., and Fersht, A. R. (1995) Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2, *J. Mol. Biol. 254*, 289−304.
12. Spector, S., Rosconi, M., and Raleigh, D. P. (1999) Conformational analysis of peptide fragments derived from the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: Evidence for nonrandom structure in the unfolded state, *Biopolymers 49*, 29−40.
13. Luisi, D. L., Wu, W. J., and Raleigh, D. P. (1999) Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: Evidence for stable native-like secondary structure in the unfolded state, *J. Mol. Biol. 287*, 395−407.
14. McKnight, C. J., Doering, D. S., Matsudaira, P. T., and Kim, P. S. (1996) A thermostable 35-residue subdomain within villin headpiece, *J. Mol. Biol. 260*, 126−34.
15. McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nat. Struct. Biol. 4*, 180−4.
16. Vardar, D., Buckley, D. A., Frank, B. S., and McKnight, C. J. (1999) NMR structure of an F-actin-binding "headpiece" motif from villin, *J. Mol. Biol. 294*, 1299−310.
17. Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., and Davies, D. R. (2005) High-resolution X-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein, *Proc. Natl. Acad. Sci. U.S.A. 102*, 7517−22.
18. Wang, M., Tang, Y., Sato, S., Vugmeyster, L., McKnight, C. J., and Raleigh, D. P. (2003) Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, *J. Am. Chem. Soc. 125*, 6032−3.
19. Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *J. Mol. Biol. 329*, 625−30.
20. Brewer, S. H., Vu, D. M., Tang, Y., Li, Y., Franzen, S., Raleigh, D. P., and Dyer, R. B. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain, *Proc. Natl. Acad. Sci. U.S.A. 102*, 16662−7.
21. Islam, S. A., Karplus, M., and Weaver, D. L. (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins, *J. Mol. Biol. 318*, 199−215.
22. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science 282*, 740−4.
23. Duan, Y., Wang, L., and Kollman, P. A. (1998) The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation, *Proc. Natl. Acad. Sci. U.S.A. 95*, 9897−902.
24. Sullivan, D. C., and Kuntz, I. D. (2002) Protein folding as biased conformational diffusion, *J. Phys. Chem. B 106*, 3255−62.

25. Shen, M. Y., and Freed, K. F. (2002) All-atom fast protein folding simulations: The villin headpiece, *Proteins: Struct., Funct., Genet. 49*, 439−45.
26. Srinivas, G., and Bagchi, B. (2002) Folding and unfolding of chicken villin headpiece: Energy landscape of a single-domain model protein, *Curr. Sci. 82*, 179−85.
27. Fernandez, A., Shen, M. Y., Colubri, A., Sosnick, T. R., Berry, R. S., and Freed, K. F. (2003) Large-scale context in protein folding: Villin headpiece, *Biochemistry 42*, 664−71.
28. Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002) Simulation of folding of a small α-helical protein in atomistic detail using worldwide-distributed computing, *J. Mol. Biol. 323*, 927−37.
29. Zagrovic, B., Snow, C., Khaliq, S., Shirts, M., and Pande, V. (2002) Native-like mean structure in the unfolded ensemble of small proteins, *J. Mol. Biol. 323*, 153−64.
30. Lin, C. Y., Hu, C. K., and Hansmann, U. H. (2003) Parallel tempering simulations of HP-36, *Proteins 52*, 436−45.
31. He, J. B., Zhang, Z. Y., Shi, Y. Y., and Liu, H. Y. (2003) Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions, *J. Chem. Phys. 119*, 4005−17.
32. Jang, S. M., Kim, E., Shin, S., and Pak, Y. (2003) Ab initio folding of helix bundle proteins using molecular dynamics simulations, *J. Am. Chem. Soc. 125*, 14841−6.
33. van der Spoel, D., and Lindahl, E. (2003) Brute-force molecular dynamics simulations of Villin headpiece: Comparison with NMR parameters, *J. Phys. Chem. B 107*, 11178−87.
34. Ripoll, D. R., Vila, J. A., and Scheraga, H. A. (2004) Folding of the villin headpiece subdomain from random structures. Analysis of the charge distribution as a function of pH, *J. Mol. Biol. 339*, 915−25.
35. Kim, S. Y., Lee, J., and Lee, J. (2005) Folding simulations of small proteins, *Biophys. Chem. 115*, 195−200.
36. De Mori, G. M. S., Colombo, G., and 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*, 459−71.
37. De Mori, G. M., Micheletti, C., and Colombo, G. (2004) All-atom folding simulations of the villin headpiece from stochastically selected coarse-grained structure, *J. Phys. Chem. B 108*, 12267−70.
38. Tang, Y. F., Rigotti, D. J., Fairman, R., and Raleigh, D. P. (2004) Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: The villin headpiece subdomain, *Biochemistry 43*, 3264−72.
39. Bi, Y., Tang, Y., Raleigh, D. P., and 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*, 234−40.
40. Kim, Y., and Prestegard, J. H. (1989) Measurement of vicinal coupling from cross peaks in COSY spectra, *J. Magn. Reson. 84*, 9−13.
41. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR 5*, 67−81.
42. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical-shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647−51.
43. Mori, S., vanZijl, P. C. M., and Shortle, D. (1997) Measurement of water-amide proton exchange rates in the denatured state of staphylococcal nuclease by a magnetization transfer technique, *Proteins: Struct., Funct., Genet. 28*, 325−32.
44. Myers, J. K., and Oas, T. G. (2001) Preorganized secondary structure as an important determinant of fast protein folding, *Nat. Struct. Biol. 8*, 552−8.
45. Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C. M., Daggett, V., and Fersht, A. R. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation, *Proc. Natl. Acad. Sci. U.S.A. 101*, 6450−5.
46. Chowdhury, S., Lei, H. X., and 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*, 9073−81.
47. Ferguson, N., Sharpe, T. D., Schartau, P. J., Sato, S., Allen, M. D., Johnson, C. M., Rutherford, T. J., and Fersht, A. R. (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family, *J. Mol. Biol. 353*, 427−46.
48. Kraulis, P. J. (1991) Molscript: A Program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−50.

BI052484N