BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2009,** *48,* 4607–4616 4607
DOI:10.1021/bi8021763

Analysis of Core Packing in a Cooperatively Folded Miniature Protein: The Ultrafast Folding Villin Headpiece Helical Subdomain†

Shifeng Xiao,‡ Yuan Bi,‡ Bing Shan,‡ and Daniel P. Raleigh*,‡,§

‡*Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794-3400, and* §*Graduate Program in Biochemistry and Structural Biology and Institute for Chemical Biology and Drug Discovery, State University of New York at Stony Brook, Stony Brook, New York 11794*

Received November 25, 2008; Revised Manuscript Received March 17, 2009

ABSTRACT: The helical subdomain of the villin headpiece is the smallest naturally occurring cooperatively folded protein. Its small size, simple three-helix topology, and very rapid folding have made it an extremely popular model system for computational and theoretical studies of protein folding. The domain has a well-packed hydrophobic core comprised in part of an unusual set of three closely packed phenylalanine residues, F47, F51, and F58 (denoted using the numbering of the larger headpiece protein). Aromatic−aromatic interactions have been thought to play a critical role in specifying the subdomain fold and have been proposed to play a general role in stabilizing small proteins. The modest stability of the subdomain has hindered studies of core packing since multiple mutations can lead to constructs which fail to fold, and even single mutants can result in poorly folded variants. Using a previously characterized hyperstable mutant of the domain, generated by targeting surface residues, a complete set of single, double, and triple core Phe to Leu mutants were characterized. A highly conserved surface Trp which is part of a Trp−Pro interaction was also examined. All mutants are well-folded as judged by CD and NMR, and all exhibit sigmoidal urea and thermally induced unfolding transitions, thus proving that aromatic−aromatic, aromatic−proline, or aromatic−hydrophobic interactions are not required for specifying the subdomain fold. Double mutant cycle analysis demonstrates that F47 and F51 make the strongest pairwise interaction. Mutations which lack F58 are the most destabilized, although even the triple mutant is folded. Interestingly, mutation of the central Phe, F51, has the smallest effect on stability even though it makes contact with both F47 and F58 and appears to form the strongest pairwise interaction.

There is considerable interest in small cooperatively folded protein domains and subdomains since their small size and often very rapid folding make them highly attractive model systems for computational, theoretical, and experimental studies of protein folding. In this regard, the helical subdomain of the villin headpiece has emerged as perhaps the most popular system for molecular dynamics (MD)[1] simulations of protein folding. The subdomain is the smallest naturally occurring protein domain which has been shown to fold cooperatively. The domain folds on the microsecond time scale and is thus one of the fastest folding proteins known. Its rapid folding coupled with its small size, 35 or 36 residues depending on the construct, and simple topology have made it the focus of a very large number of computational and theoretical studies, with more than 20 independent research groups publishing the results of their folding calculations (*1*−*40*). There have been considerably fewer published experimental studies on the kinetics of its folding, but we and the NIH group have used laser temperature jump methods to show that it folds on the microsecond time scale (*6*, *7*, *40*, *41*).

The villin headpiece helical subdomain consists of residues Leu42−Phe76 of villin and is denoted HP35 (Figure 1). The subdomain sequence is normally numbered according to the numbering used for the intact villin headpiece. Thus, the first residue of HP35 is Leu42 and its C-terminus is residue 76 (*1*). Recombinant versions of HP35 retain the N-terminal Met and are thus 36 residues in length. That construct, which is the parent sequence for these studies, is denoted HP36, and the N-terminal Met is designated Met41. The domain has a simple topology made up of three α-helices (Figure 1). HP35 and HP36 have very well packed hydrophobic cores and contain a striking triad of buried phenylalanine residues at positions 47, 51, and 58 which pack against each other in the native state (*2*, *42*). The protein also contains a conserved Trp at position 64 which has been proposed to play a critical role in dictating the fold via an aromatic proline interaction (*38*). Interactions involving aromatic residues are observed in many, but not all, naturally occurring or designed

†This work was supported by NSF Grant MCB 614365 to D.P.R.

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

[1]Abbreviations: AUC, analytical ultracentrifugation; CD, circular dichroism; C_M, midpoint of the denaturant-induced unfolding transition; DM HP36, hyperstable N68A/K70M double mutant of HP36; HP36, villin headpiece helical subdomain corresponding to residues Leu42−Phe76 of the headpiece with an additional Met residue at the N-terminus; MALDI-TOF, matrix-assisted laser desorption and ionization time-of-flight mass spectrometry; MD, molecular dynamics; NIH, National Institutes of Health; NMR, nuclear magnetic resonance; T_m, midpoint of the thermal unfolding transition.

Published on Web 4/8/2009

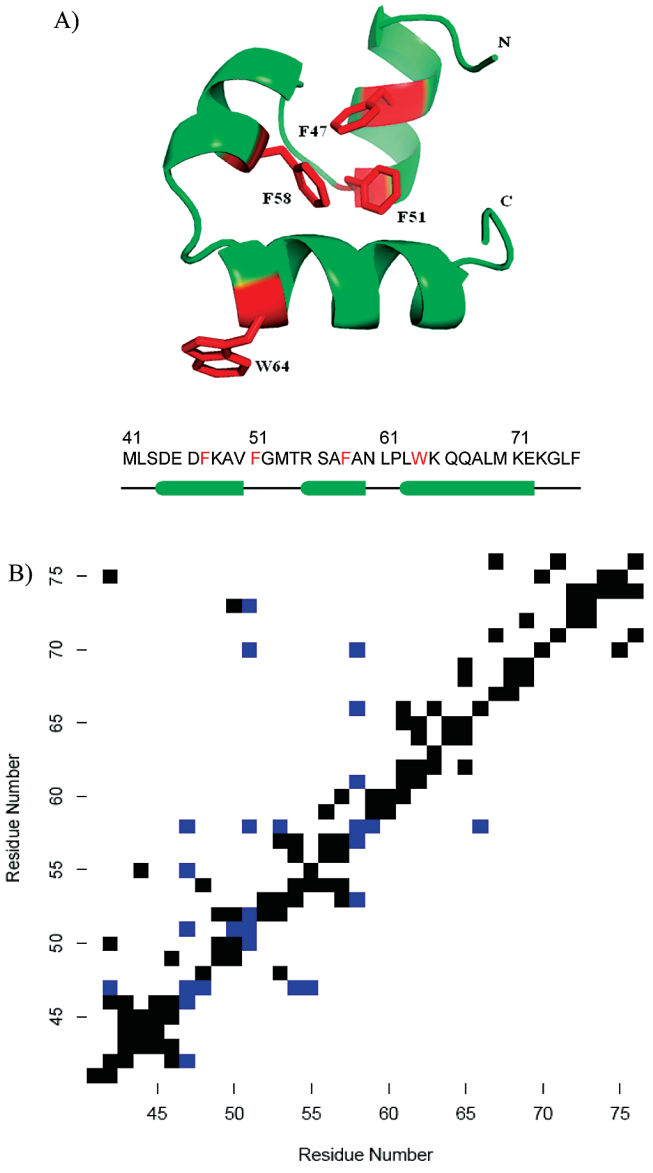

FIGURE 1: (A) Ribbon diagram of the structure of DM HP36 generated using PyMOL, version 0.99. The N- and C-termini are labeled, and the Phe hydrophobic core residues and the surface Trp residue are shown in stick format and are labeled. The primary sequence is shown together with a diagram of secondary structures. α-Helices are represented as cylinders. (B) Contact map for DM HP36 side chain−side chain contacts shown above the diagonal and side chain−main chain contacts below the diagonal. Contacts involving the three Phe residues are colored blue. A 6 Å cutoff was used.

miniature proteins (*43*−*45*), and the hypothesis that they play a particularly critical role in stabilizing and specifying their folds has been advanced (*38*, *43*−*46*). Aromatic−aromatic interactions have also been proposed to play an essential role in the ordered self-assembly of small amyloidogenic polypeptides (*47*, *48*). The side chains of the three core aromatic residues in HP36 are the three most buried in the hydrophobic core with percent fractional accessibilities of 6, 4, and 7%, respectively. Each of the core phenylalanines is highly conserved. Positions 51 and 58 are almost always Phe. The rare exceptions are Leu at position 51 in two sequences and Leu, Tyr, or Trp at position 58 in a total of three sequences, while Phe 47 appears to be completely conserved (*3*). There are no known examples of double mutants which lack aromatic residues at both positions.

Surprisingly, there have been very few mutational studies of the role of the hydrophobic core in the folding or stability of the subdomain, presumably in part because the wild type is not particularly stable in terms of $\Delta G°$, although as expected it has a high T_m (*7*, *26*, *38*, *42*, *49*, *50*). The modest $\Delta G°$ of folding for the wild type has the unfortunate consequence that even moderately destabilizing single or multiple mutants may prevent the subdomain from folding. The failure to fold could arise because the mutations disrupt critical interactions which are absolutely required to specify the fold, or they may simply destabilize the native state so much that only a very very small fraction of molecules are folded. In the latter scenario, the small fraction of molecules which are folded do adopt the wild-type structure, while in the former, the mutants are incapable of adopting the wild-type structure even under stabilizing conditions because critical interactions are missing. The difference in the two scenarios gets to the heart of the difference between the protein folding and the protein stability problems (*51*). In the case of HP35 and HP36, the failure of certain Phe mutants to fold naturally leads to the question of whether aromatic−aromatic interactions are required for specifying the fold in this unusually small protein domain. In addition, studies of the conserved Trp at position 64 have led to the proposal that it participates in critical "gatekeeper" interactions which play a key role in specifying the fold (*38*). Those studies, like the previous analysis of the core aromatics, were conducted in the normal wild-type background and lead to significant destabilization.

Recently, we described a hyperstable double mutant of the domain which was generated by mutating two surface residues whose side chains are exposed to solvent. The high-resolution crystal structure of the double mutant shows that its structure and core packing are identical to those of the wild type (*26*). Here we take advantage of this hyperstable variant to conduct an analysis of core aromatic packing in the villin headpiece helical subdomain and to analyze the role of the surface Trp. Many of the mutants characterized here so destabilize the wild-type protein that it cannot fold. We show that interactions involving aromatic residues are not required to specify the HP36 fold.

MATERIALS AND METHODS

Protein Expression and Purification. The plasmid (pET3a-NTL9-FXa-HP36) containing the gene for HP36 was prepared as described previously (*26*). The mutant proteins were expressed as fusion proteins with the N-terminal domain of L9, and the fusion protein was cleaved with Factor Xa and purified as described previously (*26*). The cleavage temperature for DM F47L/F58L, DM F51L/F58L, and DM F47L/F51L/F58L was 4 °C, while the cleavage temperature for the four other mutants was 23 °C. All proteins were more than 95% pure. The identities of all proteins were confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). The observed and expected molecular weights were as follows: DM F47L, observed 4114.6, expected 4115.9; DM F58L, observed 4113.6, expected 4115.9; DM F51L, observed 4115.4, expected 4115.9; DM F47L/F51L, observed 4084.9, expected 4081.9; DM F47L/F58L, observed 4080.8, expected 4081.9; DM F51L/F58L, observed 4081.2, expected 4081.9; DM F47L/F51L/F58L, observed 4043.7, expected 4047.9; DM W64L, observed 4076.0, expected 4076.9.

NMR. All NMR spectra were recorded on a Varian Instruments Inova 600 MHz spectrometer at 25 °C with approximately 1 mM

protein. Samples were prepared in 90% H_2O, 10% D_2O, 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0 and referenced against sodium 3-(trimethylsilyl)tetradeuteriopropionate.

Circular Dichroism. Wavelength scans were recorded from 260 to 195 nm as the average of five repeats. Spectra were collected at 25 °C with samples of 15−30 μM protein in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0. Thermal unfolding experiments were performed from 2 to 96 °C with a 2 °C interval, and the signals were monitored at 222 nm. The buffer and protein concentration was the same as that used in wavelength scan experiments. Thermal unfolding curves were fit to the standard equation for two-state folding to obtain T_m and $\Delta H°(T_m)$. The fit includes terms to describe linear pre- and post-transition baselines. Urea unfolding curves were recorded at 25 °C and 222 nm using an AVIV Instruments model 202SF CD instrument equipped with an automatic titrator. The concentration of urea was increased from 0 to ~10 M in ~0.25 M steps and was measured by the refractive index. The buffer and protein concentration was the same as that used in wavelength scan experiments. Urea unfolding curves were fit to the standard linear extrapolation model in which

$$\Delta G°(\text{urea}) = \Delta G°(H_2O) - m[\text{urea}] \quad (1)$$

to obtain m values and $\Delta G°$ in the absence of denaturant. The fit assumes that the pre- and post-transition baselines are linear functions of urea concentration.

Analytical Ultracentrifugation. Data were collected with a Beckman Optima XL-A analytical ultracentrifuge at 25 °C and 280 nm using rotor speeds of 38000 rpm for 24 h and 48000 rpm for an additional 24 h. Six-channel, 12 mm path length, charcoal-filled Epon cells with quartz windows were used. Samples of 20, 40, and 50 μM protein were prepared in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0. The partial specific volume and solution density were calculated with SEDNTERP. Data were analyzed using the HeteroAnalysis program from the Analytical Ultracentrifugation Facility at the University of Connecticut (Storrs, CT).

RESULTS

The N68A/K70M double mutant of HP36 was used as the pseudo-wild type and is designated DM HP36 to distinguish it from the true wild type. Both N68 and K70 are surface residues and exposed to solvent; thus, they may be replaced without altering the core packing, a feature which has been confirmed by comparison of the high-resolution crystal structures of DM HP36 and wild-type HP35 (*26*). A set of seven Phe to Leu mutants consisting of the complete set of single mutants, all three double mutants, and the triple mutant. The sole Trp was also mutated to Leu. Aromatic to Leu rather than aromatic to Ala mutations were made so that aromatic residues could be removed without deleting significant side chain volume and thus potentially leading to significant repacking effects. A Phe to Leu substitution does result in a decrease in side chain volume of approximately 35 $Å^3$; however, it is the smallest change associated with mutation to a hydrophobic residue and thus the most conservative (*52*). All of the mutants expressed well and were soluble. CD spectroscopy indicates that all are folded, and the far-UV CD spectra of all of the mutants are very similar to that recorded for DM HP36 and HP36 (Figure 2). DM HP36 contains a set of ring current-shifted methyl resonances that are very diagnostic of its fold. Thus, ^{1}H NMR spectra provide a good probe of the integrity of the tertiary structure. However, the three buried Phe residues are responsible for the ring current shifts; thus, Phe to Leu mutants may exhibit only a subset of the ring current-shifted peaks yet still be folded. As expected, no ring current-shifted methyls were observed in the ^{1}H NMR spectra of the DM F47L/F51L/F58L triple mutant; however, strongly upfield shifted methyl peaks were observed for each single Phe to Leu mutant and for the DM F47L/F51L, DM F47L/F58L, and DM F51L/F58L double mutants. HP36 contains one Trp residue at position 64. The indole ring is partially exposed to solvent; however, one face is packed against the body of the protein, and the indole N−H resonance appears at the characteristic chemical shift of 10.5 ppm in the folded state. All of the Phe mutants display a single peak for this proton located between 10.4 and 10.5 ppm, providing further evidence that the proteins are well-folded. In addition, the peaks in the aromatic−amide region of the spectra are well-dispersed as expected for a folded protein (Figure 3). The ^{1}H NMR spectrum of the W64L mutant displays the characteristic ring current-shifted resonances observed in the wild type, indicating that core packing is not disrupted. Previous work reported the characterization of a subset of Phe to Leu mutants in a M53L background and showed that the single mutants were monomeric in solution (*42*). Some potential association was detected for the F47L/F58L and F51L/F58L double mutants in this system; however, the M53L background is even less stable than the wild type, and thus, it was not clear if the hydrodynamic behavior of those mutants represented a rapid equilibrium between the folded and unfolded states or weak self-association in their respective native states. Consequently, we conducted analytical centrifugation (AUC) studies of each double mutant and the DM F47L/F51L/F58L triple mutant. Apparent molecular weights were determined from a single-species fit of data collected from a 50 μM sample. The apparent molecular weight determined from a single-species fit is 4920 for DM F47L/F51L, 4360 for DM F51L/F58L, and 4640 for DM F47L/F58L, and the expected weight is 4082. The apparent molecular weight for the triple mutant is 4840, and the expected weight is 4048. The apparent molecular weights deduced from a single-species fit to the AUC data are thus 8−20% higher than the true monomeric weights, indicating only weak self-association at worst. The fact that gel filtration data collected for F47L/F58L and F51L/F58L in the destabilized M53L background gave apparent molecular weights of ~7500 reflects the fact that the double Leu mutants likely lead to a significant population of unfolded protein in this system (*42*).

Having established that all of the mutants are folded and well-behaved in solution, we turned to stability measurements to characterize the effects of the aromatic to Leu substitutions. All proteins displayed sigmoidal thermally induced unfolding transitions which can be readily fit to the standard expression for two-state unfolding with linear pre- and post-transitions. There are several practical issues which are worth noting. First, the DM HP36 pseudo-wild type is so stable that it is not fully unfolded even at temperatures approaching 100 °C. A variant with a lower thermal stability could be used as the background for making core mutations; however, that leads to the problem that the most destabilizing mutants are not well folded, and thus, use of DM HP36 is a necessary compromise. Nevertheless, the T_m of DM HP36 has been previously estimated from the derivative of the thermal unfolding curve to be 90.6 °C. A second issue is that small proteins normally have broad unfolding curves because of the small $\Delta H°$ of unfolding. This is the case for HP36.

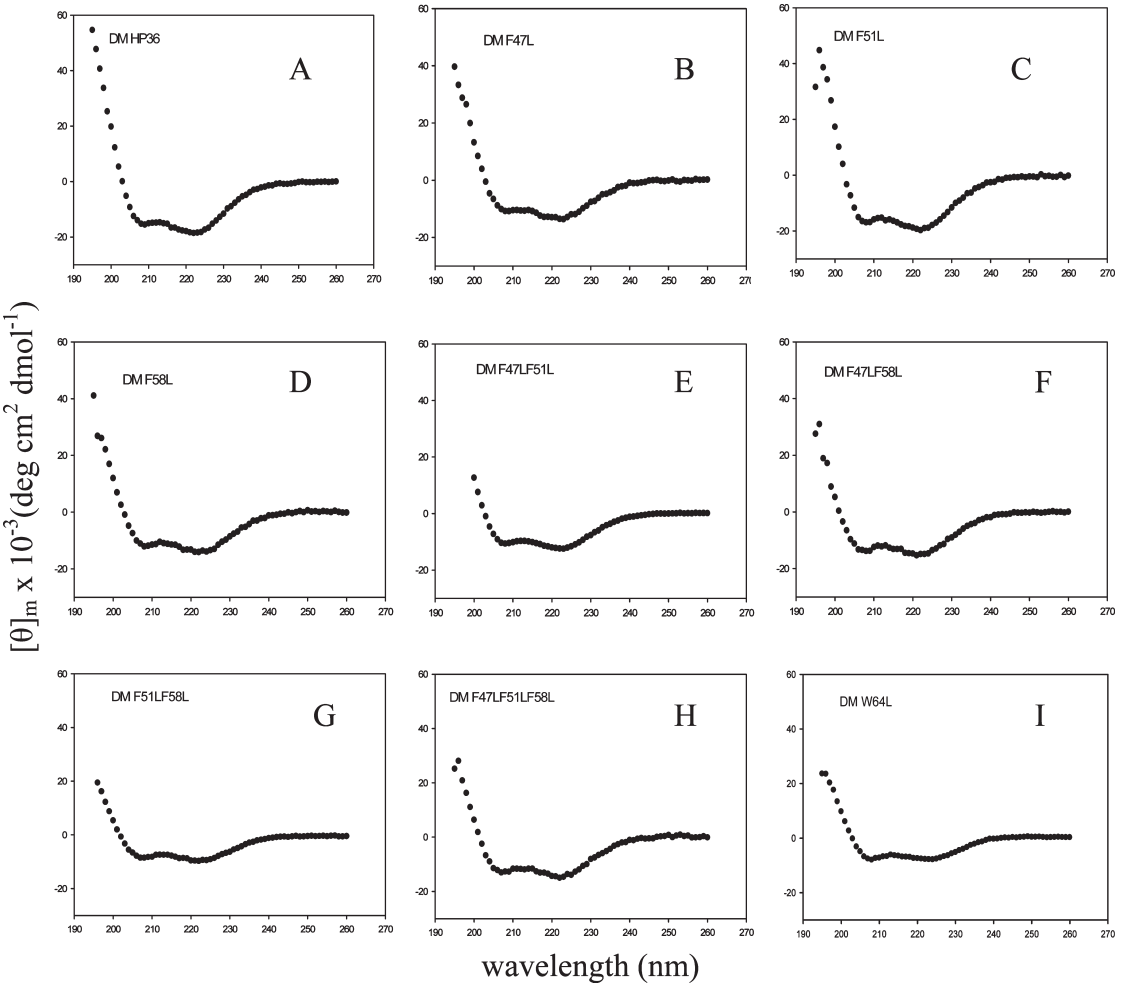

FIGURE 2: Far-UV CD spectra show all mutants are folded: (A) DM HP36, (B) DM F47L, (C) DM F51L, (D) DM F58L, (E) DM F47L/F51L, (F) DM F47L/F58L, (G) DM F51L/F58L, (H) DM F47L/F51L/F58L, and (I) DM W64L. All spectra were recorded at 25 °C in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0.

Furthermore, mutations which lead to core packing that is less optimal than that of the wild type should have even smaller $\Delta H°$ values of unfolding. Thus, some of the mutant unfolding curves are relatively broad, but again all can be reliably fit to yield T_m and $\Delta H°(T_m)$. The thermal unfolding curves are shown in Figure 4 as plots of CD signal versus T. We display the data in this way instead of as plots of fraction folded because using fraction folded can disguise problems with baselines, particularly for small proteins. Showing the actual fits gives a more objective view of the data. A more subtle but more fundamental point is that the analysis is based on a global parameter, mean residue ellipticity, which reports on the overall properties of the molecule. Noncoincidental unfolding curves have been observed for some helical proteins in NMR-monitored unfolding studies even when sigmoidal melting curves are observed using global probes, although the basis for the observed effects is extremely controversial (*53*, *54*). Unfortunately, the unfolding transitions of DM HP36 or its mutants cannot be followed throughout the entire unfolding curve by NMR due to a combination of high T_m values, fast exchange between folded and unfolded forms at higher temperatures, and issues of spectral resolution. Thus, we use CD-monitored unfolding as a probe of the global integrity of the fold and use the measured T_m values as a semiquantitative probe of the effects of the mutations on stability (Figure 4).

Each single Phe to Leu mutation leads to a less thermal stable subdomain as does the Trp to Leu mutation. The F51L mutation has the smallest effect, reducing the apparent T_m by approximately 15 °C, while the F47L or F58L mutant clearly have larger effects, reducing the T_m by 21 or 30 °C, respectively. The rank order observed here agrees with that observed by McKnight and colleagues in the M53L background (*42*). It is interesting to note that substitution of F58 has a greater effect than substitution of F51 even though F51 is sandwiched by the two other Phe rings. In fact, the F51L mutation has the smallest effect on T_m of any of the Phe mutants. The effect of the Trp to Leu mutant on T_m is similar to the effect of F51L, confirming that W64 makes stabilizing interactions as noted previously (*38*) even though it is a surface residue. Precise thermodynamics parameters could not be obtained for the W64 substitution in prior investigations because the mutation led to incompletely defined native baselines in the background used. In contrast, the DM HP36 background leads to well-defined baselines for the Trp mutation. The same trends detected in the single mutants are observed in the double mutants. Namely, any double mutant which includes the F58L substitution is very destabilizing; the apparent T_m values are reduced by more than 40 °C, while the T_m of the F47L/F51L mutant, i.e., the double mutant which retains Phe58, has a $T_m \geq 20$ °C higher than those of the other double mutants. The thermodynamics of the unfolding of both of the most destabilized Phe double mutants could not be characterized in earlier studies because of the lower stability of the pseudo-wild-type background used. The data presented here show unambiguously that

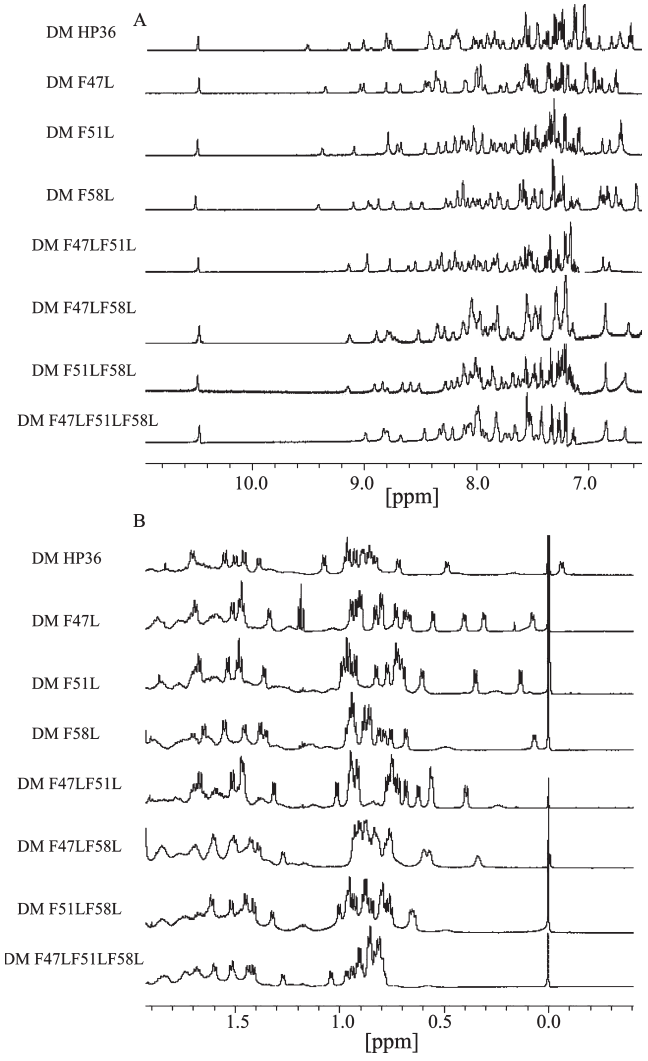

FIGURE 3: ^{1}H NMR spectra confirm that all mutants are folded. (A) Stacked plot of the downfield region of the spectra of the DM HP36 mutants. Spectra are labeled to denote the mutant to which they correspond. (B) Stacked plot of the upfield region of the spectra of DM HP36 mutants. The sharp peak at 0.00 ppm is the chemical shift standard. Spectra were recorded at 25 °C in 90% H_2O, 10% D_2O, 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0.

the deleted phenylalanines are not required for specifying the fold. Interestingly, the triple Phe to Leu mutant DM F47L/F51L/F58L is folded and is no less thermally stable than the F47L/F58L or F51L/F58L double mutant. In fact, the T_m of the triple mutant is actually slightly higher than the T_m of the F51L/F58L double mutant and is comparable to that of the F47L/F58L double mutant. This may reflect the fact that the more flexible Leu side chains can slightly repack in the core of the triple mutant relative to double mutants which contain one Phe. The triple leucine mutant has not been previously characterized, and the results are very interesting since they directly demonstrate that the Phe residues are not required for folding or for the generation of a cooperative (sigmoidal) thermal unfolding transition. Thus, while packing among the three phenylalanines is stabilizing, aromatic−aromatic interactions are not required for HP36 to fold and Trp−Pro interactions are also not required.

The fits to the thermal unfolding curves also yield estimates of the enthalpy of unfolding at T_m, $\Delta H°(T_m)$. Since the different proteins have different T_m values, $\Delta H°(T_m)$ values cannot be meaningfully compared to each other. However, $\Delta H°$ can be estimated at a common temperature provided $\Delta C_p°$ is known. pH-dependent studies of wild-type HP36 provide an estimated $\Delta C_p°$ of 0.38 kcal mol^{-1} deg^{-1}, while analysis of the $\Delta H°$ of surface mutants gives an estimated $\Delta C_p°$ of 0.23 kcal mol^{-1} deg^{-1} (*26*). Clearly, there is considerable uncertainty in the estimation of $\Delta C_p°$. Furthermore, there is no reason that the value of $\Delta C_p°$ for the Phe to Leu mutants need be exactly the same. For example, the Phe to Leu mutants might alter residual structure in the unfolded state which could lead to change in $\Delta C_p°$ even in the absence of native state effects (*7*). Table 1 summarizes the apparent $\Delta H°$ values at 50 °C, a temperature chosen because it involves a shorter extrapolation than 25 or 37 °C and hence reduces any potential systematic error associated with uncertainty in $\Delta C_p°$. For these reasons, we are hesitant to overinterpret the $\Delta H°$ values; however, we note that replacement of Phe58 has the largest effect among the single mutants, and the double mutants which include F58L also show a larger effect than double mutants which do not. Again, the importance of F58 is highlighted. The conclusion that the F58L mutant has the largest effect on $\Delta H°$ is independent of the choice of $\Delta C_p°$ over the range of values considered. The Trp64 to Leu mutation has the smallest effect upon $\Delta H°$ which is not surprising given that it lies on the surface of the protein.

We also conducted urea-induced unfolding studies of all of the mutants. Some of the issues that were relevant for the thermal unfolding come into play here as well. First, the DM HP36 protein is too stable to give a complete urea unfolding curve, but stability can be estimated using the midpoint, C_M, determined from the derivative of the curve and the wild-type *m* value. In addition, the stability of this protein has been examined previously by amide H−D exchange (*26*). All of the mutants give complete unfolding curves; however, the pretransition baselines are not as optimally defined for the two most destabilizing double mutants or the triple mutant, although clear sigmoidal curves are measured. We note this is part of the motivation for using DM HP36 as the background. A less stable "wild type" would lead to very poorly defined transitions for the destabilizing mutants. This has plagued all prior attempts to analyze the role of the aromatic residues and prevented a complete description of the role of the aromatic triad. The parameters deduced from the urea unfolding, C_M values, apparent $\Delta G°$ values in the absence of urea, and *m* values are listed in Table 1, which also summarizes the thermal unfolding data. Of the single Phe mutants, both F47L and F51L have much smaller effects than F58L. The effect of the W64L mutation is also much smaller than that of the F58L substitution. This is entirely consistent with the thermal unfolding data. Fits to the urea unfolding curves also provide *m* values which are formerly equal to $\partial\Delta G°/\partial$[urea]. *m* values are generally believed to be related to the change in solvent accessible surface area between the folded and unfolded forms, and changes in *m* values have been proposed to reflect, at least in some cases, changes in the unfolded state ensemble (*55*, *56*). The *m* value for wild-type HP36 has been determined to be 0.52 kcal mol^{-1} M^{-1} under these conditions, but the *m* value for the hyperstable DM HP36 could not be determined because the protein is too stable. The *m* values of the triple mutant and F47L/F58L and F51L/F58L double mutants are similar but are all smaller than the *m* value for the F47L/F51L double mutant which retains F58. The *m* value of DM W64L is similar to that of the wild type. *m* values are less reliably extracted from fitting broad unfolding curves, and thus, we avoid attempting a molecular-level interpretation of the changes. In addition, recent theoretical work

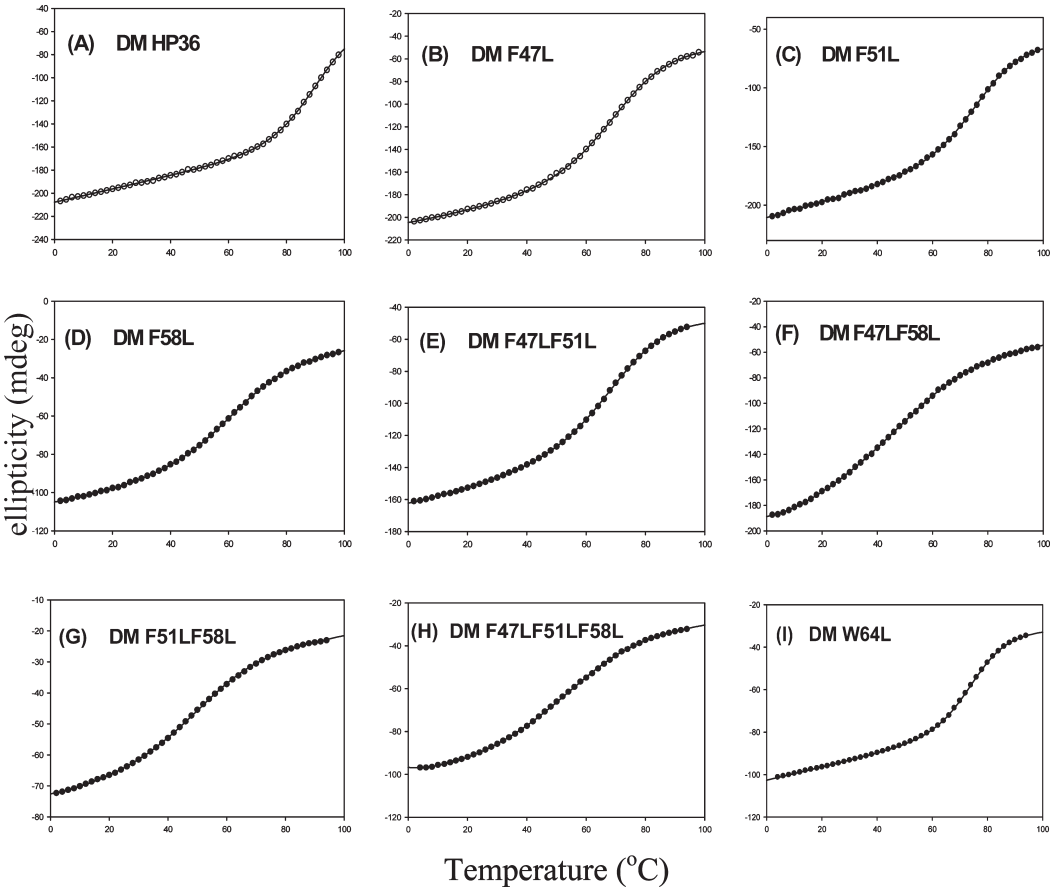

FIGURE 4: CD-monitored thermal unfolding curves for DM HP36 and its mutants: (A) DM HP36, (B) DM F47L, (C) DM F51L, (D) DM F58L, (E) DM F47L/F51L, (F) DM F47L/F58L, (G) DM F51L/F58L, (H) DM F47L/F51L/F58L, and (I) DM W64L. Samples contained between 15 and 30 μM protein in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0. The CD signal at 222 nm was detected.

suggests that urea-induced unfolding may be more complex than often assumed (*57*).

Analysis of the urea-induced unfolding of the double mutants yields conclusions which are similar to those obtained from the thermal unfolding; namely, double mutants which remove F58 are drastically destabilized relative to the double mutant which retains F58. In addition, the triple Phe to Leu mutant is no less stable than either of the double mutants which lack F58. In fact, it is slightly more stable. Mutations can also affect residual interactions in the denatured state ensemble and contribute to changes in stability via denatured state effects. This may be relevant for HP36 since studies with peptide fragments provide evidence of hydrophobic clusters in the unfolded state ensemble which involves some of the Phe residues and which can be partially disrupted by mutation to Leu (*36*, *37*). How might such effects impact the interpretation of the experimentally determined changes in T_m and $\Delta G°$? Clearly, the conclusion that interactions involving the aromatic residues are not required for folding is not altered. Transiently populated hydrophobic clustering in the unfolded state is expected to be relatively weak energetically, but it should stabilize the unfolded state ensemble. Thus, disrupting such interactions in the absence of any native state effects would stabilize the protein because the free energy of the unfolded state would be increased. The fact that all of the mutations studied here significantly destabilize the protein argues that native state effects dominate. In addition, it is worth noting that the same rank order of the effect of single Phe mutations is observed in the DM background, in the single-Phe mutant backgrounds, and in the M53L background analyzed previously, which strongly suggests that native state effects dominate and the rank ordering of stability is robust.

Analysis of $\Delta\Delta G°$ values for the single and double mutants or examination of ΔT_m values provides a measure of the effect of mutation(s) upon stability but does not provide direct evidence of energetically significant interaction between the aromatic residues. Double mutant cycle analysis was developed to provide a method for measuring the interaction free energy, $\Delta\Delta G°_{int}$, between two residues by eliminating the effects of mutations upon other interactions (*58*).

$$\Delta\Delta G°_{int}(A,B) = \Delta G°_{WT} - \Delta G°_{A} - \Delta G°_{B} + \Delta G°_{AB} \quad (2)$$

where $\Delta G°_{WT}$ is the free energy of unfolding of the wild type, $\Delta G°_{A}$ and $\Delta G°_{B}$ are the free energies of unfolding of the respective single mutants, and $\Delta G°_{AB}$ is the unfolding free energy of the double mutant. The double mutant cycle does not provide the contribution that a particular pairwise interaction makes to the native state because it ignores or neglects two important effects. First, the analysis of double mutant cycles makes the inescapable assumption that the mutations do not alter the free energy of the unfolded ensemble. Double mutant cycles can be analyzed in the conceptual context of a double mutant cycle "square" or thermodynamic cycle if unfolded state effects are not important. However, a thermodynamic cube is required if unfolded state effects are present. In this case, knowledge of changes in the free energy of both the folded states and unfolded states of the three mutants is needed, and this is generally impossible to obtain. This effect may be important for HP36 since the subdomain appears to have a partially structured

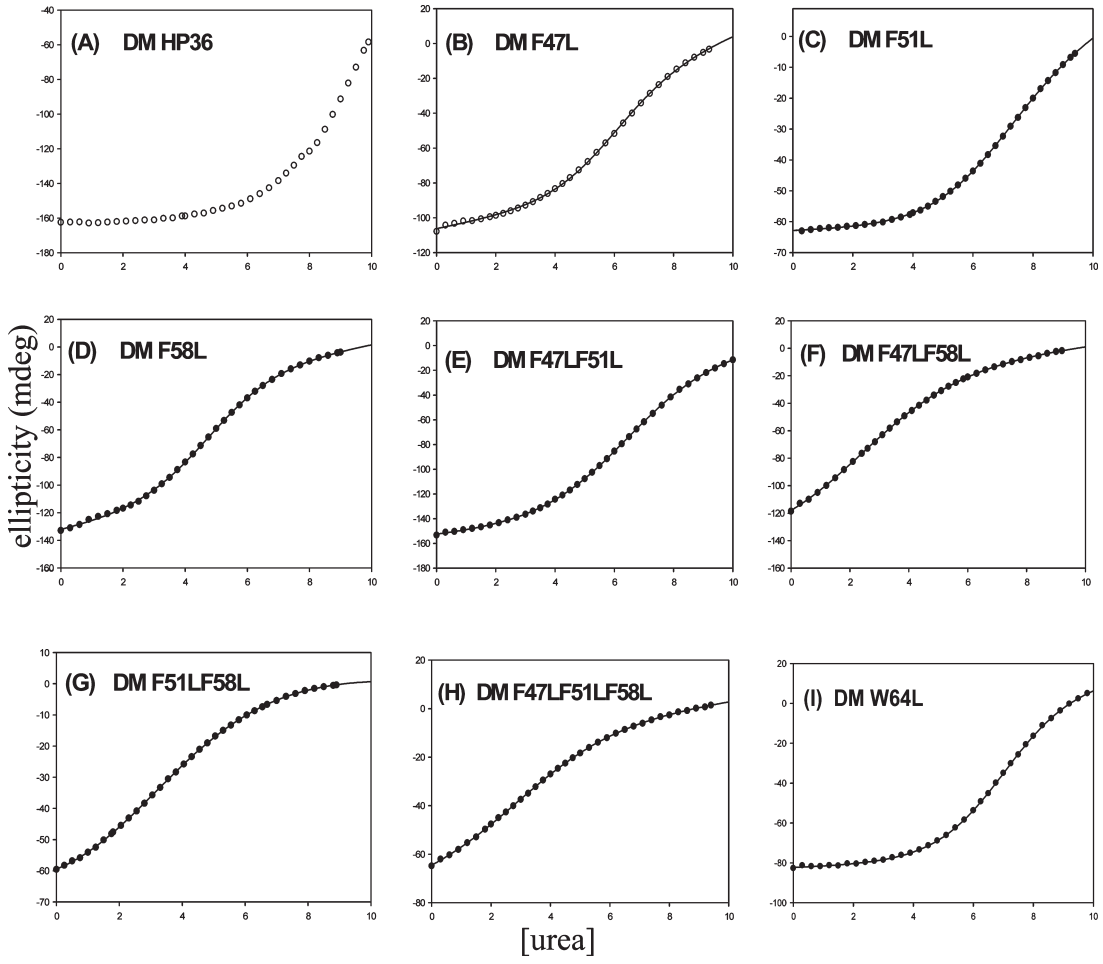

FIGURE 5: CD-monitored urea-induced unfolding curves for DM HP36 and its mutants: (A) DM HP36, (B) DM F47L, (C) DM F51L, (D) DM F58L, (E) DM F47L/F51L, (F) DM F47L/F58L, (G) DM F51L/F58L, (H) DM F47L/F51L/F58L, and (I) DM W64L. Samples contained between 15 and 30 μM protein in 10 mM sodium acetate and 150 mM sodium chloride (pH 5.0). The CD signal at 222 nm was detected.

Table 1: Thermodynamic Parameters for the Unfolding of DM HP36 Mutants

protein	T_m (°C)	$\Delta H°(T_m)$ (kcal/mol)	$\Delta H°$(50 °C) (kcal/mol) ($\Delta C_p° = 0.23$)	$\Delta H°$(50 °C) (kcal/mol) ($\Delta C_p° = 0.38$)	$\Delta G°$ (kcal/mol)	m value (kcal mol^{-1} M^{-1})	C_M (M)
DM HP36	90.6[a]	–[b]	25.6	19.5	4.94–5.06[c]	–[d]	9.60[e]
DM F47L	69.6	25.9	21.4	18.4	2.85	0.46	6.00
DM F51L	76.0	28.4	22.4	18.5	3.01	0.44	7.00
DM F58L	60.9	21.3	18.8	17.2	2.62	0.55	4.75
DM F47L/F51L	68.1	23.2	19.0	16.3	2.54	0.40	6.50
DM F47L/F58L	48.9	13.3	13.6	13.7	0.75	0.34	2.85
DM F51L/F58L	46.4	15.6	16.4	17.0	0.72	0.29	3.30
DM F47L/F51L/F58L	49.8	13.3	13.3	13.4	1.02	0.34	3.50
DM W64L	74.4	36.6	31.0	27.3	3.58	0.52	6.75

[a] Determined from the derivative of the curve of the CD signal vs temperature. [b] Not determined because of an incomplete unfolding transition. [c] Calculated from the C_M value determined by numerical differentiation of the curve together with the wild-type m value. Data for DM H36 are from ref *26*. [d] Not determined because of an incomplete unfolding transition. [e] Determined by numerical differentiation of the plot of CD signal vs urea concentration.

denatured state ensemble in which the aromatic residues may be involved in hydrophobic clusters (*7*, *36*, *37*). The second important effect is that the double mutant cycle cancels desolvation effects to first order (*59*). Nonetheless, the approach is still powerful and in this case should be able to deduce which pairs of aromatic residues interact most strongly. Table 2 summarizes the experimental interaction free energies determined using $\Delta G°$ values from urea denaturation studies. The interaction free energies between F47 and F58 and between F51 and F58 are both small, while the interaction energy between F47 and F51 is large. The same result is obtained if the interaction free energy between F47 and F51 is measured in a DM F58L background. The fact that a self-consistent rank order is obtained independent of the background used argues that, while unfolded state effects may play a role, the qualitative conclusion that the F47–F51 interaction is stronger than the native state F47–F58 or F51–F58 interactions is likely to be robust.

Table 2: Double Mutant Cycle Analysis of Apparent Interaction Free Energies[a]

background	ΔΔ*G*°(F47,F51) (kcal/mol)	ΔΔ*G*°(F47,F58) (kcal/mol)	ΔΔ*G*°(F51,F58) (kcal/mol)
DM HP36	1.68	0.28	0.09
DM F47L	−	−	0.58
DM F51L	−	0.77	−
DM F58L	2.17	−	−

[a] Δ*G*° values from the urea unfolding studies were used. Experimental data were collected at 25 °C and pH 5.0 in 10 mM sodium acetate and 150 mM sodium chloride.

DISCUSSION

By using a hyperstable double mutant which retains the same fold and core packing as the wild-type subdomain, we have been able to define the role of the core phenylalanine residues in HP36 and to probe the role of a conserved surface aromatic residue. Previous attempts to do so have been hindered by the low stability of the wild-type subdomain. Although mutation of any one of the core phenylalanine results in a decrease in stability, none are critical for folding and even a triple Phe to Leu mutant folds. The W64L mutant confirms the important role proposed for this residue but indicates that the aromatic−proline interaction is not strictly required to specify the fold (*38*). F58 is consistently found to play a larger role than either F47 or F51 in the content of both single and double mutants. These conclusions are independent of whether thermal or urea unfolding data are used to assess stability. It is interesting to note that mutation of F58 clearly has a larger effect than mutation of F51 even though F51 is the central residue in the triad of phenylalanines. This likely reflects the fact that F58 makes contacts with a larger number of other residues, including F47, F51, M53, A57, A59, L61, Q66, and M70 while F51 is involved in long-range contacts with a smaller number of residues, F47, V50, G52, F58, M70, and K73. The double mutant cycle analysis implies a significant favorable interaction between F47 and F51. Interaction free energies involving F58 and the other phenylalanines are much smaller even though mutation of F58 has the largest effect upon stability. Again, this likely reflects the fact that F58 makes more extensive contacts with other hydrophobic core residues than with F47 or F51. The important role of F58 is broadly consistent with MD simulations which have examined the folding of HP36 (*30*). F58 is located near the N-terminus of helix 2 and make contacts with both helix 1 and helix 3. The folding scenario derived from the MD studies is that helix 1 folds after helix 2. In this case, a correctly positioned F58 in helix 2 will help to stabilize helix 1.

The data presented here clearly demonstrate that aromatic−proline, aromatic−hydrophobic, and aromatic−aromatic interactions, while important for stability, are not required to specify the overall fold of the villin headpiece helical subdomain. Many, but not all, small cooperatively folded domains exhibit aromatic−aromatic or aromatic−proline interactions, and this has led to some speculation that they may be critical determinants of "mini-protein" folding (*43*). Clearly, this is not the case for HP36. It is interesting to speculate if unusual interactions are generally required to stabilize miniature protein domains. We believe that this is likely not the case at least for proteins the size of HP36 or larger since there are several examples, including HP36, in which aromatic−aromatic or aromatic−proline interactions are not critical for specifying the fold (*60*, *61*). Important early studies on protein thermodynamics led to the suggestion that proteins needed to exceed a minimum size, estimated to be 50 residues, to be cooperatively folded (*62*). There are obviously a number of examples of proteins that fold cooperatively yet are much smaller, and this has promoted the idea that small cooperatively folded proteins require special interactions. However, it is important to remember that the early thermodynamic analysis was designed to calculate the minimum size required so that the Δ*G*° of folding was significantly larger than *RT*. In fact, all miniature proteins characterized to data have Δ*G*° values which are only several times *RT* (*63*). Futhermore, their thermodynamic properties are predicted well by databases derived from larger proteins (*62*, *64*). Thus, there is no inherent contradiction between the early thermodynamic analysis and the fact the proteins of ≤40 residues can be found which fold cooperatively. Of course, we do not want to imply that aromatic−aromatic or aromatic−proline interactions may not be required in specific cases, but they are clearly not mandatory in all cases. We hope that the data presented here will provide additional important benchmarks for theoretical and computational studies of the folding of HP35 and HP36.

ACKNOWLEDGMENT

We thank Mr. Wenli Meng for helpful discussions and Mr. Vadim Patsalo for his help with the contact map calculation. We also thank Professor C. J. McKnight for his advice and continued interest in this project.

SUPPORTING INFORMATION AVAILABLE

Plots of AUC data for DM F47L/F51L, DM F47L/F58L, DM F51L/F58L, and DM F47L/F51L/F58L, a table of apparent molecular weights determined by AUC, and the upfield region of the ^{1}H NMR spectrum of DM W64L. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. 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*, 126–134.
2. McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Biol. 4*, 180–184.
3. 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*, 1299–1310.
4. 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*, 7517–7522.
5. 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*, 6032–6033.
6. Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations. *J. Mol. Biol. 329*, 625–630.
7. 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*, 16662–16667.
8. 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.
9. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science 282*, 740–744.
10. Duan, Y., Wang, L., and Kollman, P. A. (1998) The early stage of folding of villin headpiece subdomain observed in a 200-ns fully

solvated molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S. A. 95*, 9897–98902.
11. Sullivan, D. C., and Kuntz, I. D. (2002) Protein folding as biased conformational diffusion. *J. Phys. Chem. B 106*, 3255–3262.
12. Shen, M. Y., and Freed, K. F. (2002) All-atom fast protein folding simulations: The villin headpiece. *Proteins: Struct., Funct., Genet. 49*, 439–445.
13. Srinivas, G., and Bagchi, B. (2002) Folding and unfolding of chicken villin headpiece: Energy landscape of a single-domain model protein. *Curr. Sci. 82*, 179–185.
14. 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*, 664–671.
15. 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*, 927–937.
16. 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*, 153–164.
17. Lin, C. Y., Hu, C. K., and Hansmann, U. H. (2003) Parallel tempering simulations of HP-36. *Proteins 52*, 436–445.
18. 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*, 4005–4017.
19. 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*, 14841–14836.
20. 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*, 11178–11187.
21. 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*, 915–925.
22. Kim, S. Y., Lee, J., and Lee, J. (2005) Folding simulations of small proteins. *Biophys. Chem. 115*, 195–200.
23. 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*, 459–471.
24. 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*, 12267–12270.
25. Godoy-Ruiz, R., Henry, E. R., Kubelka, J., Hofrichter, J., Munoz, V., Sanchez-Ruiz, J. M., and Eaton, W. A. (2008) Estimating free energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data. *J. Phys. Chem. B 112*, 5938–5949.
26. Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H., and Raleigh, D. P. (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain. *Biochemistry 46*, 7497–7505.
27. Ensign, D. L., Kasson, P. M., and Pande, V. S. (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.
28. Hansmann, U. H. E. (2002) Protein-folding simulations in generalized ensembles. *Int. J. Quantum Chem. 90*, 1515–1521.
29. Jayachandran, G., Vishal, V., Garcia, A. E., and Pande, V. S. (2007) Local structure formation in simulations of two small proteins. *J. Struct. Biol. 157*, 491–499.
30. Lei, H. X., and Duan, Y. (2007) Two-stage folding of HP-35 from ab initio simulations. *J. Mol. Biol. 370*, 196–206.
31. Micheletti, C., and Colombo, G. (2008) Study of the Villin Headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. *Proteins: Struct., Funct., Bioinf. 70*, 309.
32. Mukherjee, A., and 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.
33. Piana, S., Laio, A., Marinelli, F., Van Troys, M., Bourry, D., Ampe, C., and Martins, J. C. (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.
34. Wei, Y., Nadler, W., and Hansmann, U. H. E. (2008) Backbone and side-chain ordering in a small protein. *J. Chem. Phys. 128*, 025105.
35. Yang, J. S., Wallin, S., and Shakhnovich, E. I. (2008) Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc. Natl. Acad. Sci. U.S.A. 105*, 895–900.
36. 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*, 3264–3272.
37. Tang, Y. F., Goger, M. J., and Raleigh, D. P. (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.
38. Vermeulen, W., Troys, M. V., Bourry, D., Dewitte, D., Rossenu, S., Goethals, M., Borremans, F. A. M., Vandekerckhove, J., Martins, J. C., and Ampe, C. (2006) Identification of the PXW sequence as a structural gatekeeper of the headpiece C-terminal subdomain fold. *J. Mol. Biol. 359*, 1277–1292.
39. Glasscock, J. M., Zhu, Y., Chowdhury, P., Tang, J., and Gai, F. (2008) Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: Application to the villin headpiece subdomain and the LysM domain. *Biochemistry 47*, 11070–11076.
40. Cellmer, T., Henry, E. R., Hofrichter, J., and Eaton, W. A. (2008) Measuring internal friction of an unltrafast-folding protein. *Proc. Natl. Acad. Sci. U.S.A. 105*, 18320–18325.
41. Brewer, S., Song, B., Raleigh, D. P., and Dyer, R. B. (2007) Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. *Biochemistry 46*, 3279–3285.
42. Frank, B. S., Vardar, D., Buckley, D. A., and McKnight, C. J. (2002) The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci. 11*, 680–687.
43. Gellman, S. H., and Woolfson, D. N. (2002) Mini-proteins Trp the light fantastic. *Nat. Struct. Biol. 9*, 408–410.
44. Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002) Designing a 20-residue protein. *Nat. Struct. Biol. 9*, 425–430.
45. Cocharan, A. G., Skelton, N. J., and Starovasnik, M. A. (2001) Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci. U.S.A. 98*, 5578–5583.
46. Bhattacharyya, R., and Chakrabarti, P. (2003) Stereospecific interactions of proline residues in protein structures and complexes. *J. Mol. Biol. 331*, 925–940.
47. Tracz, S. M., Abedini, A., Driscoll, M., and Raleigh, D. P. (2004) Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry 43*, 15901–15908.
48. Gazit, E. (2002) A possible role for π-stacking in the self-assembly of amyloid fibrils. *FASEB J. 16*, 77–83.
49. Doering, D. S., and Matsudaira, P. T. (1996) Cysteine scanning mutagenesis at 40 of 76 position in villin headpiece maps the F-actin binding site and structural features of the domain. *Biochemistry 35*, 12677–12685.
50. Wickstrom, L., Bi, Y., Hornak, V., Raleigh, D. P., and Simmerling, C. (2007) 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 46*, 3624–3634.
51. Lattman, E. E., and Rose, G. D. (1993) Protein folding: What's the question?. *Proc. Natl. Acad. Sci. U.S.A. 90*, 439–441.
52. Richards, F. M., and Lim, M. (1994) An analysis of packing in the protein folding problem. *Q. Rev. Biophys. 26*, 423–498.
53. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M., and Munoz, V. (2002) Experimental identification of downhill protein folding. *Science 298*, 2191–2195.
54. Ferguson, N., and Fersht, A. R. (2003) Early events in protein folding. *Curr. Opin. Struct. Biol. 13*, 75–81.
55. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins. *Annu. Rev. Biochem. 60*, 795–825.
56. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138–2148.
57. Hua, L., Zhou, R., Thirumalai, D., and Berne, B. J. (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. U.S.A. 105*, 16928–16933.
58. Horovitz, A., and Fersht, A. R. (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol. 214*, 613–617.
59. Luisi, D. L., Snow, C. D., Lin, J., Hendsch, Z. S., Tidor, B., and Raleigh, D. P. (2003) Surface salt bridge, double-mutant cycle, and protein stability: An experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein L9. *Biochemistry 42*, 7050–7060.

60. Horng, J., Moroz, V., Rigotti, D. J., Fairman, R., and Raleigh, D. P. (2002) Characterization of large peptide fragments derived from the N-terminal domain of the ribosomal protein L9: Definition of the minimum folding motif and characterization of local electrostatic interactions. *Biochemistry 41*, 13360–13369.
61. Spector, S., Young, P., and Raleigh, D. P. (1999) Native like structure and stability in a truncation mutant of a protein minidomain: The peripheral subunit-binding domain. *Biochemistry 38*, 4128–4136.
62. Privalov, P. L. (1989) Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem. 18*, 47–69.
63. Horng, J., Moroz, V., and Raleigh, D. P. (2003) Rapid cooperative two-state folding of a miniature α-β protein and design of a thermostable variant. *J. Mol. Biol. 326*, 1261–1270.
64. Rees, D. C., and Robertson, A. D. (2001) Some thermodynamic implications for the thermostability of proteins. *Protein Sci. 10*, 1187–1194.