J. Mol. Biol. (1996) **260**, 126–134

JMB

COMMUNICATION

A Thermostable 35-Residue Subdomain within Villin Headpiece

C. James McKnight[1,2]*, Don S. Doering[2], Paul T. Matsudaira[2] and Peter S. Kim[1,2]

[1]*Howard Hughes Medical Institute and* [2]*Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology Nine Cambridge Center Cambridge, Massachusetts MA 02142, USA*

**Corresponding author*

The actin-bundling protein villin contains, at its extreme C terminus, a compact f-actin binding domain called ''headpiece''. This 76-amino acid domain from chicken is highly thermostable. Here, we show that the stable folded structure in headpiece is localized to a subdomain formed by the C-terminal 35 residues. The subdomain, denoted HP-35, is monomeric and retains high thermostability, with a T_m of 70($\pm$1) °C at pH 7.0. There are no cysteine residues in HP-35 and its folding is not dependent on the binding of metals or other ligands. HP-35 is not a molten globule, but instead, has properties expected for a fully folded protein with a unique structure. In particular, the slowly exchanging amide protons in HP-35 have protection factors that are slightly larger than those predicted if exchange occurred only from globally unfolded molecules. NMR studies indicate that the headpiece subdomain contains three short α-helices, and that these same helices are present in the corresponding regions of intact headpiece. HP-35 is the smallest monomeric polypeptide characterized consisting of only naturally occurring amino acids that autonomously folds into a unique and thermostable structure without disulfide bonds or ligand binding.

Keywords: subdomain; thermostability; protein folding; NMR; villin

Subdomains that fold independently are important tools for solving the protein folding problem (Oas & Kim, 1988; Rose, 1979). Subdomains are units of folded structure larger than an isolated helix or sheet, but smaller than an entire domain. The shorter amino acid sequences of subdomains reduce complexity, relative to intact proteins. This relative simplicity facilitates experimental studies and also comparison between experiments and theoretical calculations. The stable structures formed by subdomains may also be important intermediates or initiation sites for protein folding. Here, we describe the identification of a short, monomeric subdomain from chicken villin that folds without disulfide bonds or ligand binding.

The f-actin bundling protein villin is implicated in the formation of microvilli in the absorbtive epithelium of the gut and the proximal tube of the kidney (Bretscher & Weber, 1979; Finidori *et al*., 1992). Villin is composed of two domains defined by partial proteolysis: a ''core'' domain of 84 kDa, and a small, 8 kDa carboxy-terminal domain called ''headpiece'' (Bazari *et al*., 1988; Glenny *et al*., 1981). Each of these domains contains an f-actin binding site, enabling villin to crosslink actin filaments into structural bundles that support the microvillar membrane (Glenny *et al*., 1981).

Present addresses: C. J. McKnight, Department of Biophysics, Center for Advanced Biomedical Research, Boston University School of Medicine, 80 East Concord Street, Boston, MA 02118-2394, USA; D. S. Doering, AquaPharm Technologies Corporation, 9110 Red Branch Road, Columbia, MD 21045, USA.

Abbreviations used: NMR, nuclear magnetic resonance; CD, circular dichroism; NOE, nuclear Overhauser enhancement; TOCSY, total correlation spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; NOESY nuclear Overhauser enhancement spectroscopy; HSQC, heteronuclear single quantum shift correlation; HSMQC, heteronuclear single and multiple quantum shift correlation; HMQC, heteronuclear multiple quantum shift correlation; T_m, transition midpoint temperature; GuHCl, guanidine hydrochloride; PCR, polymerase chain reaction; HPLC, high pressure liquid chromatography; TFA, trifluoroacetic acid; FMOC, 9-fluorenlymethycarbonate; p.p.m., parts per million; 2D, 3D, two- and three-dimensional, respectively.

0022–2836/96/270126–09 $18.00/0

Headpiece binds f-actin in the absence of the core domain (Glenny *et al*., 1981; Pope *et al*., 1994). The modular nature of the 76-amino acid headpiece domain is demonstrated by its presence in the f-actin bundling protein dematin (band 4.9), which has a core domain that is unrelated to villin (Rana *et al*., 1993). The amino acid sequence of headpiece (Bazari *et al*., 1988) is not similar to other known actin-binding motifs (Vanderkerckhove, 1990). The sequence of the chicken villin headpiece is: VFTATTTLVPTKLETFPLDVLVNTAAEDLPRGV-DPSRKENHLSDEDFKAVFGMTRSAFANLPLWK-QQNLKKEKGLF. Residue 1 of headpiece corresponds to residue 750 of intact chicken villin. The underlined residues correspond to the 35-residue subdomain described in this study.

Table 1. Backbone resonance assignments[a] at pH 5.0, 30°C

Headpiece				HP-36			
Residue	N	HN	Hα	Residue	N	HN	Hα
43 Ser				43 Ser	122.32	9.53	4.52
44 Asp	123.88	9.15	4.33	44 Asp	123.62	9.14	4.37
45 Glu	120.65	8.81	4.09	45 Glu	121.06	8.77	4.10
46 Asp	123.85	7.97	4.55	46 Asp	123.86	7.99	4.51
47 Phe	123.13	9.18	3.77	47 Phe	123.06	9.06	3.86
48 Lys	119.42	7.91	4.32	48 Lys	119.93	7.81	4.36
49 Ala	123.61	7.75	4.08	49 Ala	123.41	7.66	4.10
50 Val	118.70	8.10	3.59	50 Val	118.94	7.92	3.55
51 Phe	114.69	8.31	4.36	51 Phe	114.96	8.45	4.23
52 Gly	109.94	8.22	3.99, 3.99	52 Gly	109.87	8.15	4.08, 3.94
53 Met	114.40	7.53	4.85	53 Met	115.14	7.57	4.84
54 Thr				54 Thr	108.64	8.18	4.55
55 Arg	122.08	8.57	3.30	55 Arg	122.25	8.76	3.26
56 Ser	114.96	8.11	4.06	56 Ser	114.96	8.15	4.08
57 Ala	126.62	7.55	4.08	57 Ala	126.52	7.59	4.11
58 Phe	122.39	8.26	4.10	58 Phe	123.14	8.42	4.11
59 Ala	119.15	7.87	3.94	59 Ala	118.86	7.80	3.97
60 Asn	114.85	7.16	4.65	60 Asn	114.79	7.17	4.66
61 Leu	123.29	7.39	4.32	61 Leu	123.20	7.40	4.33
62 Pro				62 Pro			
63 Leu	127.13	8.81	4.32	63 Leu	127.04	8.82	
64 Trp	116.03	7.95	4.38	64 Trp	116.09	7.92	4.40
65 Lys	125.07	6.07	3.66	65 Lys	125.22	6.04	3.66
66 Gln	120.27	7.54	3.38	66 Gln	120.09	7.57	3.44
67 Gln	116.63	8.28	4.00	67 Gln	116.73	8.20	4.01
68 Asn	120.47	7.97	4.45	68 Asn	120.58	7.94	4.52
69 Leu	122.47	8.44	4.26	69 Leu	122.53	8.44	4.28
70 Lys	118.57	8.33	3.91	70 Lys	119.32	8.25	4.07
71 Lys				71 Lys	120.56	8.08	4.13
72 Glu	121.37	8.29	4.08	72 Glu	121.15	8.27	4.08
73 Lys	115.98	7.66	4.31	73 Lys	115.97	7.75	4.31
74 Gly	109.03	7.81	4.02, 3.85	74 Gly	108.90	7.87	4.03, 3.85
75 Leu	121.96	7.96	4.55	75 Leu	121.86	7.78	4.50
76 Phe	125.66	7.68	4.41	76 Phe	125.40	7.56	4.41

NMR data were collected on a Bruker AMX-500 spectrometer at 30°C. ^{15}N-HP-45 and ^{15}N-HP-36 were expressed and purified as described in the legend to Figures 1 and 2, except that cells were grown in M9T minimal media supplemented with 1 g/l ^{15}N-ammonium sulfate (Muchmore *et al*., 1989). The NMR samples used for assignments were 6.6 mM for ^{15}N-headpiece, 4.0 mM for ^{15}N-HP-45, and 3.2 mM for ^{15}N-HP-36. Samples contained 5% $^{2}H_2O$ and the pH was adjusted using a glass electrode with no correction for the effect of $^{2}H_2O$ on the measured pH (Bundi & Wüthrich, 1979).

Partial NMR assignments of 15N-enriched headpiece were made at pH 5.0, 30°C by examination of spectra from 3D (^{1}H, ^{1}H, ^{15}N) NOESY-HSQC (Bodenhausen & Ruben, 1980; Marion *et al*., 1989), 3D (^{1}H, ^{1}H, ^{15}N) TOCSY-HSQC (Bodenhausen & Ruben, 1980; Griesinger *et al*., 1988), and HSQC (Bodenhausen & Ruben, 1980) experiments. NMR assignments for ^{15}N-HP-36 at pH 3.7, and ^{15}N-HP-45 at pH 5.0 (not shown), were based on standard 2D homonuclear methods using DQF-COSY, NOESY and TOCSY spectra (Wüthrich, 1986). Heteronuclear 2D (^{1}H, ^{15}N) spectra collected to aid resonance assignments included HMQC-COSY (Gronenborn *et al*., 1989), HSQC-TOCSY (Bodenhausen & Ruben, 1980; Griesinger *et al*., 1988; Gronenborn *et al*., 1989), and HSMQC-NOESY (Gronenborn *et al*., 1989; Zuiderweg, 1990). The backbone assignments of ^{15}N-HP-36 were transferred from pH 3.7 to pH 5.0 by comparison of the HSQC spectra and then verified with HSMQC-NOESY. HMQC-COSY spectra were recorded to extract H^{α} proton assignments of ^{15}N-HP-35 at pH 5.0.

[a] Chemical shifts (p.p.m.) are relative to 3-(trimethylsilyl)-1-propanesulfonic acid (–0.01 p.p.m.) for ^{1}H, and NH_4Cl (24.93 p.p.m.) for ^{15}N. Unassigned resonances are left blank.

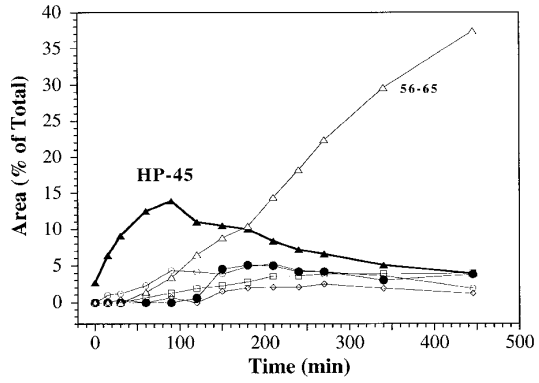


Figure 1. Generation of C-terminal peptides by trypsinolysis of headpiece. Trypsinolysis was carried out in 1 M GuHCl. After quenching at the specified times by addition of acetic acid to 5% (v/v), digestion products were analyzed by reverse-phase high pressure liquid chromatography (HPLC). The data are plotted as a percentage of the total absorbance of all peaks at 229 nm. The 56 to 65 fragment is the complete digestion product. Filled triangles, 32 to 76; filled circles, 38 to 76; open triangles, 56 to 65; open circles, 56 to 73; open squares, 56 to 71; open diamonds, 56 to 70.

Methods: Headpiece was expressed from the plasmid pDDHP10 (Doering, 1992) in *E. coli* using a phage T7 expression system (Studier *et al*., 1990). BL21(DE3)pLysS cells (Novagen) transformed with the desired plasmid were grown in LB media containing ampicillin at 100 μg ml^{-1}. Expression was induced by addition of isopropyl β-D-thiogalactopyranoside (120 μg ml^{-1} final concentration) at an optical density between 0.5 and 0.8 at 590 nm. Cells were harvested by centrifugation two to three hours after induction.

Cell pellets from one liter bacterial cultures were brought up in 25 ml of buffer A (3 M urea, 200 mM NaCl, 0.1 mM EDTA, 50 mM Tris (pH 8.0)) and sonicated with a probe sonicator for five minutes to lyse the cells. The samples were then centrifuged at 180,000 ***g*** for 30 minutes and the supernatants were applied to an ACA34 size exclusion column (1000 cm × 5 cm, Sepracor) running in buffer A. Fractions were collected and assayed by sodium dodecyl sulfate/gel electrophoresis (Schägger & von Jagow, 1987). Fractions containing the desired peptides were pooled, dialyzed against water and then concentrated by lyophilization.

The crude peptide was solublized in 5% acetic acid and purified by reverse-phase HPLC on a C18 column (Vydac, 2 cm × 25 cm) eluted with a water/acetonitrile gradient containing 0.1% (v/v) trifluoroacetic acid (TFA). Peptide identity was confirmed by laser desorption mass spectrometry (Finnigan MAT). Peptide concentrations were determined by UV absorbance at 280 nm (Edelhoch, 1967).

Trypsin digestion was carried out with chymotrypsin-free bovine trypsin. Bovine trypsin (Sigma) was solvated to 5 mg ml^{-1} in 100 mM Tris, 20 mM $CaCl_2$ (pH 7.5), and 10 ml was loaded on to 5 cm × 2 cm soybean trypsin inhibitor-linked agarose column (Pharmacia). The column was washed with 20 ml of running buffer containing tryptamine (2.5 mg ml^{-1}) to elute any contaminating chymotrypsin. The purified trypsin was eluted with 3 mM HCl, 20 mM $CaCl_2$, and stored in aliquots frozen at –20°C.

The digestion reaction contained 100 μM headpiece in 1 M GuHCl, 100 mM ammonium bicarbonate (pH ~9). Trypsin was added to a final concentration of 3 μM. Aliquots were removed at given time points, quenched by addition of glacial acetic acid to a final concentration of 5%, and frozen in liquid nitrogen. The quenched samples were analyzed by reverse-phase HPLC using a C18 column (Vydac, 0.46 cm × 30 cm) and a 1% min^{-1}, water/acetonitrile gradient containing 0.1% (v/v) TFA. Peaks were monitored by absorbance at 229 nm. The molecular masses of the fragments were determined by laser desorption mass spectrometry (Finnigan MAT). Reported peak identities were confirmed by N-terminal sequence analysis (model 470A, Applied Biosystems, Inc.).

Identification of the subdomain

Only partial NMR assignments of headpiece can be obtained readily (Table 1). Resonances from the amino-terminal half of headpiece are split into multiple peaks and/or are in intermediate exchange, indicating the presence of multiple interconverting conformations that may arise, at least in part, from *cis–trans* isomerization about the three proline residues at positions 17, 30 and 35. Sequential resonance assignments are possible for the C-terminal residues of headpiece, starting at Asp44 and continuing through to Phe76, with breaks at residues 54 and 71 due to ambiguities (Table 1). These NMR data suggest the presence of a C-terminal subdomain in headpiece.

To delineate the subdomain, headpiece was subjected to limited proteolysis by trypsin. Trypsinolysis was carried out in 1 M guanidine hydrochloride (GuHCl) to destabilize any residual folded structure that might exist in the N-terminal region of headpiece (Riviere *et al*., 1991). A 45-amino acid fragment, denoted HP-45 and corresponding to residues 32 to 76 of headpiece, is the most populated digestion product, accumulating to 14% of the total sample absorbance at 229 nm by 90 minutes (Figure 1). At later time points, the population of HP-45 decreases, while several smaller C-terminal fragments become populated at low levels. The relative resistance to trypsin digestion of HP-45, which contains eight potential trypsin cleavage sites, argues strongly for the presence of stably folded structure at the C terminus of headpiece.

The thermal unfolding of headpiece, as monitored by circular dichroism (CD), is reversible, with a transition midpoint (T_m) of 74°C (Doering, 1992). This T_m value is the same in the presence of the metal chelator ethylenediaminetetraacetic acid (EDTA) (Figure 2a), indicating that the thermostable structure of headpiece is not dependent on metal binding. The T_m value for the unfolding transition of HP-45 is 72°C, only 2°C lower than for intact headpiece (Figure 2a). Thus, the 31 N-terminal residues of headpiece are not required for the thermostability of HP-45. However, the differences in the slopes in the transition regions of headpiece compared to HP-45 in the thermal unfolding experiment may indicate a significantly higher molar enthalpy for headpiece unfolding than for HP-45. This may indicate that there is a larger difference in

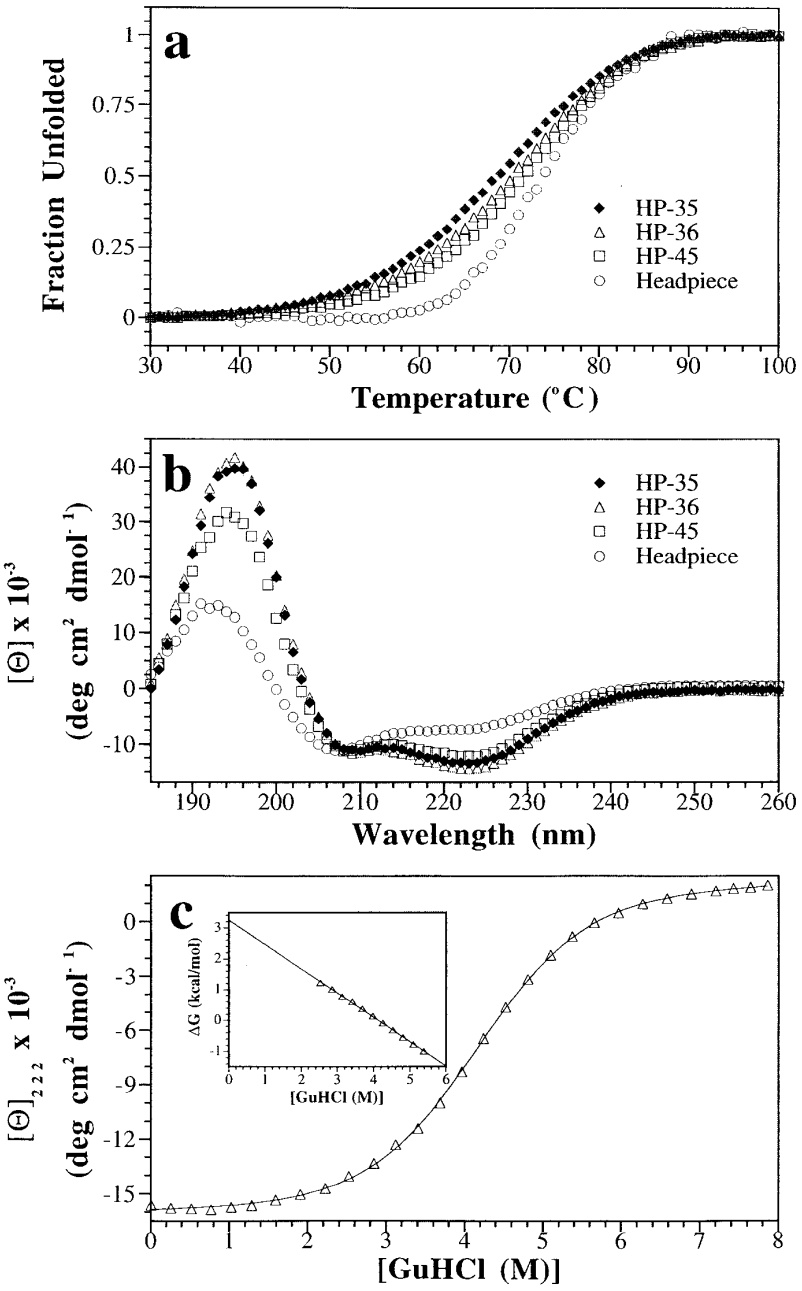


Figure 2. CD measurements of headpiece and C-terminal peptides. a, Thermal unfolding monitored by CD. For ease of comparison, the data have been normalized to fraction unfolded after subtraction of the folded and unfolded baselines. Protein concentrations were 20 μM in 10 mM phosphate (pH 7.0), with 150 mM NaCl and 500 μM EDTA. b, CD spectra of headpiece and headpiece fragments at 25°C. Protein concentrations were 20 μM in 10 mM phosphate (pH 7.0). c, Guanidine hydrochloride denaturation of HP-36 in H_2O. The samples were 20 μM HP-36 in 150 mM NaCl, 25 mM acetate (pH 5.0), at 4°C. The line is the best fit to the equation described by Santoro & Bolen (1988). For this fit $\Delta G = 3.3$ kcal mol^{-1}, and the m value was 800 cal mol^{-1} M^{-1}. Insert shows the linear extrapolation of the ΔG values determined in the transition region to 0 M GuHCl.

Methods: The expression plasmid for HP-45, pVHP32-76, was constructed by PCR from the parent headpiece expression plasmid pDDHP10 (Doering, 1992). Expressed HP-35 retains the additional N-terminal initiator methionine residue, so the resultant peptide is 36 residues long and is referred to as HP-36. An initial expression plasmid for HP-36, pVHP42-76, was constructed by single-stranded mutagenesis of pVHP32-76, deleting the first ten codons after the initiator methionine codon. This plasmid expressed poorly, so the coding region was recloned into the pDDHP10 plasmid backbone (pVHP42-76b). The pVHP42-76b plasmid was found to express the HP-36 fragment at levels sufficient for purification. Recombinant DNA techniques were based on standard protocols (Sambrook *et al*., 1989). Constructs were confirmed by DNA sequencing (Sanger *et al*., 1977).

HP-45 and HP-36 were expressed and purified as described in the legend to Figure 1, except that buffer A was replaced by 50 mM phosphate buffer (pH 7.0), and the ACA34 column was replaced by a Sephadex G-50 column (1000 × 2.5 cm) run in 50 mM phosphate. The synthetic peptide HP-35 was synthesized using standard solid phase FMOC peptide synthesis protocols on an Applied Biosystems model 430A automated peptide synthesizer. The peptide was cleaved from the resin with TFA and purified by HPLC as described in the legend to Figure 1.

CD spectra were recorded on an AVIV 60DS or 62DS spectrometer equipped with a thermoelectric sample temperature controller. Samples for wavelength spectra contained 20 μM peptide in 10 mM phosphate buffer (pH 7.0). The cell path length was 0.1 cm. The spectra were collected as the average of six scans, using a three-second integration time at 1.0 nm wavelength increments. Spectra were baseline-corrected against the cuvet with buffer alone.

Samples for the thermal unfolding experiments were 20 μM peptide in 150 mM NaCl, 500 μM EDTA, 10 mM phosphate buffer (pH 7.0). A 1.0 cm path-length cell was used with continuous stirring. Data points were collected at one-degree intervals with an integration time of 20 seconds after a one-minute equilibration period at the desired temperature. The thermal unfolding curves were baseline-corrected by subtraction of a buffer only control experiment. In all cases the thermal unfolding was over 95% reversible and the T_m values did not change after repeated thermal

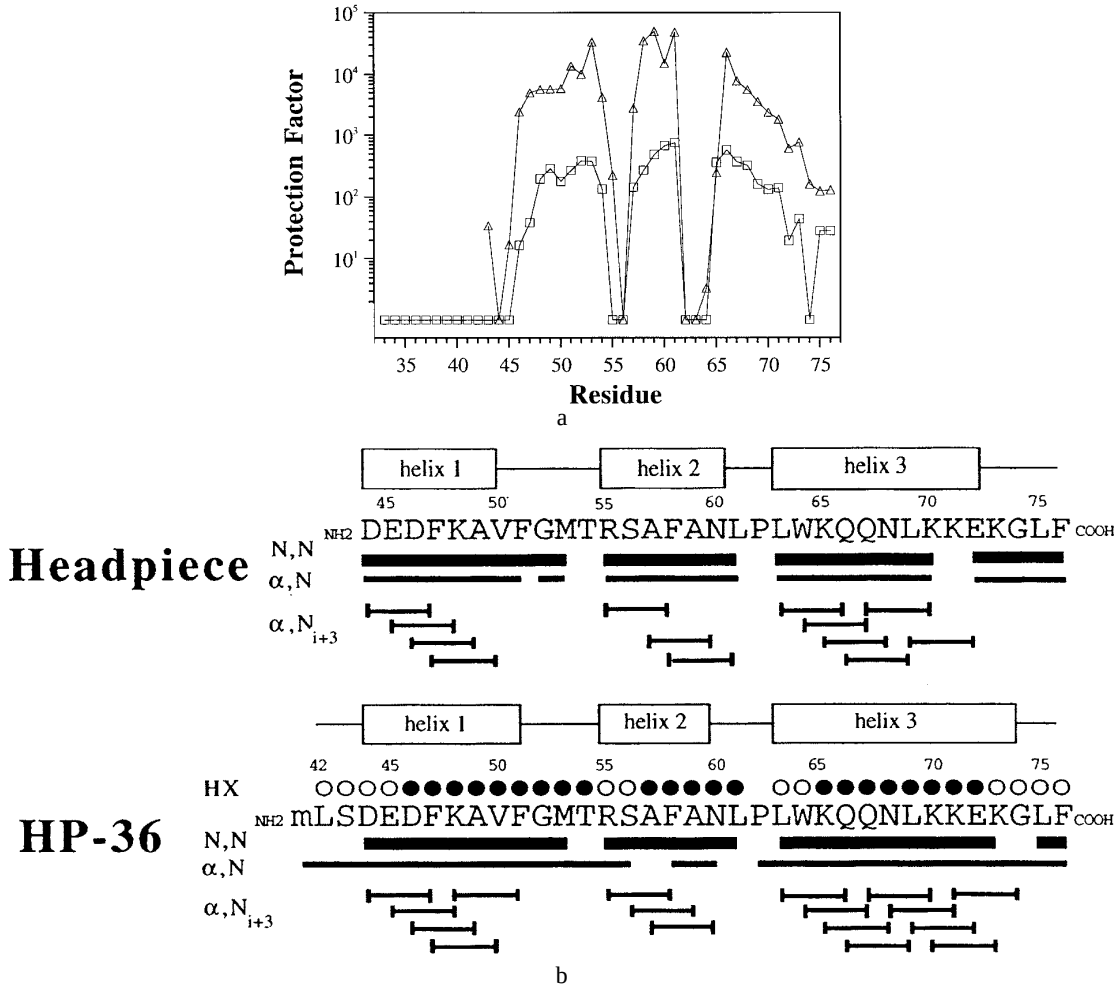


Figure 3 a–b *(caption overleaf)*

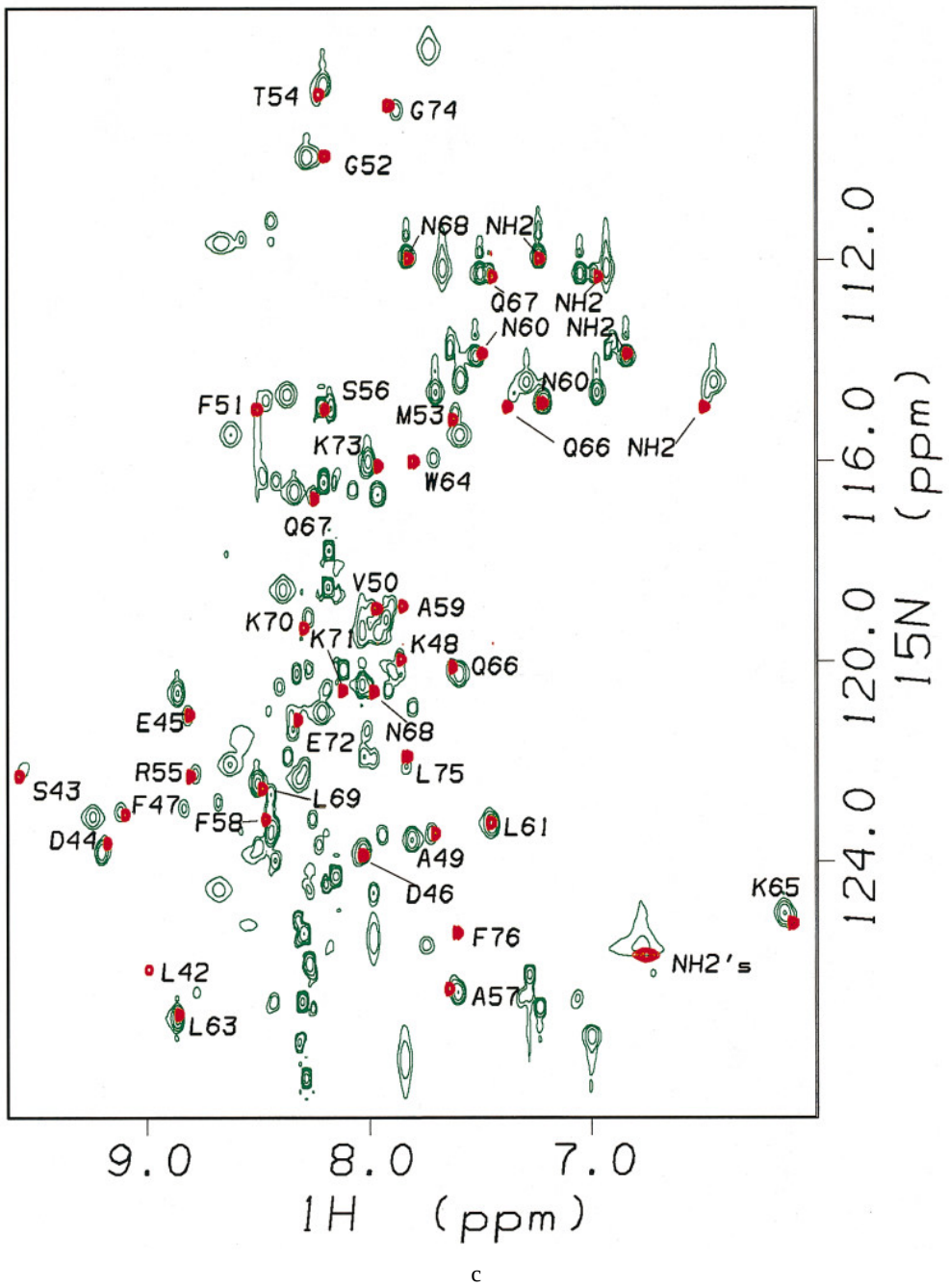
ΔG value between headpiece and the HP-45 fragment than that indicated by the small difference in T_m. Accurate determination of the ΔG value for intact headpiece by CD denaturant unfolding experiments is complicated by difficulties establishing a folded baseline.

From the magnitude of the CD signal at 222 nm (Figure 2b), we estimate that expressed HP-45 contains approximately the same number of residues in a helical conformation as intact headpiece ($\sim$18 residues compared with $\sim$19 for intact headpiece). Helix content was estimated from the CD signal by dividing the mean residue ellipticity at 222 nm by the value expected for 100% helix formation by short helices, –30,000 deg.cm^2 dmol^{-1} (Chen *et al*., 1974).

The N-terminal 14 residues of HP-45 have rapid amide–proton exchange rates (Figure 3a) and do not give rise to significant non-sequential NOE peaks (not shown), suggesting that this region is unfolded. A peptide (HP-35) was synthesized to test whether the N-terminal ten residues were required for the folding of HP-45. We began HP-35 at Leu42, since the backbone carbonyl oxygen of

unfolding (not shown). The T_m values were determined by examination of the derivative of the $[\theta]_{222}$ values with respect to the reciprocal of the temperature (Cantor & Schimmel, 1980). The error in estimation of T_m is $\pm$1°C.

Denaturation of HP-36 by GuHCl was performed in H_2O and ^2H_2O under conditions similar to those of the amide exchange measurements (pH 5.0, 4°C). Seven stock solutions spanning 0 to $\sim$8 M GuHCl, in H_2O or ^2H_2O, were prepared containing 20 μM HP-36, 150 mM NaCl and 25 mM acetate and then adjusted to pH 5.0. No corrections were made for the effect of ^2H_2O on the measured pH (Bundi & Wüthrich, 1979). The GuHCl concentration of each stock was determined by refractive index (Pace, 1986). Linear combinations of the stock solutions were used to prepare samples at intermediate GuHCl concentrations. The samples were equilibrated at 4°C overnight. CD data were collected in kinetics mode for five minutes, with a five second averaging time, at 222 nm, in a 0.5 cm path-length cell thermostated at 4°C. The average of the data points for the last 100 seconds of collection was taken for the final value. Selected samples in the transition region of the GuHCl denaturation experiment in H_2O were remeasured after an additional 24 hours of equilibration at 4°C and were superimposable with those at 12 hours, indicating that the samples were fully equilibrated (not shown). In addition to linear extrapolation, data were fit to a six-parameter equation to determine the slopes and intercepts of the folded and unfolded baselines, the m value, and $\Delta G_{pH\ 5,4C}$ (Santoro & Bolen, 1988).



c

Figure 3. NMR characterization of headpiece, HP-45, and HP-36. a, Hydrogen-exchange protection factors for HP-45 and HP-36. Amide groups that exchanged too fast to measure are shown with protection factors of 1 for reference only. Open squares, HP-45; open triangles, HP-36. Exchange was measured at pH 5.0, 30°C for HP-45, and 4°C for HP-36. HSQC spectra were recorded at various times after addition of ^2H_2O to previously lyophilized H_2O samples. The volumes of the peaks were fit to exponentials to determine the exchange rates, and then converted to protection factors after calculation of the intrinsic, random-coil exchange rates (Bai *et al*., 1993). b, NMR assignment summaries of villin headpiece and HP-36. Bars under the sequence show unambiguous sequential (α,$N_{(i+1)}$, and N,$N_{(i+1)}$) and helical (α,$N_{(i+3)}$) NOE connectivities. Filled circles indicate residues with greater than 1000-fold amide exchange protection. The assignments were made at pH 5.0, 30°C for headpiece and at pH 3.7, 30°C for HP-36. c, HSQC spectra of headpiece and HP-36 at pH 5.0 and 30°C. Indicated assignments are for HP-36. Headpiece spectrum is in green, HP-36 is in red.

this residue is expected to be the hydrogen bond acceptor for the amide proton of Asp46, the most N-terminal amide proton protected from exchange in HP-45.

The CD spectrum of HP-35 (Figure 2b) indicates that it folds, and that ~16 residues are in a helical conformation. Most importantly, the T_m for the unfolding of HP-35 (70°C) is only 4°C less than intact headpiece (Figure 2a). Therefore, the thermostable folded structure of headpiece is located within the C-terminal 35 residues.

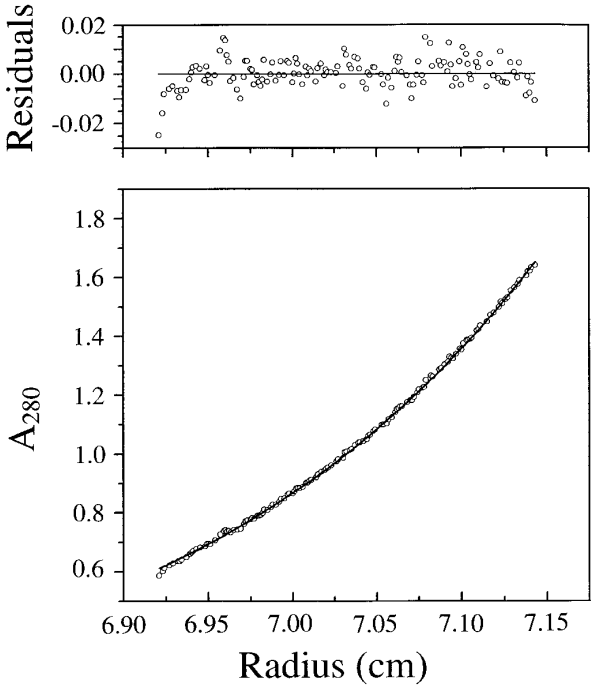


Figure 4. Sedimentation equilibrium of HP-35. This sample consisted of 100 μM HP-35, in 150 mM NaCl, 10 mM phosphate buffer (pH 7.0) 25°C. The spinning rate for this experiment was 38,000 rpm. The observed molecular mass was 3980; the expected mass is 4059. The random distribution of residuals (above) indicates the data are fit well by an ideal single-species model.
Methods: Analytical ultracentrifugation of HP-35 was performed on a Beckman X-LA ultracentrifuge with an AN-60 Ti rotor. Molecular masses were determined with initial peptide concentrations of 10, 30, and 100 μM in 150 mM NaCl, 10 mM phosphate buffer (pH 7.0). Data sets were collected at 25°C after equilibration (>15 hours) at rotor speeds of 38,000 and again at 48,000 rpm. The data were analyzed as described by Laue *et al*. (1992).

The HP subdomain is monomeric and has a fixed structure

Figure 4 shows a sample of the sedimentation equilibrium data collected on HP-35 to determine its oligomeric state. Over a tenfold range of peptide concentration (10, 30, and 100 μM), the average experimental molecular mass of HP-35 is 3890(±130) (s.d.), as determined by analytical ultracentrifugation (Laue *et al*., 1992). This value is in good agreement with the calculated molecular weight of 4059, indicating that HP-35 is monomeric in solution.

The 35-residue subdomain was expressed in *Escherichia coli* to facilitate further studies. The expressed peptide, HP-36, retains the additional N-terminal initiator methionine. Both the CD spectrum and thermal unfolding of HP-36 are similar to that of HP-35 (Figure 2a and b). The only significant difference is a slightly elevated thermal unfolding temperature (T_m = 72°C) for the longer, recombinant HP-36.

Linear extrapolation of the GuHCl denaturation of HP-36 to 0 M GuHCl (Figure 2c) yields a $\Delta G_{pH\ 5,\ 4°C}$ value of 3.3(±0.4) kcal mol^{-1} (1 kcal = 4.184 Joules). In ^2H_2O, $\Delta G_{pH\ 5,\ 4°C}$ is 4.1(±0.3) kcal mol^{-1}, with an m value of 1040 cal mol^{-1} M^{-1} (not shown). These values are likely to be the lower limits of ΔG because of the simple linear extrapolation fitting procedure used (Pace, 1986; Santoro & Bolen, 1988).

If the villin C-terminal subdomain has a unique, well-defined structure, it should contain slowly exchanging amide protons, with protection factors equal to, or greater than, those predicted from the global stability determined by chemical denaturation (Lumb & Kim, 1995). For HP-36, this would correspond to amide-exchange protection factors in excess of ~2000 (using $\Delta G = -RT \ln K$, and ΔG = 4.1 kcal mol^{-1}, in ^2H_2O). HP-35 contains five amides with protection factors between 20,000 and 50,000 in ^2H_2O (Figure 3a). Thus, HP-36 exhibits the amide exchange behavior expected of a protein in a fixed, unique conformation.

In the C-terminal region of headpiece, the pattern of $\alpha,N_{(i+3)}$ NOEs, characteristic of α-helices, indicates the presence of three short helical segments (Figure 3b). This same pattern is also seen in HP-36, indicating that it contains the same three helical segments. Further support for the similarity in the structure between HP-36 and the corresponding residues in headpiece is provided by the close correspondence of the respective backbone chemical shifts (Figure 3c and Table 1).

Conclusions

The 35-residue subdomain from villin headpiece is fully folded in a single, discrete conformation and is not a molten globule. Unlike molten globules (Kuwajima *et al*., 1976; Ptitsyn, 1992), HP-35 undergoes a cooperative thermal unfolding transition and has an NMR spectrum with substantial chemical shift dispersion. Proteins that fold into a unique conformation generally contain amide protons with hydrogen-exchange protection factors that are an order of magnitude greater than expected if exchange occurred only from globally unfolded molecules (Bai *et al*., 1994; Loh *et al*., 1993; Marmorino *et al*., 1993; Mayo & Baldwin, 1993; O'Shea *et al*., 1993; Roder, 1989). Conversely, molten globules and most designed proteins or peptides have amide-proton exchange rates that are substantially faster than that predicted from a global unfolding exchange mechanism, suggesting that ill-defined, fluctuating structures are formed (see Discussion in Lumb & Kim, 1995). As expected for a folded protein with a unique, fixed structure, HP-36 contains amides with hydrogen-exchange protection factors in excess of those predicted from the global stability of the subdomain determined by chemical denaturation.

Although some short monomeric helical peptides are marginally thermostable (Marqusee *et al*., 1989), peptides of length comparable to HP-35 generally

require oligomerization, ligand binding, co-solvents, or disulfide-bond formation for stable folding. In contrast, HP-35 folds independently and autonomously in the absence of these additional stabilizing mechanisms. Recently, the NMR structure of a monomeric, 23-residue synthetic polypeptide corresponding to a zinc finger has been reported (Struthers *et al*., 1996). However, in order to fold in the absence of zinc, the peptide requires one unusual amino acid (D-proline) and one non-naturally occurring amino acid (3-(1,10-phenanthrol-2-yl)-L-alanine).

After HP-35, the shortest sequence of naturally occurring amino acids that folds autonomously into a stable structure, of which we are aware, is the 43-residue peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from pyruvate dehydrogenase (Kalia *et al*., 1993). Although the NMR structure of that subdomain contains a 33-amino acid core, it has not been determined whether the 33-amino acid core can fold to form a stable structure, in isolation. Indeed, residues at both the N and C-terminal regions, outside the 33-amino acid core, contain amides that are protected from hydrogen exchange (Kalia *et al*., 1993), suggesting that, while less ordered in the NMR structure, these additional residues contribute to the folding and/or stabilization of the subdomain.

Like HP-35, the acetyltransferase subdomain folds to form three helices. However, the lack of sequence similarity, coupled with the longer sequence length and the wider gap between the central and C-terminal helices in acetyltransferase, suggests that the folds of these two subdomains may differ.

Without a detailed three-dimensional structure of the HP-35 subdomain, it is difficult to account for the high thermostability of this short, three-helix motif. The native-like amide exchange data and chemical-shift dispersion strongly suggest, however, that these helices fold to form a tightly packed hydrophobic core.

Finally, our NMR results indicate that the corresponding residues in intact headpiece fold into the same overall three-helix conformation found in the HP-35 subdomain. It is of interest that all of the residues identified as essential for actin binding (Doering, 1992; Friederich *et al*., 1992) are in the C-terminal helix of the HP-35 subdomain. HP-35 should be a useful subdomain for experimental and theoretical studies of protein folding, and for the design of small actin-binding proteins.

Acknowledgements

We thank Lawrence McIntosh for assistance with the NMR studies, Mike Burgess for peptide synthesis, and Kevin Lumb and Rheba Rutkowski for their careful reading of the manuscript. We acknowledge financial support from the American Cancer Society (C.J.M.), the National Institutes of Health (P.T.M.), and the Howard Hughes Medical Institute (P.S.K.).

References

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75–86.

Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* **29**, 4–14.

Bazari, W. L., Matsudaira, P. T., Wallek, M., Smeal, T., Jakes, R. & Ahmed, Y. (1988). Villin sequence and peptide map identify six homologous domains. *Proc. Natl Acad. Sci. USA*, **85**, 4986–4990.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.

Bretscher, A. & Weber, K. (1979). Villin: the major microfilament-associated protein of the intestinal microvillus. *Proc. Natl Acad. Sci. USA*, **76**, 2321–2325.

Bundi, A. & Wüthrich, K. (1979). ^{1}H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers*, **18**, 285–297.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry*, W. H. Freeman and Co., New York.

Chen, Y.-H., Yang, J. T. & Chau, K. H. (1974). Determination of the helix and B form of proteins in aqueous solution by circular dichroism. *Biochemistry*, **13**, 3350–3359.

Doering, D. S. (1992). *Functional and Structural Studies of a Small F-actin Binding Domain*. PhD thesis, Massachusetts Institute of Technology.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–1954.

Finidori, J., Friederich, E., Kwiatkowski, D. J. & Louvard, D. (1992). *In vivo* analysis of functional domains from villin and gelsolin. *J. Cell Biol.* **116**, 1145–1155.

Friederich, E., Vancompernolle, K., Huet, C., Goethals, M., Finidori, J., Vandekerckhove, J. & Louvard, D. (1992). An actin-binding site containing a conserved motif of charged amino acid residues is essential for the morphogenic effect of villin. *Cell*, **70**, 81–92.

Glenny, J. R. J., Geisler, N., Kaulfus, P. & Weber, K. (1981). Demonstration of at least two different actin-binding sites in villin, a calcium-regulated modulator of f-actin organization. *J. Biol. Chem.* **256**, 8156–8161.

Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870–7872.

Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. (1989). A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N–^{1}H multiple quantum coherence spectroscopy. *FEBS Letters*, **243**, 93–98.

Kalia, Y. N., Brocklehurst, S. M., Hipps, D. S., Appella, E., Sakaguchi, K. & Perham, R. N. (1993). The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *J. Mol. Biol.* **230**, 323–341.

Kuwajima, K., Nitta, K., Yoneyama, M. & Sugai, S. (1976). Three-state denaturation of α-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* **106**, 359–373.

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 90–125, The Royal Society of Chemistry, Cambridge, UK.

Loh, S. N., Prehoda, K. E., Jinfeng, W. & Markely, J. L. (1993). Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry*, **32**, 11022–11028.

Lumb, K. J. & Kim, P. S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry*, **34**, 8642–8648.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989). Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.

Marmorino, J. L., Auld, D. S., Betz, S. F., Doyle, D. F., Young, G. B. & Pielak, G. J. (1993). Amide proton exchange rates of oxidized and reduced *Saccharomyces cerevisiae* iso-1-cytochrome c. *Protein Sci.* **2**, 1966–1974.

Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Natl Acad. Sci. USA*, **86**, 5286–5290.

Mayo, S. L. & Baldwin, R. L. (1993). Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science*, **262**, 873–876.

Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E, & Dahlquist, F. W. (1989). Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **177**, 44–73.

Oas, T. G. & Kim, P. S. (1988). A peptide model of a protein folding intermediate. *Nature*, **336**, 42–48.

O'Shea, E. K., Lumb, K. J. & Kim, P. S. (1993). Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658–667.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.

Pope, B., Way, M., Matsudaira, P. T. & Weeds, A. (1994). Characterization of the F-actin binding domains of villin: classification of F-actin binding proteins into two groups according to their binding sites on actin. *FEBS Letters*, **338**, 58–62.

Ptitsyn, O. B. (1992). The molten globule state. In *Protein Folding* (Creighton, T. E., ed.), pp. 243–300, W. H. Freeman and Co., New York.

Rana, A. P., Ruff, P., Maalouf, G. J., Speicher, D. W. & Chishti, A. H. (1993). Cloning of human erythroid dematin reveals another member of the villin family. *Proc. Natl Acad. Sci. USA*, **90**, 6651–6655.

Riviere, L. R., Fleming, M., Elicone, C. & Tempst, P. (1991). Study and applications of the effects of detergents and chaotropes on enzymatic proteolysis. In *Techniques in Protein Chemistry II* (Villafranca, J. J., ed.), pp. 171–180, Academic Press Inc., San Diego.

Roder, H. (1989). Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol.* **176**, 446–473.

Rose, G. D. (1979). Hierarchic organization of domains in globular proteins. *J. Mol. Biol.* **134**, 447–470.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Course*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368–379.

Struthers, M. D., Cheng, R. P. & Imperiali, B. (1996). Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science*, **271**, 342–345.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89.

Vanderkerckhove, J. (1990). Actin-binding proteins. *Curr. Opin. Cell Biol.* **2**, 41–50.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acid*, John Wiley and Sons, New York.

Zuiderweg, E. P. R. (1990). A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity. *J. Magn. Reson.* **86**, 346–357.

Edited by P. E. Wright

(Received 2 January 1996; received in revised form 4 April 1996; accepted 18 April 1996)