Protein Science (1994), *3*:70–81. Cambridge University Press. Printed in the USA.

Solution structure of villin 14T, a domain conserved among actin-severing proteins

MICHELLE A. MARKUS,[1] TOMOKO NAKAYAMA,[2]
PAUL MATSUDAIRA,[2] AND GERHARD WAGNER[1]

[1] Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, Massachusetts 02138, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115
[2] Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

(RECEIVED September 9, 1993; ACCEPTED October 19, 1993)

Abstract

The solution structure of the N-terminal domain of the actin-severing protein villin has been determined by multidimensional heteronuclear resonance spectroscopy. Villin is a member of a family of actin-severing proteins that regulate the organization of actin in the eukaryotic cytoskeleton. Members of this family are built from 3 or 6 homologous repeats of a structural domain of approximately 130 amino acids that is unrelated to any previously known structure. The N-terminal domain of villin (14T) contains a central β-sheet with 4 antiparallel strands and a fifth parallel strand at one edge. This sheet is sandwiched between 2 helices on one side and a 2-stranded parallel β-sheet with another helix on the other side. The strongly conserved sequence characteristic of the protein family corresponds to internal hydrophobic residues. Calcium titration experiments suggest that there are 2 binding sites for Ca^{2+}, a stronger site near the N-terminal end of the longest helix, with a K_d of 1.8 ± 0.4 mM, and a weaker site near the C-terminal end of the same helix, with a K_d of 11 ± 2 mM. Mutational and biochemical studies of this domain in several members of the family suggest that the actin monomer binding site is near the parallel strand at the edge of the central β-sheet.

Keywords: actin-binding protein; calcium-binding protein; modular domain; NMR; protein structure; villin

Actin-severing proteins regulate actin filament length in the cytoplasm of eukaryotic cells by severing actin filaments, capping filament growth, and nucleating polymerization (Matsudaira & Janmey, 1988). In turn, these proteins are regulated by binding Ca^{2+} and polyphosphoinositides. This family of proteins is based on a fundamental 130-residue domain that contains conserved hydrophobic residues. This domain is repeated 3 times in severin and fragmin and 6 times in gelsolin and villin (Janmey & Matsudaira, 1988; Matsudaira & Janmey, 1988; Way & Weeds, 1988). The organization of proteins in this family is depicted in Figure 1. Calcium induces conformational changes in villin that are detected by differences in the pattern of proteolysis between domains. Despite their homology, the domains differ in their actin-binding activity. The N-terminal domain binds actin monomers, whereas the second domain binds actin filaments. Mutants of gelsolin that lack the N-terminal domain decorate actin filaments but lose severing activity. Because single domains do not cap and sever actin filaments, these activities probably result from cooperativity between monomer and filamentous actin binding sites on adjacent domains (Pope et al., 1991).

Villin differs from other members of the severing family in several important respects. First, villin contains a seventh domain, the headpiece, at its C-terminus. The headpiece contains a second binding site for filamentous actin that confers on villin a unique actin-cross-linking activity. Second, the capping and severing activities of villin require a higher free Ca^{2+} concentration for activity (Janmey & Matsudaira, 1988; Way & Weeds, 1988). Like gelsolin, these activities are dependent on the N-terminal domain (14T). The reduced calcium affinity and the additional headpiece domain enable villin to maintain actin bundles in intestinal microvilli not only in normal calcium concentrations (submicromolar) but also in elevated calcium concentrations (less than 0.1 millimolar) that would activate the actin-severing activity of gelsolin. The actin monomer-binding activity of the isolated N-terminal domain and its critical role in actin severing and capping raise several questions: What is the conformation of the fundamental domain? Are the conserved sequences involved in actin binding or in the structural framework of the domain? How does the conformation change when bound to calcium and actin? How does the conformation of the isolated domain compare to the domain when a part of villin? We have

Reprint requests to: Gerhard Wagner, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115.

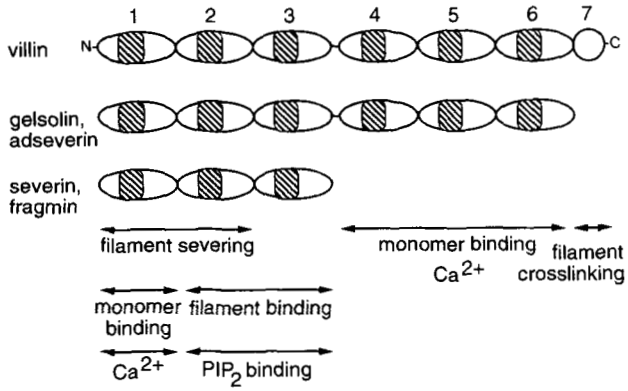

Fig. 1. Schematic drawing of the structural organization and distribution of functions in the actin-severing proteins. Ovals represent the conserved domain; the circle represents the bundling domain in villin. Hatched boxes represent regions of strong sequence homology. Domains are numbered starting from the N-terminal end of the sequence; "N" denotes the amino-terminus and "C" denotes the carboxyl-terminus of the protein. Regions associated with particular functions are denoted with arrows; Ca^{2+} refers to calcium-dependent activity, and PIP_2 is an abbreviation for phosphatidyl inositol 4,5-bisphosphate. Examples of proteins with each of the organizations shown are named at the left.

studied the 3-dimensional solution structure of 14T, a prototypic domain of the actin-severing family, as a first step in describing the mechanisms of actin filament capping, severing, and bundling by villin. We have also made a preliminary comparison of the 14T structure to the recent structure of the corresponding domain of gelsolin, derived from a co-crystal with actin (McLaughlin et al., 1993). An atomic level comparison of these structures is the next step toward understanding the mechanism of actin-monomer binding and the differences between these domains.

Results

Assignments

Sequential assignments for villin 14T were obtained by a triple-resonance, backbone-directed strategy and extended by traditional nuclear Overhauser effect-based strategies (Wüthrich, 1986). Backbone assignments for amide nitrogen, α-carbon, amide proton, and α-proton chemical shifts are complete except for the nitrogen chemical shifts for the 3 proline residues, the amide nitrogen and proton chemical shifts for the N-terminal valine, and the assignments for Asp^{12} and Lys^{13}. Amide nitrogen to proton crosspeaks for Asp^{12} and Lys^{13} were not observed in the ^{15}N heteronuclear correlation spectra and were not identified in subsequent spectra. Crosspeaks for neighboring Leu^{11} and Thr^{14} were noticeably broader than most, suggesting that local exchange broadening may have obscured the missing crosspeaks.

Side chain assignments were obtained from various types of correlation spectroscopy. So far, side chain assignments including chemical shifts for nitrogen, proton-attached carbon, and protons, but excluding the carbons of the aromatic rings, the terminal amines of the arginines and lysines, and the imidazole amines of the histidine rings, are complete for 94 of 126 residues. The experiments used for the assignments are listed with references in the Materials and methods section. Full details of the assignments will be presented elsewhere (Markus MA, Nakayama T, Matsudaira P, Wagner G, manuscript in prep.).

Structure calculations and residual violations of restraints

Information for structure calculations based on the NMR data is summarized in Table 1. NOE crosspeaks were identified in homonuclear 2-dimensional and heteronuclear 3-dimensional data sets. Dihedral angle restraints for the backbone angle ϕ were based on measurements of the vicinal coupling constant $^3J_{HNH\alpha}$. Restraints on the side chain torsion angle χ_1 and stereospecific assignments for β-methylene protons were derived from measurements of the vicinal coupling constants $^3J_{H\alpha H\beta}$ and $^3J_{NH\beta}$. Stereospecific assignments for the methyl groups of valine and leucine were obtained from a ^{13}C heteronuclear single quantum coherence (^{13}C-HSQC) spectrum recorded on a 10% labeled sample (Neri et al., 1989; Senn et al., 1989; Szyperski et al., 1992). The restraints described above were supplemented with 86 general distance restraints representing 43 backbone hydrogen bonds assigned on the basis of hydrogen exchange experiments. A complete list of the restraints is included in the supplementary material on the Diskette Appendix (file Markus.NMR, SUPLEMNT directory). Based on these restraints, structures were calculated with the distance geometry program DGII (Havel, 1991). Backbone atom representations of 10 distance geometry structures are superimposed in Figure 2A.

For the 10 best structures from a calculation of 20, residual NOE restraint violations were small, with no violations above 0.9 Å in any structure. The number of violations above 0.5 Å, quoted as mean ± SD, is 2.6 ± 1.2 violations per structure and the number above 0.2 Å is 23.4 ± 4.7. Violations of dihedral angle restraints were also small, with only 1 structure with a single angle violation above 10° and an average of 2.9 ± 1.5 violations above 1° per structure.

Description of the structure and evaluations of its precision

The structure of 14T is shown as a ribbon diagram in Figure 2B and as numbered backbone traces in Figure 2C and D and in Ki-

Table 1. *Summary of data used in structure calculations*

Restraint type	Number of restraints
Assigned NOE crosspeaks	772
NOE distance restraints (total)	720
Interproton distances	
Intraresidue	17
Interresidue sequential ($\|i-j\| = 1$)	300
Interresidue medium range ($2 \leq \|i-j\| \leq 4$)	97
Interresidue long range ($\|i-j\| > 4$)	306
Hydrogen bond distances	86
Dihedral angles	
ϕ ($C'_{(i-1)} - N_i - C_{\alpha i} - C'_i$)	67
χ_1 ($N_i - C_{\alpha i} - C_{\beta i} - C_{\gamma i}$)	46
Stereospecific assignments	
β Methylene	35
γ Terminal methyl groups for valine	8 of 9
δ Terminal methyl groups for leucine	5 of 7

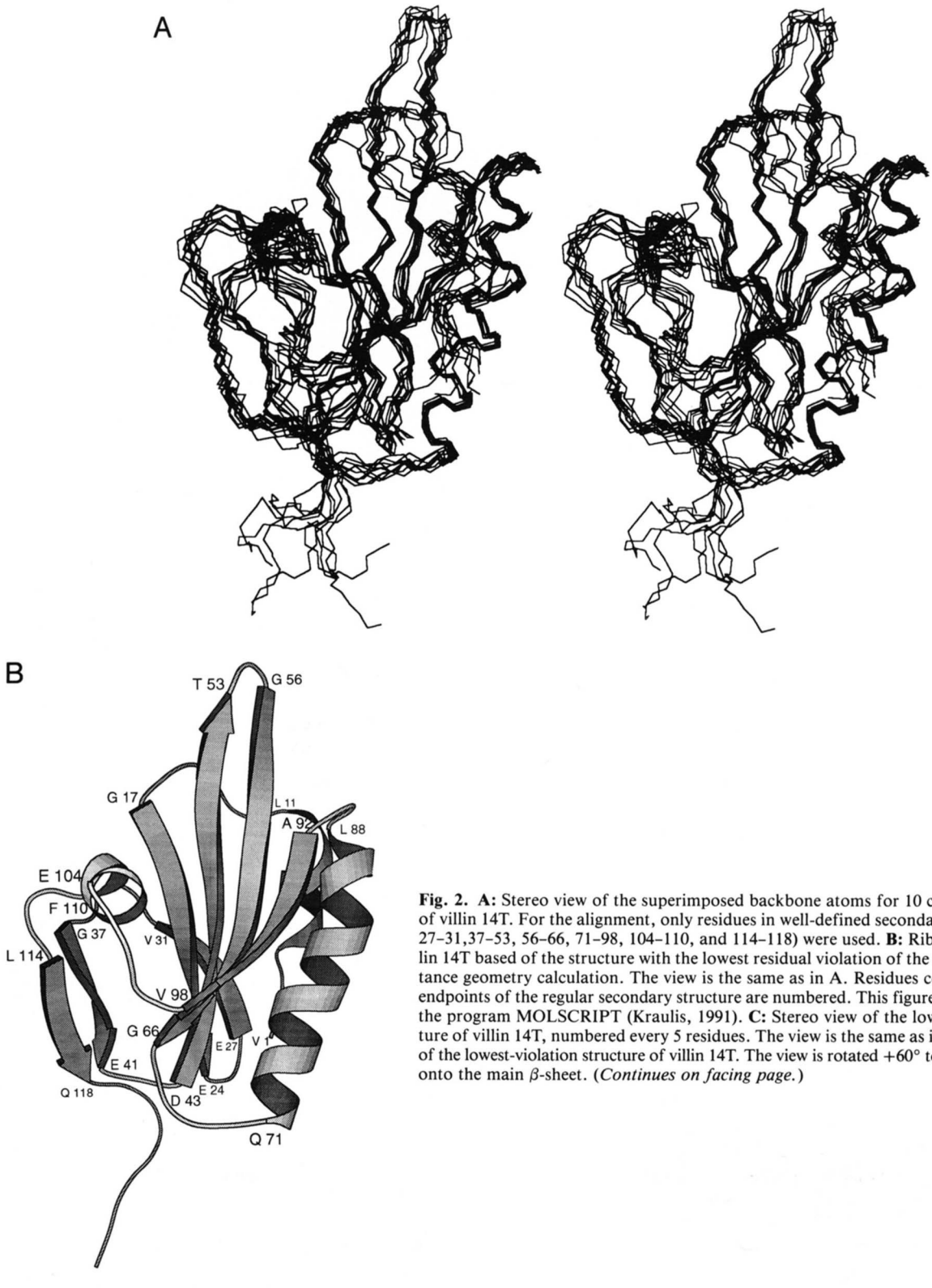

Fig. 2. A: Stereo view of the superimposed backbone atoms for 10 calculated structures of villin 14T. For the alignment, only residues in well-defined secondary structure (16–24, 27–31,37–53, 56–66, 71–98, 104–110, and 114–118) were used. **B:** Ribbon diagram of villin 14T based of the structure with the lowest residual violation of the restraints in the distance geometry calculation. The view is the same as in A. Residues corresponding to the endpoints of the regular secondary structure are numbered. This figure was generated with the program MOLSCRIPT (Kraulis, 1991). **C:** Stereo view of the lowest-violation structure of villin 14T, numbered every 5 residues. The view is the same as in A. **D:** Stereo view of the lowest-violation structure of villin 14T. The view is rotated +60° to look straight down onto the main β-sheet. (*Continues on facing page.*)

nemage 1. Villin 14T is organized around a central β-sheet comprised of 4 antiparallel strands (residues 17–24, 27–31, 43–53, and 56–66) and 1 parallel strand (92–98) that is surrounded on one side by a long α-helix (71–88) together with a shorter helix (1–11) and on the other side by a short parallel β-sheet (37–41 and 114–118) and another short helix (104–110). The N-terminal 11 amino acids are described as a helix because there are a few medium-range NOE crosspeaks characteristic of a helix (i to $i+3$ and i to $i+4$) and the $^3J_{HNH\alpha}$ values are small. However, there are not enough NOE and dihedral restraints in this region

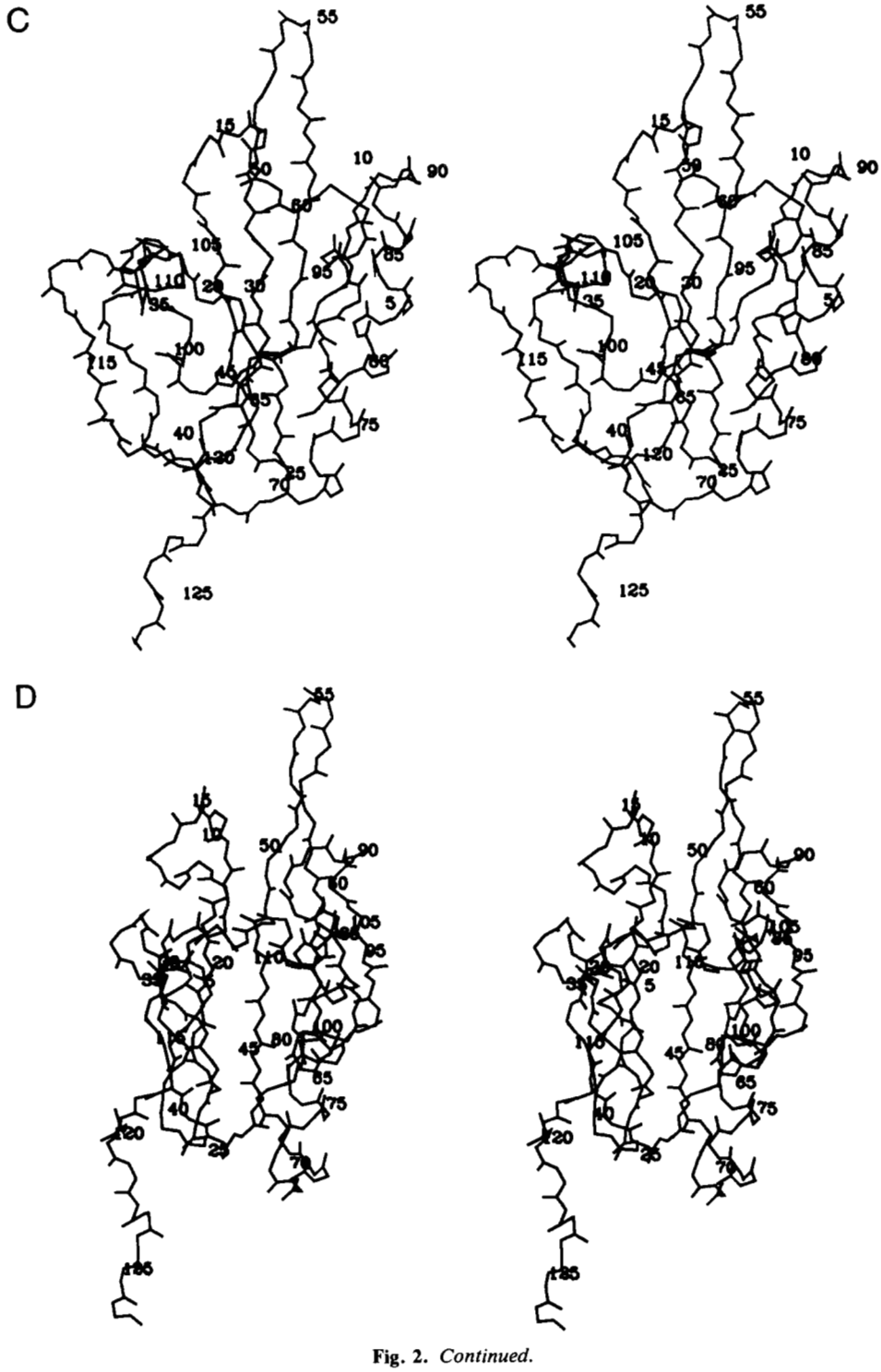

Fig. 2. *Continued.*

to produce a well-defined helix in the distance geometry calculations, and the hydrogen exchange data do not support explicit hydrogen bond restraints. The highly disordered residues at the C-terminal end of this helix correspond to the residues that lack backbone assignments, Asp^{12} and Lys^{13}.

The residues involved in regular secondary structure are structurally well defined; superposition of the central β-sheet, the longest helix, the parallel β-sheet, and the C-terminal helix gives an average RMS deviation (⟨RMSD⟩) of 0.70 ± 0.10 Å from the average structure for backbone atoms. Superposition of backbone atoms in residues 1–121 has an ⟨RMSD⟩ from the mean structure of 1.15 ± 0.17 Å; residues 122–126 are excluded because no medium- or long-range NOE crosspeaks have been assigned for them to date. For all heavy atoms in residues 1–121, the ⟨RMSD⟩ is 1.63 ± 0.16 Å. To examine which parts of the structure are well defined in more detail, a plot of the ⟨RMSD⟩ per residue as a function of sequence is given in Figure 3A. The lowest ⟨RMSD⟩s correspond to the regular secondary structure,

with values for the backbone generally less than 1.0 Å and some less than 0.5 Å. The N-terminal helix and the turns between elements of secondary structure are somewhat less defined. The highest ⟨RMSD⟩s, with values above 2.0 Å, are found near the unassigned residues at 12 and 13 and after residue 120.

The pattern of well-defined structure in regions of regular secondary structure is clear in the analysis of backbone dihedral angles and order parameters (Hyberts et al., 1992), shown in Figure 3B. The plot of ϕ and ψ as a function of sequence confirms that the regions identified as strands of β-sheet show the angles characteristic of β-sheets, and the α-helices, except for the N-terminal helix, show the angles characteristic of α-helices (Hyberts et al., 1992). The order parameters for both ϕ and ψ in these regions are generally above 0.9, indicating 1 consistent conformation in all 10 structures. The N-terminal helix is somewhat disordered due to lack of NOE restraints, as already discussed. The turns between the elements of secondary structure generally have lower-order parameters.

The Ramachandran plots for non-glycine residues in all 10 structures are superimposed in Figure 4A. While most pairs of dihedral angles fall within the allowed ranges, a significant number have positive ϕ angles. However, if we use the angular order parameters to identify well-defined angles and plot only these angles, there are no pairs with positive ϕ angles (Fig. 4B). This suggests that positive ϕ angles point to regions where more restraints are needed rather than to unfavorable steric contacts and that the Ramachandran plot will improve as the structure is refined.

Calcium binding by villin 14T

Because villin requires calcium for actin-binding activity (Janmey & Matsudaira, 1988), we titrated calcium into an ^{15}N-labeled sample of 14T to identify possible Ca^{2+}-binding sites. The structural studies were done at pH 4.15; however, Ca^{2+} binding requires a more neutral pH. Therefore, the first step was a pH titration to transfer amide proton and nitrogen assignments to pH 6.91. Almost half (46%) of the peaks in the ^{15}N-HSQC experiment do not move as the pH is raised. Most peaks move less than 0.20 ppm in ^{1}H chemical shift and less than 1.5 ppm in ^{15}N chemical shift. The stationary peaks are distributed throughout the sequence, suggesting that the protein does not undergo any major conformational change as the pH changes. The peaks that move the most are Cys^{44}, Ala^{92}, and Val^{121}, with changes in ^{1}H chemical shift ranging from 1.00 to 1.05 ppm and changes in ^{15}N chemical shift ranging from 1.4 to 3.2 ppm. As Ca^{2+} is added, half the peaks do not move, again suggesting that the overall conformation of the protein remains the same. However, several backbone amide crosspeaks move by more than 0.50 ppm in ^{1}H chemical shift, including Asp^{43} (0.53 ppm), Ser^{70} (0.59 ppm), Ala^{92} (0.66 ppm), Val^{121} (1.09 ppm), and Ala^{122} (0.53 ppm). Several also move by more than 2.0 ppm in ^{15}N chemical shift, including Glu^{41} (2.3 ppm), Asp^{43} (5.7 ppm), Val^{91} (2.9 ppm), Ala^{92} (2.2 ppm), Val^{121} (5.6 ppm), Ala^{122} (4.7 ppm), and Ser^{123} (3.5 ppm). Based on the change in chemical shift, the fraction of the protein bound to Ca^{2+} can be calculated. Plots of the fraction of protein bound to Ca^{2+} as a function of the total Ca^{2+} concentration are shown for 2 representative crosspeaks in Figure 5. Most of the peaks that move follow the curve for Val^{121}. However, Ala^{92} and Val^{91} follow the curve shown in Figure 5 for Ala^{92}. This shows that there are at least 2 binding sites for calcium, distinguished by their dissociation constant. Calculation of the dissociation constants, assuming 2 independent Ca^{2+} binding sites, gives a K_d of 1.8 ± 0.4 mM for the stronger site, near Val^{121}, and 11 ± 2 mM for the weaker site, near Ala^{92}.

Discussion

Topology of villin 14T

The general features of the structure, a central β-sheet surrounded by helices, resemble several other proteins, including the actin-binding protein profilin (Archer et al., 1993). However, the precise topology of 14T is unique. This topology is diagrammed in Figure 6. The protein can be described in terms of 3 layers, represented by the planes in Figure 6. The N-terminal helix, α1, from the bottom layer, leads into strand β1 (17–24) in the middle layer. Strand β1 is connected by a tight turn to strand β2 (27–31), which leads into strand β3 (37–41) in the top layer. Strand β3 leads back into the middle layer and to strand β4 (43–53), which connects to strand β5 (56–66) in a tight turn. From strand β5, the chain drops back to the bottom layer to helix α2 (71–88). Helix α2 leads back to the middle layer to strand β6 (92–98). Strand β6 leads back to the top layer for helix α3 (104–110) and strand β7 (114–118). The topology of the central β-sheet is β2, β1, β4, β5, and β6. The small parallel β-sheet in the top layer is β3 and β7. Since parallel strands are usually buried within protein structures (Richardson, 1981), the parallel strand β6 at the edge of the main β-sheet and the small parallel sheet are possible sites for docking the remaining domains of villin or for binding actin.

Hydrophobic core

All 3 helices are amphipathic in character, consistent with their positions on the surface of 14T. Helix α1 extends hydrophilic side chains Glu^2, Lys^5, and Asp^{12} toward solvent and packs hydrophobic side chains Leu^3, Val^7, and Leu^{11} in toward the rest of the protein. Helix α2 has hydrophilic functional groups from Gln^{74}, Gln^{83}, Asp^{85}, and Glu^{86} on its solvent side and hydrophobic groups from Ala^{77}, Thr^{81}, Met^{84}, and Leu^{88} facing inward. Asn^{68} is in the position of a classical N-cap residue for helix α2. Helix α3 exposes charged side chains Glu^{104} and Arg^{107} to solvent and buries aromatic rings Phe^{106}, Tyr^{109}, and Phe^{110} in toward the central β-sheet. Side chains on the central β-sheet form hydrophobic patches that are covered by the hydrophobic faces of the helices. Thus, residues on one side of the central β-sheet form a hydrophobic region that interacts with the hydrophobic faces of helix α1 and helix α2, as sketched in Figure 7A with side chains extending from the sheet shown as ovals and the helices shadowed in over the sheet. As seen in detail in Kinemage 2, side chains from Val^7 and Leu^{11} from helix α1 (shown in green) pack to Trp^{21} (strand β1) and to Leu^{48} (strand β4) (both in magenta). Ala^{77}, Thr^{81}, Met^{84}, Asp^{85}, and Leu^{88} from helix α2 (in blue) pack to Ile^{23} of strand β1, Cys^{44}, Val^{46}, and Leu^{48} of strand β4, Tyr^{59}, Ile^{61}, and Tyr^{63} of strand β5, and Gln^{94} of strand β6. These 2 helices also interact with each other, with Leu^3, Val^7, and Leu^{11} from helix α1 interacting with Tyr^{80}, Met^{84}, Tyr^{87}, and Leu^{88} from helix α2. On the other side of the central β-sheet, the hydrophobic region is covered by Phe^{106}, Tyr^{109}, and Phe^{110} from helix α3 and Phe^{39} and Leu^{114} from strands β3 and β7, respectively, of the short

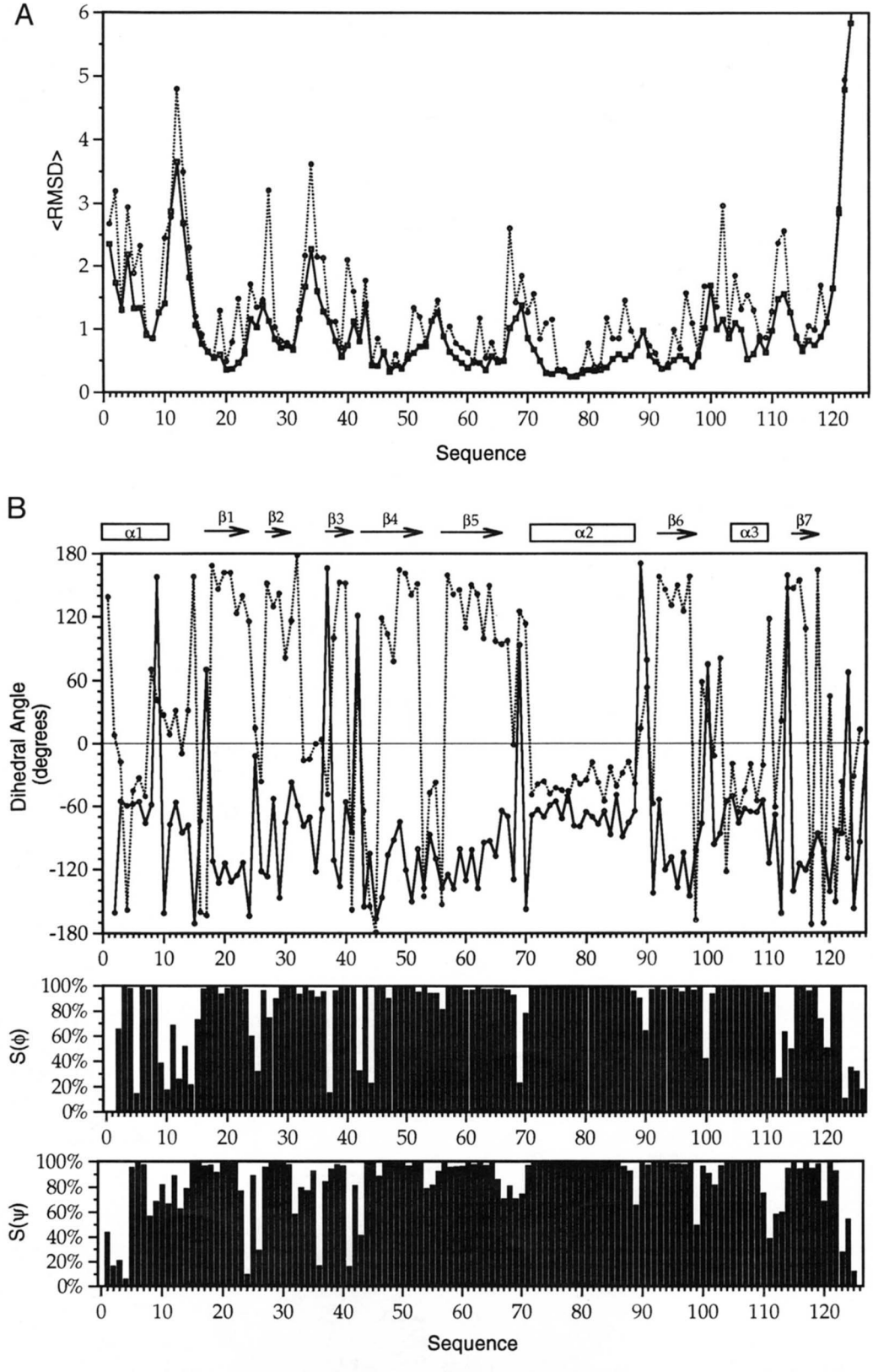

Fig. 3. A: Plot of the average RMS deviation (⟨RMSD⟩) per residue for the 10 best structures. The solid line gives the ⟨RMSD⟩ as a function of sequence for the backbone atoms N, C_α, C′, and O. The dashed line gives the ⟨RMSD⟩ for all heavy atoms in the residue. **B:** Plot of the average dihedral angles ϕ and ψ versus sequence. In the top panel, the solid line represents the ϕ angles and the dashed line represents ψ. At the top of the panel, the secondary structure is represented with arrows for strands and open boxes for α-helices. In the lower panels, the angular order parameters are plotted. Values near 1.00 mean that the angles are well defined. For a completely disordered angle, the order parameter should approach $1/\sqrt{N}$ (Hyberts et al., 1992).

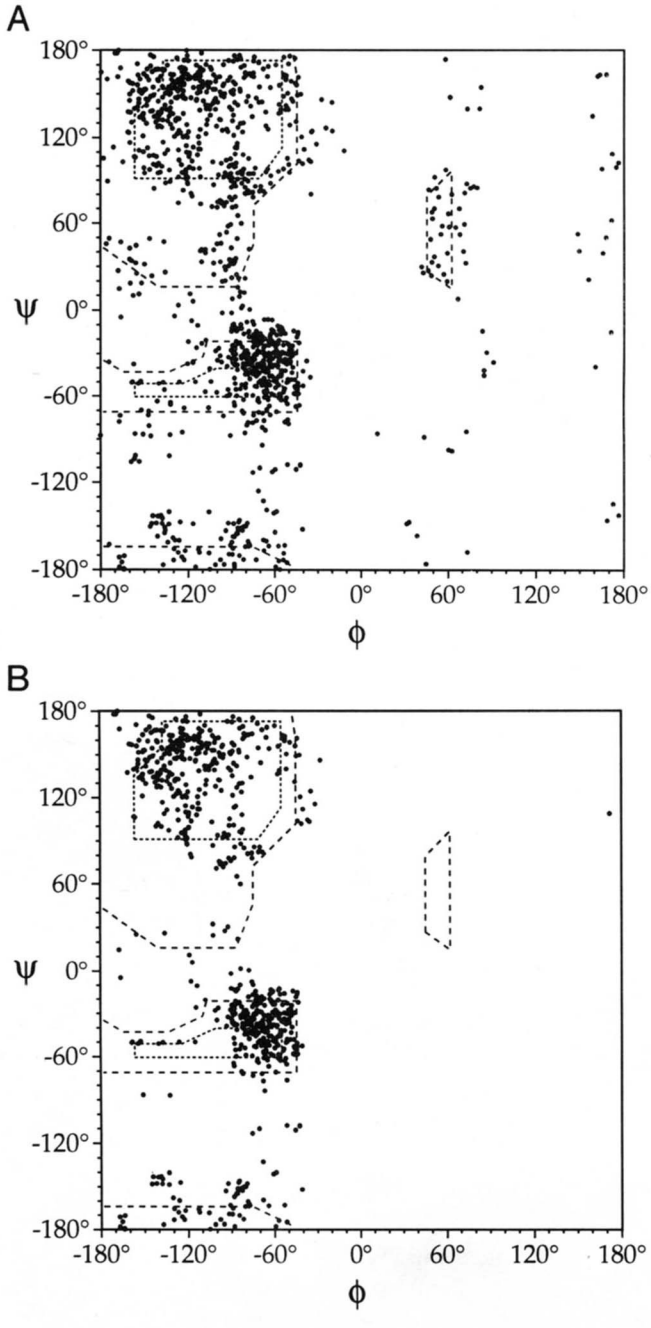

Fig. 4. Ramachandran plots for (**A**) pairs of ϕ and ψ angles for all non-glycine residues in all 10 structures and (**B**) only pairs of angles with order parameters above 0.9.

parallel β-sheet. These interactions are sketched in Figure 7B and shown in detail in Kinemage 2.

Conserved turns

Villin 14T has a relatively large number of glycines (14 of 126 or 11%), most of which are located near reverses in the polypeptide chain. Thus, Gly^{17} forms part of the transition from helix $\alpha 1$ to strand $\beta 1$ of the central β-sheet, Gly^{42} from strand $\beta 3$ of the short parallel β-sheet to strand $\beta 4$ of the central β-sheet, Gly^{54} and Gly^{56} between strands $\beta 4$ and $\beta 5$ of the central β-sheet, Gly^{66} from strand $\beta 5$ of the central β-sheet into helix $\alpha 2$, Gly^{89}

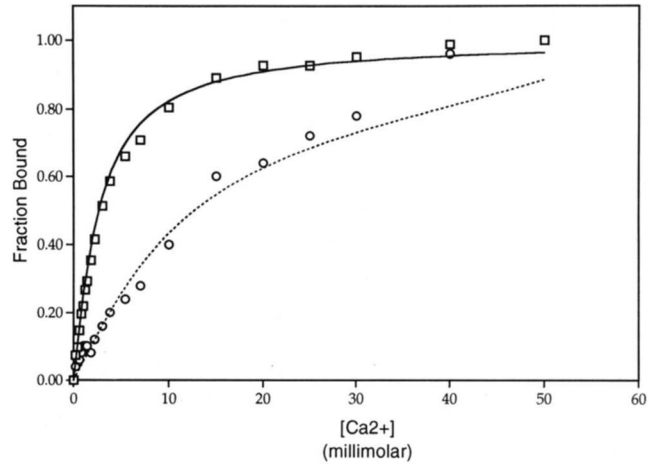

Fig. 5. Fraction of protein in the Ca^{2+}-bound form as a function of total calcium concentration. The circles represent data points for the proton chemical shift of Ala^{92}, and the boxes represent data for Val^{121}. The solid line and the dashed line represent the curves fit to these data with dissociation constants of 1.8 mM and 11 mM.

out of helix $\alpha 2$ into strand $\beta 6$ of the central β-sheet, Gly^{100} at a tight turn at the end strand $\beta 6$, and Gly^{113} from helix $\alpha 3$ into strand $\beta 7$ of the parallel β-sheet. Of these, Gly^{66} is highly conserved among domains of the actin-severing proteins. Further, glycines at position 42, near 54 or 56, near 89, and at 100 are commonly found in domains one and four. Because turns establish boundaries for elements of secondary structure, their conservation points to elements of secondary structure that may be conserved. The proline residues, which all adopt a *trans* conformation, are also implicated in turns. Pro^{16} is part of the link between helix $\alpha 1$ and the central β-sheet, and Pro^{32} is at the beginning of the loop leading into the parallel β-sheet. Pro^{30} is still part of the central β-sheet, though it is near the end of strand $\beta 2$. Pro^{30} is strongly conserved among domains one and four and is also seen, though less consistently, in the other domains.

Proposed sites of calcium binding

Calcium-binding requires from 6 to 8 oxygen atoms, which can be provided by side chains (including asparagine, aspartate, glutamine, glutamate, serine, and threonine), backbone carbonyls,

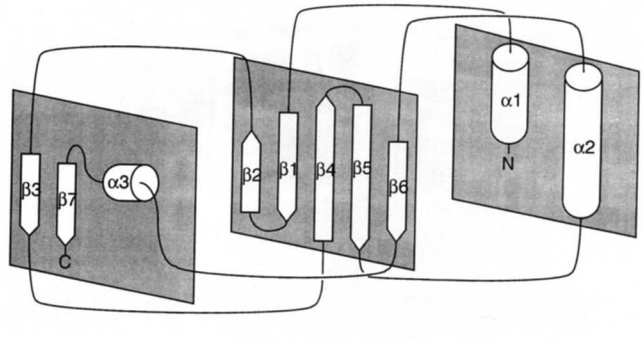

Fig. 6. Topology diagram for villin 14T. Arrows represent strands of β-sheet; cylinders represent α-helices. Sequential connectivities are made by lines. The 3 layers of the structure are represented by the shaded planes.

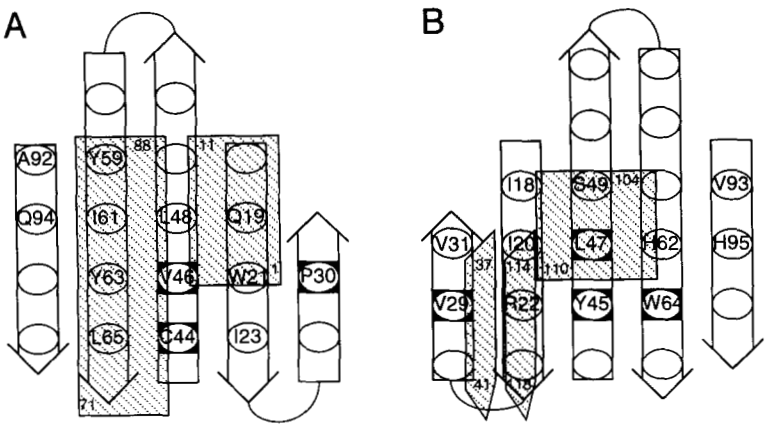

Fig. 7. Hydrophobic regions on the β-sheet and their interactions with other structural elements. **A:** The face of the sheet interacting with helices α1 and α2. **B:** The other face of the sheet, interacting with helix α3 and strand β3. The hydrophobic regions are depicted schematically. The strands of the β-sheet are depicted as arrows. Amino acid residues with side chains pointing up from the sheet are drawn as ovals. Residues that form the hydrophobic region are labeled. Strongly conserved residues are enclosed in black boxes. The α-helices that interact with the central β-sheet are shown as hatched rectangular boxes. The strands of parallel β-sheet that interact with the central β-sheet are shown as hatched angled arrows. Note that the 2 β-sheets are not parallel, but meet at an angle. The elements of secondary structure that interact with the central β-sheet are labeled with numbers denoting their initial and final residues.

and water molecules (Strynadka & James, 1991). The calcium titration we performed does not report directly on calcium ligands; it reports on the backbone amide protons that could be involved in hydrogen bonds to calcium ligands. The changes in chemical shift probably reflect the rearrangement of hydrogen bonds as calcium is ligated. This idea is consistent the observation that the residues that had the largest chemical shift changes upon Ca^{2+} binding also had the largest changes during the pH titration. Consider the pH titration from low to neutral pH. It appears that deprotonation of the carboxyl side chains involved in Ca^{2+} binding leads to formation of hydrogen bonds from the carboxylates to some of the amide groups. Upon addition of calcium, Ca^{2+} binding competes with these intramolecular hydrogen bonds, and the Ca^{2+} titration has the opposite effect on the chemical shifts from the pH titration. The result is that the ^{15}N-HSQC spectrum of the Ca^{2+}-loaded protein at pH 6.91 is similar to that of the Ca^{2+}-free protein at pH 4.15.

Based on the calcium titration data, we have localized 2 calcium binding sites. To localize the binding sites, we focused our attention on side chain carboxylates, since the negative charge of the carboxylate at pH 6.91 would offset the positive charge of the calcium ion, and surface charges have been found to be an important factor in ion binding for the calcium-binding protein calbindin D_{9k} (Linse et al., 1988). Figure 8A and Kinemage 3 show the region of the protein involved in the stronger calcium-binding site. The amide nitrogens and protons that shift the most upon calcium binding are shown in blue. All carboxylate side chains within 6 Å of these amide nitrogens and protons are shown in red; backbone carbonyls within this range are shown in orange. Three carboxylate side chains stand out in the center of this region: Glu^{24}, Asp^{43}, and Glu^{73}. The same analysis is shown for the weaker site in Figure 8B and Kinemage 4. The important side chains for this site are identified as Asp^{85} and Glu^{86}. Although we have localized the calcium-binding sites, the relatively low calcium affinity makes the biological role for this binding unclear. The dissociation constant for calbindin D_{9k}, another protein found in intestinal epithelial cells, is on the order of 0.1 μM in similar salt concentrations (Linse et al., 1991). The calcium affinity of villin 14T may increase in the context of the full-length protein or in the presence of actin.

Mapping the region conserved among actin-severing proteins

Actin-severing proteins share a region of high sequence homology that maps to residues 29–77 in villin 14T (Bazari et al., 1988) (highlighted in magenta in Kinemage 1). In the structure, these residues correspond to the turn at the end of strand β2 of the central β-sheet into strand β3 of the parallel β-sheet, strands β4 and β5 of the central β-sheet, and the beginning of helix α2, and include about half the residues that comprise the hydrophobic patches on both sides of the sheet. Because the region of sequence homology corresponds to internal and hydrophobic residues, the conserved sequence corresponds to the structural core of the domain. Consistent with this idea are the results of chemical modification experiments that show that methionine and cysteine residues within this region resist chemical modification and are therefore inaccessible to solvent (Nakayama T, Way M, Weeds A, Matsudaira P, manuscript in prep.; Matsudaira P, unpubl. results).

Mapping the actin monomer-binding site

A less well-conserved region, characteristic of actin monomer-binding proteins, maps to residues 82–102, finishing helix α2 and including strand β6 of the central β-sheet (Way et al., 1992). These residues are shown in blue on the space-filling model in Figure 9 and in Kinemage 1. Chemical crosslinking of cysteine-replacement mutants of villin (Nakayama T, Way M, Weeds A, Matsudaira P, manuscript in prep.) to monomer actin shows that a cysteine residue at position 90 in 14T (green in Fig. 9) can be crosslinked to actin Cys^{374}. This position at the N-terminal end of helix α2 is near a conserved pentapeptide (residues 84–88) that had been advanced as a possible actin-binding site (Vandekerckhove, 1989). However, site-directed mutagenesis of gelsolin domain one (Way et al., 1992) suggests that this pentapeptide has a minimal effect on actin binding. Rather, residues corresponding to His^{95}, Arg^{96}, Glu^{97}, and Val^{98} in the villin sequence, part of strand α6 near the C-terminal end of the helix, seem to affect actin binding (Way et al., 1992). Taken with the chemical crosslinking, this localizes the actin monomer-binding site to strand α6 of the central β-sheet. Interestingly, His^{95} is across the sheet from His^{62} and within 15 Å of the only other histidine in 14T, His^{101}. (The histidines are pink in Figure 9.) This is somewhat reminiscent of the actin-binding protein hisactophilin, for which the large number of histidine residues was postulated to be important for actin binding (Habazettl et al., 1992).

Comparison with the actin-severing protein gelsolin

While this manuscript was in preparation, the structure of another member of the actin-severing family, gelsolin, was pub-

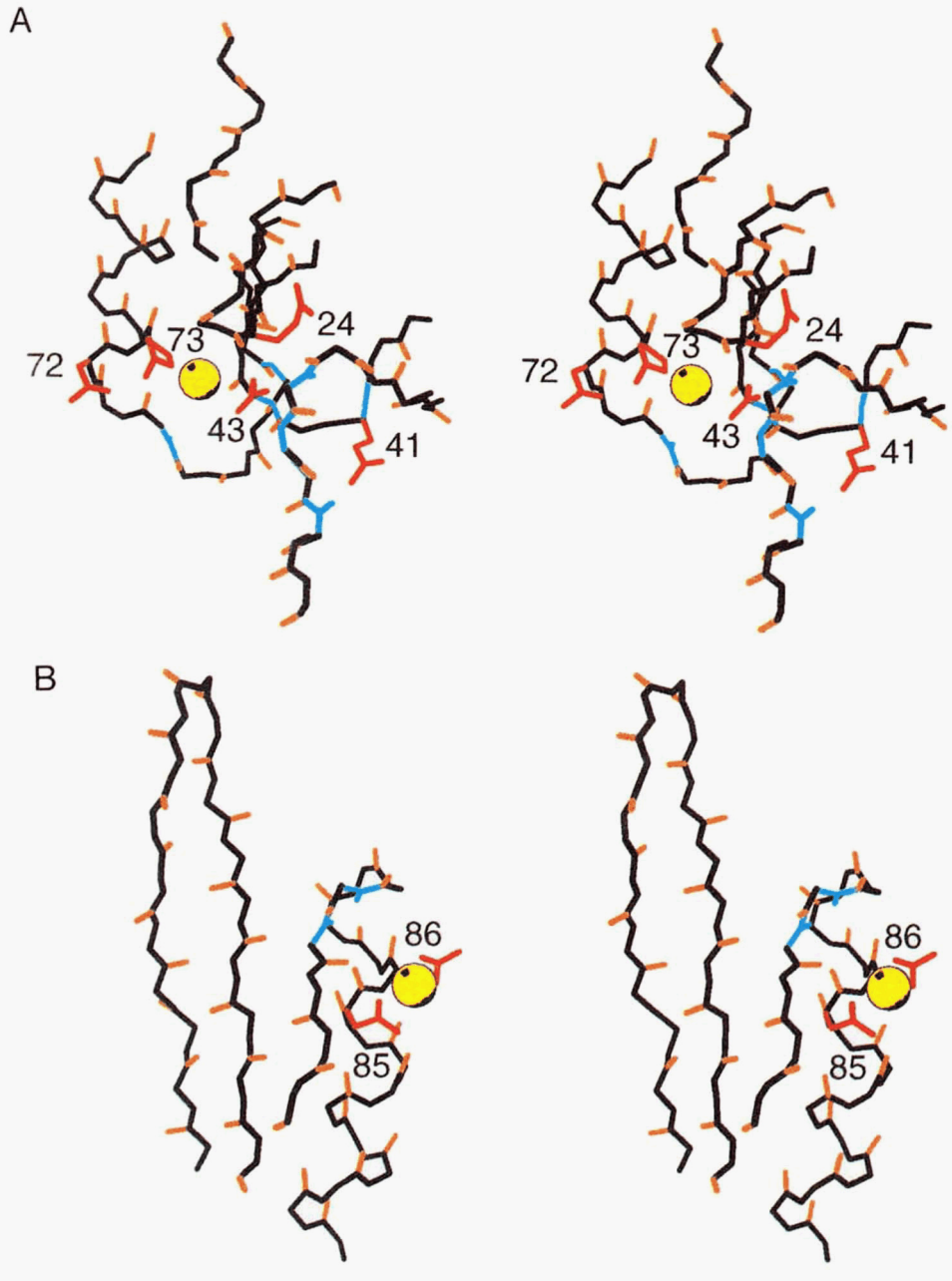

Fig. 8. Localizing the calcium-binding sites. **A:** Region near the stronger binding site. **B:** Region near the weaker binding site. Amide nitrogens and protons that change significantly in chemical shifts are shown in blue. Nearby carboxylate side chains are shown in red. Nearby backbone carbonyl oxygen atoms are in orange. Only side chains for aspartate and glutamate are shown. The yellow sphere in each figure represents a possible location for the calcium ion.

lished (McLaughlin et al., 1993). The structure is based on a co-crystal of the first domain of gelsolin (segment 1) with actin. Segment 1 has 54% sequence identity to villin 14T. Comparison of the structures reveals that they have the same overall topology, though strand $\beta 2$, at the edge of the main sheet, and strand $\beta 3$, in the small parallel sheet, are not present or were not identified in the gelsolin structure. A detailed comparison is given in Table 2. The gelsolin structure contains 2 calcium ions, one in an intramolecular site and another shared between gelsolin and actin. The stronger calcium-binding site suggested by the NMR data corresponds to the intramolecular site in gelsolin, while the weaker site corresponds to the intermolecular site. (See Table 2.) A more detailed, atomic-level comparison of the structures will provide insight into both the mechanism for binding monomer actin and the reasons for the different actin and calcium affinities of villin 14T versus gelsolin segment 1.

Summary

Villin 14T is a compact domain organized around a hydrophobic core of conserved residues. The conservation of the core suggests that all domains of the actin-severing proteins are based

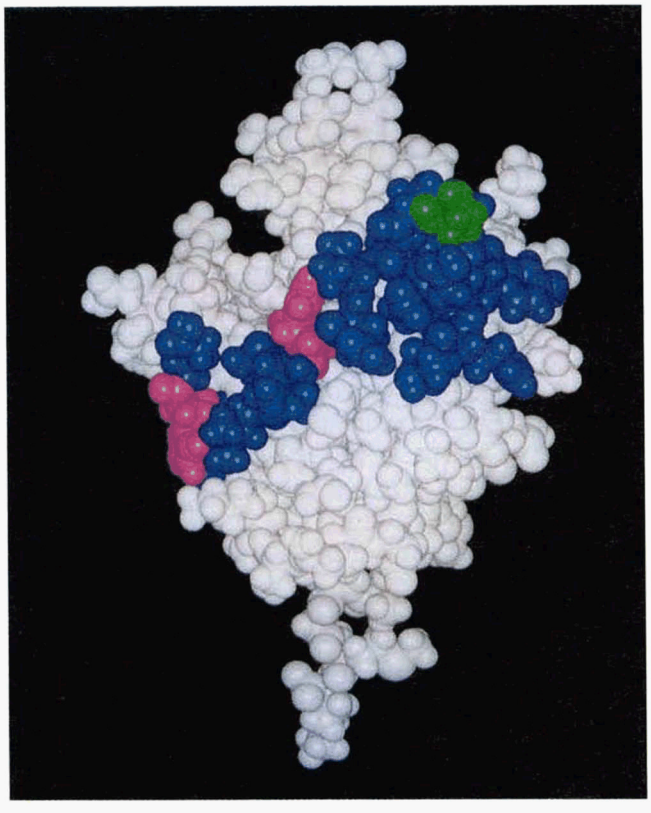

Fig. 9. Space-filling model of villin 14T showing the putative actin-binding surface. All atoms are shown; the orientation is rotated by 45° about the vertical axis from Figure 2A. The residues conserved among actin monomer-binding proteins are blue. The residue in position 90, which when mutated to cysteine can be crosslinked to actin, is in green. Histidine residues His^{62}, His^{95}, and His^{101} are shown in pink. Note that His^{62} and His^{95} are close together, on adjacent strands of the central β-sheet, and appear as one pink region in the center of the figure. All three histidine side chains of villin 14T are on the same face of the molecule, and mutagenesis of gelsolin implies that His^{95} may be involved in actin monomer binding.

on the central β-sheet and long α-helix found in the fold of 14T. Nonconserved sequences, primarily along the surface-exposed edge of the central β-sheet, form potential sites for binding actin monomers. Two calcium binding sites are identified, one at either end of the long helix. In the absence of actin, the structure of this prototypic domain provides a basis for understanding structural changes in villin as it caps, severs, and bundles actin filaments.

Materials and methods

Preparation of protein samples

Villin 14T (126 amino acids) was produced in *Escherichia coli*; expression and purification of villin 14T are described in a forthcoming paper (Nakayama T, Way M, Weeds A, Matsudaira P, manuscript in prep.). To label protein with ^{15}N and ^{13}C isotopes for spectroscopy, cells were grown in minimal medium with ^{13}C glucose (Isotec, Miamisburg, Ohio) as the sole carbon source and $^{15}NH_4Cl$ or $(^{15}NH_4)_2SO_4$ (Isotec) as the sole nitrogen source (McIntosh & Dahlquist, 1990). Samples included protein labeled with ^{15}N, ^{13}C and ^{15}N, and ^{13}C to 10% as well as unlabeled protein. The protein was judged to be more than 95% pure by SDS-PAGE. Yields were as high as 90 mg of purified protein from 1 L of cell growth, even in minimal medium. For structure determination, the protein was dialyzed into NMR buffer (50 mM NaH_2PO_4, pH 4.15, 100 mM NaCl, and 0.1 mM NaN_3) and concentrated to a sample volume of approximately 500 μL. Sample concentrations ranged from 1.65 to 6.72 mM, depending on the availability of material.

Table 2. *Comparison of the structures of villin 14T and gelsolin segment 1*[a]

Structural element	Sequence position	
	Villin 14T	Gelsolin segment 1
α1	1–11	30–34 (6–10)
β1	17–24	41–45 (18–22)
β2	27–31	Not identified
β3	37–41	Not identified
β4	43–53	67–74 (44–51)
β5	56–66	81–90 (57–66)
α2	71–88	95–112 (71–88)
β6	92–98	119–122 (95–98)
α3	104–110	128–132 (104–108)
β7	114–118	138–141 (112–117)
Strong Ca^{2+} site	E24, D43, or E73	D66 and E97 (D43 and E73)
Weak Ca^{2+} site	D85 or E86	D109 (D85)

[a] In the Gelsolin segment 1 column, residue numbers using the human gelsolin numbering are followed in parentheses by the corresponding numbers in the villin sequence, using the alignment of Way and Weeds (1988). The carboxylate side chains near the strong calcium-binding site in villin 14T are compared with the carboxylate side chains found in the intramolecular binding site in gelsolin segment 1. The weak site is compared with the intermolecular site.

NMR data collection and processing

NMR measurements were performed on AMX500 and AMX600 spectrometers (Bruker, Karlsruhe, Germany) at 25 °C. Data were processed using the Felix software package (Hare Research Inc., Bothell, Washington). The HNCA and HN(CO)CA pair of experiments was used for sequential assignments (Kay et al., 1990; Bax & Ikura, 1991; Grzesiek & Bax, 1992). For side chain assignments, experiments included ^{15}N-dispersed total correlation spectroscopy (TOCSY) (Marion et al., 1989a) and a ^{13}C heteronuclear cross-polarization experiment (Majumdar et al., 1993), as well as conventional TOCSY (Braunschweiler & Ernst, 1983) and double quantum filtered correlation spectroscopy (COSY) (Piantini et al., 1982).

NOE restraints

NOE restraints were based on experiments including 3D ^{15}N- and ^{13}C-dispersed NOE spectroscopy (NOESY) (Marion et al., 1989a, 1989b) in addition to 2D NOESY in H_2O and D_2O. The spectra were recorded with mixing times of 100 and 150 ms, and crosspeaks were only loosely quantitated by counting plotted contour levels. Crosspeak intensity was correlated with distance for use in distance geometry calculations by examining the in-

tensity of crosspeaks between backbone protons involved in regular secondary structure and assuming regular geometries. Peaks were then classified as strong, with an associated upper distance limit of 3.0 Å, intermediate, with a limit of 4.0 Å, and weak, with a limit of 5.0 Å. The difference between the number of assigned NOE crosspeaks and the number of distance restraints for structure calculation is due to methylene protons that lack stereospecific assignments.

Restraints on the dihedral angles and stereospecific assignments

The α-proton to amide proton scalar coupling constant ($^3J_{HNH\alpha}$) was determined from crosspeak splitting in the proton dimension of an ^{15}N heteronuclear correlation spectrum and confirmed with data from the HNHA experiment (Vuister & Bax, 1993). The backbone torsion angle ϕ was then restrained to the range $-90°$ to $-40°$ for $^3J_{HNH\alpha} \leq 5.5$ Hz and $-160°$ to $-80°$ for $^3J_{HNH\alpha} > 8.0$ Hz. The side chain torsion angle χ_1 was restrained based on the α-proton to β-proton coupling constant ($^3J_{H\alpha H\beta}$), classified as large or small from the COSY data, and the amide nitrogen to β-proton coupling constant ($^3J_{NH\beta}$), classified as large or small from the HNHB experiment (Archer et al., 1991). Stereospecific assignments for methyl groups of valine and leucine were obtained from spectra acquired on the 10% ^{13}C-labeled sample (Neri et al., 1989; Senn et al., 1989; Szyperski et al., 1992).

Hydrogen bond restraints

The hydrogen bonds were assigned for regions of regular secondary structure based on hydrogen exchange data (Markus MA, Nakayama T, Matsudaira P, Wagner G, manuscript in prep.) and characteristic NOE patterns. They are distributed as follows: 22 bonds in the central β-sheet, 3 in the parallel β-sheet, 13 in helix α2, and 5 in the helix α3. For each hydrogen bond, one restraint holds the distance between the amide proton and the oxygen to less than 2.3 Å. A second restraint holds the amide nitrogen to oxygen distance in the range 2.5–3.3 Å to maintain a somewhat linear bond geometry.

Structure calculations

The structure calculations used the distance geometry (DG) algorithm implemented in the program DGII (Havel, 1991), used as part of the INSIGHT II software package (Biosym Technologies, San Diego, California). For optimization in DGII, an initial energy of 4,600 kcal/mol with a time step of 0.39 ps gave convergence for 12 out of 20 structures to a final DG error of less than 2.00. Two further structures were discarded due to large local distance restraint violations. The coordinates for the 10 best structures have been deposited with the Brookhaven Protein Data Bank (file 1VIL). Superposition and analysis of the structures made use of the INSIGHT II software.

Calcium titration experiments

For the calcium titration, villin 14T was exchanged into 50 mM deuterated acetic acid, 100 mM NaCl, and 0.1 mM NaN_3 in 7% D_2O by ultrafiltration (Amicon Division, W.R. Grace & Co., Beverly, Massachusetts). Local changes in the protein were monitored with ^{15}N-HSQC experiments (Bodenhausen & Ruben, 1980), run with 8 scans and 256 t_1 increments. First the pH of the sample was titrated from 4.10 to 6.91 to allow the transfer of assignments. Then small volumes of $CaCl_2$, diluted with the acetate buffer, were added to the sample. Some of the peaks in the ^{15}N-HSQC spectra move as calcium is added. (See the Results section.) No broadening or splitting of peaks is observed, which suggests that the exchange between the free and Ca^{2+}-bound states of the protein is fast on the NMR time scale. If we assume that the change in chemical shift is due to calcium binding and that the titration goes to full occupancy of the Ca^{2+}-binding site, we can calculate the fraction of protein in the Ca^{2+}-bound state as the change in chemical shift divided by the change in chemical shift at the highest calcium concentration. Binding constants are calculated from the fraction bound for each site as a function of total calcium concentration, assuming 2 independent binding sites.

Supplementary material on Diskette Appendix

A complete list of the restraints used to calculate the structures appears on the Diskette Appendix (file Markus.NMR, SUPLEMNT directory). Four kinemages (file Markus.kin, KINEMAGE directory) are also provided. Kinemage 1 shows the backbone of villin 14T with conserved regions highlighted, Kinemage 2 shows the side chains of the hydrophobic core, Kinemage 3 shows the stronger calcium-binding site, and Kinemage 4 shows the weaker calcium-binding site.

Acknowledgments

This work was supported by NIH grant GM38608 to G.W. and grant DK35306 to P.M. Computer facilities used for the structure calculations were funded in part by the W.M. Keck Foundation. We thank Michael Way for the plasmid used in expression of villin 14T and Lawrence McIntosh for acquiring the initial spectra and stimulating interest in the project. We also thank Andrzej Krezel for advice and assistance with computers, Daniel Wyss for assistance with analysis of structures, and N.R. Nirmala, K. Chandrasekhar, and Peter Schmieder for assistance with spectroscopy.

References

Archer SJ, Ikura M, Torchia DA, Bax A. 1991. An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. *J Magn Reson 95*:636–641.

Archer SJ, Vinson VK, Pollard TD, Torchia DA. 1993. Secondary structure and topology of *Acanthamoeba* profilin I as determined by heteronuclear nuclear magnetic resonance spectroscopy. *Biochemistry 32*:6680–6687.

Bax A, Ikura M. 1991. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue in uniformly $^{15}N/^{13}C$ enriched proteins. *J Biomol NMR 1*:99–104.

Bazari WL, Matsudaira P, 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 DJ. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett 69*:517–552.

Braunschweiler L, Ernst RR. 1983. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J Magn Reson 53*:521–528.

Grzesiek S, Bax A. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson 96*:432–440.

Habazettl J, Gondol D, Wiltscheck R, Otlewski J, Schleicher M, Holak TA. 1992. Structure of hisactophilin is similar to interleukin-1β and fibroblast growth factor. *Nature 359*:855–858.

Havel TF. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Progr Biophys Mol Biol 56*:43–78.

Hyberts SG, Goldberg MS, Havel TF, Wagner G. 1992. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci 1*:736–751.

Janmey PA, Matsudaira PT. 1988. Functional comparison of villin and gelsolin. *J Biol Chem 263*:16738–16743.

Kay LE, Ikura M, Tschudin R, Bax A. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson 89*:496–514.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Linse S, Brodin P, Johansson C, Thulin E, Grundström T, Forsén S. 1988. The role of protein surface charges in ion binding. *Nature 335*:651–652.

Linse S, Johansson C, Brodin P, Grundström T, Drakenberg T, Forsén S. 1991. Electrostatic contributions to the binding of Ca^{2+} in calbindin D_{9k}. *Biochemistry 30*:154–162.

Majumdar A, Wang H, Morshauser RC, Zuiderweg ERP. 1993. Sensitivity improvement in 2D and 3D HCCH spectroscopy using heteronuclear cross-polarization. *J Biomol NMR 3*:387–397.

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM. 1989a. Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1β. *Biochemistry 28*:6150–6156.

Marion D, Kay LE, Sparks SW, Torchia DA, Bax A. 1989b. Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J Am Chem Soc 111*:1515–1517.

Matsudaira P, Janmey P. 1988. Pieces in the actin-severing protein puzzle. *Cell 54*:139–140.

McIntosh LP, Dahlquist FW. 1990. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q *Rev Biophys 23*:1–38.

McLaughlin PJ, Gooch JT, Mannherz HG, Weeds AG. 1993. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. *Nature 364*:685–692.

Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. 1989. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry 28*:7510–7516.

Piantini U, Sorensen OW, Ernst RR. 1982. Multiple quantum filters for elucidating NMR coupling networks. *J Am Chem Soc 104*:6800–6801.

Pope B, Way M, Weeds AG. 1991. Two of the three actin-binding domains of gelsolin bind to the same subdomain of actin: Implications for capping and severing mechanisms. *FEBS Lett 280*:70–74.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–337.

Senn H, Werner B, Messerle BA, Weber C, Traber R, Wüthrich K. 1989. Stereospecific assignment of the methyl ^1H NMR lines of valine and leucine in polypeptides by nonrandom ^{13}C labelling. *FEBS Lett 249*:113–118.

Strynadka NCJ, James MNG. 1991. Towards an understanding of the effects of calcium on protein structure and function. *Curr Opin Struct Biol 1*:905–914.

Szyperski T, Neri D, Leiting B, Otting G, Wüthrich K. 1992. Support of ^1H NMR assignments in proteins by biosynthetically directed fractional ^{13}C-labeling. *J Biomol NMR 2*:323–334.

Vandekerckhove J. 1989. Structural principles of actin-binding proteins. *Curr Opin Cell Biol 1*:15–22.

Vuister GW, Bax A. 1993. Quantitative J correlation: A new approach for measuring homonuclear three-bond $J(H^NH^\alpha)$ coupling constants in ^{15}N-enriched proteins. *J Am Chem Soc 115*:7772–7777.

Way M, Pope B, Weeds AG. 1992. Are the conserved sequences in segment 1 of gelsolin important for binding actin? *J Cell Biol 116*:1135–1143.

Way M, Weeds AG. 1988. Nucleotide sequence of pig plasma gelsolin: Comparison of protein sequence with human gelsolin and other actin-severing proteins shows strong homologies and evidence for large internal repeats. *J Mol Biol 203*:1127–1133.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: John Wiley & Sons.