Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 4986–4990, July 1988
Biochemistry

Villin sequence and peptide map identify six homologous domains

(actin-binding proteins/calcium-binding proteins/conformational changes/intestinal proteins)

WENDY LEVOY BAZARI*, PAUL MATSUDAIRA*†, MALGORZATA WALLEK*, TOD SMEAL*, ROSS JAKES‡, AND YASHI AHMED*

*Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142; and ‡Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, United Kingdom

Communicated by Harvey F. Lodish, March 3, 1988

ABSTRACT **Site-specific proteases and antisera to the amino terminus of villin have been used to show that villin is organized into seven protease-resistant domains. Six are contained in the amino-terminal M_r 87,000 villin core, a Ca^{2+}-regulated actin-severing fragment, whereas the carboxyl-terminal domain includes the villin "headpiece," a fragment involved in bundling of actin filaments. Ca^{2+} inhibits proteolytic cleavage between domains in the amino-terminal half of villin. The protein sequence of villin deduced from a single cDNA clone contains a conserved sequence that is repeated six times and is found in each domain of the villin core. The conserved repeats are found in other actin-severing proteins but not in the villin headpiece. Our results suggest that actin-severing proteins are organized around a common M_r 14,000–17,000 domain.**

Villin (M_r 95,000) is a major cytoskeletal protein in microvilli from brush-border cells of intestine and kidney (1). Binding studies have shown that villin crosslinks actin filaments into bundles at low Ca^{2+} concentrations, but at high (greater than micromolar) Ca^{2+} concentrations, villin caps and severs actin filaments to short lengths (1, 2). Limited proteolysis of native villin generates large amino-terminal fragments, the villin core (M_r 87,000) (3, 4) and 44T (M_r 44,000) (5), that retain Ca^{2+}-regulated actin-severing properties and a carboxyl-terminal M_r 8700 fragment (villin "headpiece") that is required for the actin-crosslinking activity (3, 4). Further proteolysis of 44T produces a M_r 14,000 amino-terminal fragment (44T-14T), which suggests the presence of smaller domains (6). Similarities in sequences, actin-binding activities, and patterns of proteolytic cleavage suggest that villin is related to gelsolin, a M_r 85,000 actin-severing protein found in vertebrate cells and sera (reviewed in refs. 7 and 8). Gelsolin is also cleaved in half by proteases, but unlike villin the amino-terminal half of gelsolin displays Ca^{2+}-insensitive actin-severing activity (9–12). Further proteolysis generates fragments whose sizes suggest that both halves of gelsolin contains small (M_r 14,000) and large (M_r 30,000) domains.

Villin and gelsolin are structurally and functionally similar to fragmin (13, 14) and severin (15, 16), M_r 44,000 proteins isolated from *Physarum* plasmodia and *Dictyostelium* amoeba. Comparisons of the gelsolin sequence (17) with the fragmin (18) and severin (19) sequences showed a 350-residue region of homology. Further analysis of the sequences of all three proteins indicated that there is a smaller (100-residue) repeated sequence within the 350-residue sequence, but there is no correlation of the reported number and size of the smaller repeats with structural or functional regions of the proteins.

We have been studying the relationship between villin structure and function to explain the role of Ca^{2+} in regulating actin-severing and -bundling properties. In the work reported here, we used amino-terminal-specific antisera and site-specific proteolysis to show that villin is organized into seven structural domains and that Ca^{2+} alters the conformation of the amino-terminal half of villin. Analysis of the villin protein sequence deduced from cDNA§ showed the presence of a 37- to 51-residue conserved sequence that is repeated six times in the villin core and gelsolin and three times in fragmin and severin. The repeats map within the six domains of the villin core, suggesting that the domains of actin-severing proteins are homologous.

MATERIALS AND METHODS

Proteins. Villin and its proteolytic fragments 44T and 51T were purified as described (5). Trypsin, *Staphylococcus aureus* V8 protease, and chymotrypsin were obtained from Boehringer Mannheim. Restriction enzymes and polymerases were obtained from Pharmacia.

Construction of Chicken Intestine cDNA Library. RNA was isolated from chicken intestines by the guanidinium isothiocyanate method (20), and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose column chromatography. cDNA was synthesized (21) from 20 μg of poly(A)$^+$ RNA, attached to *Eco*RI linkers, and size-fractionated by electrophoresis in agarose gels. cDNAs longer than 1.0 kilobase were ligated into the *Eco*RI site of the bacteriophage expression vector λgt11 and packaged *in vitro* (22) (Gigapack, Stratagene, San Diego, CA). The resulting library, containing 2×10^5 independent recombinants, was amplified and stored at 4°C.

Cloning and Sequencing of Villin cDNA. Transformants (a total of 10^4 clones) were screened with rabbit polyclonal antisera (R211.3 and R200.2) that are crossreactive with the amino- and carboxyl-terminal residues of villin (6). A full-length clone (3.0 kilobases) that was crossreactive with both antibodies was identified and its insert was subcloned into phage M13mp8. Both strands of the clone were sequenced by the dideoxy chain-termination method using DNA polymerase I Klenow fragment (23) or phage T7 polymerase (24) (Sequenase, United States Biochemical, Cleveland) and dITP in place of dGTP. The entire nucleotide sequence was determined on both strands from a single clone. A single base deletion at nucleotide 1332 was discovered in the cDNA sequence as a frameshift in the protein sequence when compared to villin and gelsolin protein sequences. To identify the missing base, two independent clones were sequenced by using a 21-base oligonucleotide primer that was complementary to a region adjacent to the deletion. Protein sequence data bases were compared with the entire villin protein sequence and conserved regions of sequences by using

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03781).

FASTHP and BESTFIT (University of Wisconsin Genetics Computer Group programs) (25).

Proteolysis of Villin, 44T, and 51T. Villin was digested for various times with trypsin (1:1000, wt/wt), chymotrypsin (1:100), or *S. aureus* V8 protease (1:2000) under conditions described previously (5). The proteolytic fragments were separated by microslab $NaDodSO_4$/PAGE (26) and stained with Coomassie blue or electrophoretically transferred onto nitrocellulose membranes and processed for immunoblotting. Amino-terminal fragments were detected with antiserum R211.3. Fragments 44T and 51T were purified from a tryptic digest of villin as described (5) except that the DEAE-Sephacel column was replaced with a Pharmacia Mono Q ion-exchange column.

Amino-Terminal Sequence Analysis. Digests of 44T and 51T (50 pmol) were subjected to $NaDodSO_4$/PAGE in a 10–20% gradient of polyacrylamide and electroblotted onto polyvinylidene difluoride membranes (Immobilon transfer membrane, Millipore) (27). The Coomassie blue-stained bands were cut from the membrane and directly sequenced in an Applied Biosystems (Foster City, CA) model 470 gas-phase machine equipped for on-line analysis of phenylthiohydantoin derivatives of amino acids.

RESULTS

Villin cDNA and Protein Sequence. Two peptide antisera (9), crossreactive with the amino and carboxyl termini of villin, identified a clone that expressed full-length villin cDNA in a λgt11 library prepared from poly(A)$^+$ RNA isolated from chicken intestine epithelial cells. The nucleotide sequence and deduced amino acid sequence are shown in Fig. 1. The cDNA encoded an 825 amino acid protein with a molecular mass of 92,248 daltons and a net negative charge. These values compare favorably with the molecular weight (M_r 95,000) estimated by $NaDodSO_4$/PAGE and with the acidic isoelectric point of the protein. The villin core consisted of residues 1–749, and the villin headpiece corresponded to residues 750–825. The amino acid sequence agreed exactly with the reported sequence of chicken intestine villin headpiece (4) and shared 54% identical residues with the protein sequence of human intestine villin headpiece deduced from a cDNA clone (28).

Analysis of the protein sequence showed that villin contained two 280-residue sequences that are 22% identical. These repeats (residues 21–314 and 402–678) mapped within the amino- and carboxyl-terminal halves of the villin core. Comparison between the villin core and human gelsolin showed a 48% identity in their protein sequences. The areas of similarity were distributed evenly throughout the sequence and confirmed proposals that the two proteins are homologous. A similar tandem duplication of sequences has been found in the gelsolin sequence (17). A search of the protein and nucleic acid data bases[¶][‖] did not reveal any other statistically significant matches. Reported similarities between gelsolin and the sequences of actin (17) and the tyrosine kinase cytoplasmic substrate p36 (29) were not found in analogous regions of the villin sequence. Although villin binds three Ca^{2+} ions (30), comparisons of the villin sequence with other Ca^{2+}-binding proteins did not identify regions that contain Ca^{2+}-binding sites.

In addition to the tandemly duplicated sequence, there was a sixfold repeated sequence (Fig. 2). The conserved sequence consisted of 37–51 residues that were located in six regions of the villin core. Comparisons between villin and other actin-severing proteins showed that similar sequences were present in six regions of gelsolin and in three regions of fragmin and severin.

Map of the Protease Cleavage Sites. The location of the repeated sequences was then compared with the organization of villin structural domains. Connecting segments between structural domains of large proteins are often preferentially cleaved by mild proteolysis (31). At low concentrations, trypsin, chymotrypsin, and V8 protease cleave villin into a similar pattern of M_r 44,000, 30,000, and 14,000 subfragments (Fig. 3). Although the proteases produced similar banding patterns on gels, we could not conclude that the three proteases cleaved the same regions in the protein. Therefore, we used antibodies specific for the villin amino terminus to map the locations of the protease cleavage sites (6). On immunoblots of the villin proteolytic fragments (Fig. 3*b*), the amino-terminus-specific antibody showed that at least two of the proteases cleaved villin at six sites located approximately 87,000, 75,000, 64,000, 44,000, 32,000, and 14,000 daltons from the amino terminus. A map of the cleavage sites identified the boundaries of seven protease-resistant domains of villin. If these domains are numbered in order from the amino terminus, then the villin core contains domains 1–6, the villin headpiece includes domain 7, fragment 44T contains domains 1–3, and 51T contains domains 4–7.

We purified 44T and 51T and cleaved them separately to identify the sites of cleavage more accurately. Partial cleavage of 44T with trypsin and chymotrypsin (Fig. 4*a*) and V8 protease (data not shown) required approximately 20-fold more protease and produced subfragments of $M_r \approx 28,000$ and 14,000 in the presence of EGTA and Ca^{2+}. The cleavage sites were identified directly by sequencing the amino termini of subfragments that were electroblotted onto Immobilon membranes. The amino-terminal sequences (Fig. 4*b*) showed that in the presence of EGTA, the proteases cleaved 44T at two sites and produced two subfragments with identical amino-terminal sequences. In the presence of Ca^{2+}, 44T was cleaved between domains 1 and 2 to yield two complementary subfragments. Tryptic and chymotryptic digests of 51T produced a similar pattern of subfragments of M_r 44,000, 28,000, 14,000, and 8500. The amino-terminal sequences showed that trypsin cleaved 51T into two complementary subfragments, whereas chymotrypsin produced three complementary subfragments.

DISCUSSION

Our results (summarized in Fig. 4*c*) show that villin is organized into seven structural domains and that the six domains of villin core are homologous. Domain 7 contains the villin headpiece and shows no similarity in sequence with other regions of villin, although actin- and Ca^{2+}-binding sites are located in both the villin headpiece and a M_r 44,000 domain of the villin core (30, 32). The short homologous repeats define a segment of each domain that is structurally similar in all actin-severing proteins, but whether these regions represent actin- or Ca^{2+}-binding sites is not known. Preliminary studies showed that the cysteine (6) and methionine residues in the homologous repeats cannot be chemically modified unless villin is unfolded by denaturants (M.W. and P.M., unpublished data). These observations suggest that the regions of homology in each domain are inaccessible to solvent and thus might not provide residues that lie in a binding site for actin or Ca^{2+}. Conserved sequences in other regions of the domain might indicate regions important for the function of the domains. Domains 1 and 4 contain sequences not shared by the other domains. For example, there are conserved tryptophan residues at positions 21 and 402 of

[¶]Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 12.
[‖]EMBL/GenBank Genetic Sequence Database (1987) GenBank (IntelliGenetics, Mountain View, CA), Tape Release 53.

```
                                                                5'-----CGGGAAGGTCCCCTGCCTGCAGCCAACACCATG   -1

GTG GAG CTC AGC AAA AAG GTC ACC GGG AAG CTG GAC AAG ACC ACG CCG GGC ATC CAG ATA TGG AGA ATC GAG AAC ATG GAG ATG GTG CCG     90
VAL GLU LEU SER LYS LYS VAL THR GLY LYS LEU ASP LYS THR THR PRO GLY ILE GLN ILE TRP ARG ILE GLU ASN MET GLU MET VAL PRO     30

GTG CCC ACT AAA AGC TAT GGC AAC TTC TAC GAG GGG GAC TGC TAC GTG CTG CTG TCG ACA CGT AAG ACT GGG AGC GGC TTC AGC TAC AAC    180
VAL PRO THR LYS SER TYR GLY ASN PHE TYR GLU GLY ASP Cys TYR VAL LEU LEU SER THR ARG LYS THR GLY SER GLY Phe SER TYR ASN     60

ATC CAC TAC TGG CTG GGT AAG AAC TCG AGC CAG GAC GAG CAG GGG GCG GCC GCG ATC TAC ACC ACG CAG ATG GAT GAG TAT CTG GGC TCT    270
ILE HIS TYR TRP LEU GLY LYS ASN Ser Ser Gln Asp Glu Gln Gly Ala Ala Ala Ile Tyr Thr Thr Gln MET ASP GLU TYR LEU GLY SER     90

GTG GCC GTG CAG CAC CGT GAG GTC CAG GGC CAT GAG AGC GAG ACG TTC CGT GCA TAC TTC AAG CAG GGA CTC ATC TAT AAG CAG GGT GGG    360
VAL ALA VAL GLN HIS ARG GLU VAL GLN GLY HIS GLU SER GLU THR PHE ARG ALA TYR PHE LYS GLN GLY LEU ILE TYR LYS GLN GLY GLY    120

GTG GCC TCA GGC ATG AAG CAC GTG GAG ACC AAC ACC TAC AAC GTC CAG CGC CTG CTG CAT GTG AAG GGC AAG AAG AAC GTG GTG GCT GCA    450
VAL ALA SER GLY Met LYS His VAL GLU THR ASN THR TYR ASN VAL GLN ARG LEU LEU HIS VAL LYS GLY LYS LYS ASN VAL VAL ALA ALA    150

GAG GTG GAG ATG AGC TGG AAA AGC TTT AAC CTG GGC GAT GTC TTC CTG CTG GAC CTC GGC CAG CTC ATC ATC CAG TGG AAC GGC CCC GAG    540
GLU VAL GLU MET Ser Trp Lys Ser Phe ASN LEU GLY Asp Val Phe Leu Leu Asp Leu Gly Gln Leu Ile Ile Gln Trp Asn Gly Pro Glu    180

AGC AAT CGT GCC GAA AGG CTG AGG GCA ATG ACC CTG GCC AAG GAC ATT CGG GAC CGA GAG CGT GCG GGA CGT GCG AAG GTT GGC GTG GTG    630
Ser Asn Arg Ala Glu Arg Leu Arg Ala Met Thr Leu Ala Lys Asp Ile Arg Asp Arg Glu Arg Ala Gly Arg Ala Lys Val Gly Val Val    210

GAG GGG GAG AAT GAG GCG GCC TCA CCT GAG CTG ATG CAG GCC TTG ACA CAT GTG CTG GGG GAG AAG AAG AAC ATC AAG GCA GCC ACA CCC    720
Glu Gly Glu Asn Glu Ala Ala Ser Pro Glu Leu Met Gln Ala Leu Thr His Val Leu Gly Glu Lys Lys Asn Ile Lys Ala Ala Thr Pro    240

GAT GAA CAA GTT CAC CAG GCG CTC AAC AGT GCC CTC AAG CTC TAC CAT GTC TCT GAT GCC AGT GGG AAC CTG GTC ATA CAA GAG GTA GCA    810
Asp Glu Gln Val His Gln Ala Leu Asn Ser Ala Leu Lys Leu Tyr His Val Ser Asp Ala Ser Gly Asn Leu Val Ile Gln Glu Val Ala    270

ATT CGC CCC TTG ACT CAA GAT ATG CTC CAG CAT GAG GAC TGC TAC ATC CTT GAT CAA GCA GGT CTC AAG ATC TTT GTG TGG AAG GGC AAG    900
Ile Arg Pro Leu Thr Gln Asp Met Leu Gln His Glu Asp Cys Tyr Ile Leu Asp Gln Ala Gly Leu Lys Ile Phe Val Trp Lys Gly Lys    300

AAT GCC AAC AAG GAG GAG AAG CAG CAG GCA ATG AGC AGG GCC CTG GGC TTC ATC AAA GCC AAG AAC TAC CTG GCC AGC ACC AGC GTG GAG    990
Asn Ala Asn Lys Glu Glu Lys Gln Gln Ala Met Ser Arg Ala Leu Gly Phe Ile Lys Ala Lys Asn Tyr Leu Ala Ser Thr Ser Val Glu    330

ACA GAG AAC GAT GGG TCT GAG TCC GCT GTC TTC AGG CAG CTC TTC CAA AAA TGG ACT GTT CCC AAC CAA ACC AGT GGG CTG GGC AAG ACC   1080
Thr Glu Asn Asp Gly Ser Glu Ser Ala Val Phe Arg Gln Leu Phe Gln Lys Trp Thr Val Pro Asn Gln Thr Ser Gly Leu Gly Lys Thr    360

CAT ACT GTG GGC AAA GTG GCT AAG GTG GAA CAG CTG AAG TTT GAC GCC ACC ACA ATG CAT GTC AAG CCT GAA GTG GCC GCC CAG CAG AAG   1170
His Thr Val Gly Lys VAL ALA LYS VAL GLU GLN VAL LYS PHE ASP ALA THR THR MET His VAL LYS PRO GLU VAL ALA ALA GLN GLN LYS    390

ATG GTG GAT GAT GGA TCT GGG GAG GCA GAG GTC TGG CGG GTG GAG AAC CAG GAG CTG GTG CCC GTG GAG AAG CGG TGG TTG GGC CAT TTC   1260
MET VAL ASP ASP GLY SER GLY GLU ALA GLU VAL Trp Arg Val Glu Asn Gln Glu Leu Val Pro Val Glu Lys Arg Trp Leu Gly His Phe    420

TAC GGC GGG GAC TGC TAC CTG GTG CTC TAC ACC TAC TAC GTG GGG CCC AAG GTG AAC CGC ATC ATC TAC ATC TGG CAG GGC CGC CAT GCC   1350
Tyr Gly Gly Asp Cys Tyr Leu Val Leu Tyr Thr Tyr Tyr Val Gly Pro Lys Val Asn Arg Ile Ile Tyr Ile Trp Gln Gly Arg His Ala    450

AGC ACG GAT GAG CTG GCC GCC TCG GCC TAC CAA GCC GTC TTC CTG GAC CAG AAG TAC AAC AAC GAG CCT GTG CAG GTG CGC GTC ACC ATG   1440
Ser Thr Asp Glu Leu Ala Ala Ser Ala Tyr Gln Ala Val Phe Leu Asp Gln Lys Tyr Asn Asn Glu Pro Val Gln Val Arg Val Thr Met    480

GGC AAG GAG CCG GCC CAC CTG ATG GCG ATC TTC AAG GGC AAG ATG GTG GTG TAC GAG AAC GGC TCC TCG CGG GCG GGC GGC ACG GAG CCG   1530
Gly Lys Glu Pro Ala His Leu Met Ala Ile Phe Lys Gly Lys Met Val Val Tyr Glu Asn Gly Ser Ser Arg Ala Gly Gly Thr Glu Pro    510

GCG TCC TCC ACT CGG CTC TTC CAT GTG CAC GGC ACC AAC GAG TAC AAC ACC AAG GCC TTC GAG GTG CCC GTC CGA GCC GCT TCT CTC AAC   1620
Ala Ser Ser Thr Arg Leu Phe His Val His Gly Thr Asn Glu Tyr Asn Thr Lys Ala Phe Glu Val Pro Val Arg Ala Ala Ser Leu Asn    540

TCC AAC GAT GTC TTT GTG CTC AAG ACG CCC AGC TCC TGC TAC CTC TGG TAT GGG AAG GGC TGC AGC GGG GAT GAG CGT GAG ATG GGC AAG   1710
Ser Asn Asp Val Phe Val Leu Lys Thr Pro Ser Ser Cys Tyr Leu Trp Tyr Gly Lys Gly Cys Ser Gly Asp Glu Arg Glu Met Gly Lys    570

ATG GTG GCC GAC ATC ATC TCC AAG ACG GAG AAG CCG GTG GTC GCT GAA GGG CAG GAG CCG CCC GAG TTC TGG GTA GCT CTG GGC GGC AAG   1800
Met Val Ala Asp Ile Ile Ser Lys Thr Glu Lys Pro Val Val Ala Glu Gly Gln Glu Pro Pro Glu Phe Trp Val Ala Leu Gly Gly Lys    600

ACC AGC TAC GCC AAC AGC AAG AGG CTA CAG GAA GAG AAT CCC TCT GTG CCC CCC CGG CTC TTT GAA TGC TCT AAC AAG ACA GGC AGG TTC   1890
Thr Ser Tyr Ala Asn Ser Lys Arg LEU GLN GLU GLU ASN PRO SER VAL PRO PRO ARG LEU PHE GLU Cys SER ASN Lys THR GLY ARG PHE    630

TTG GCC ACT GAG ATC GTC GAC TTC ACC CAG GAT GAC CTG GAC GAG AAC GAC GTT TAC CTG CTG GAC ACT TGG GAC CAG ATT TTC TTC TGG   1980
LEU ALA Thr Glu Ile Val Asp Phe Thr Gln Asp Asp Leu Asp Glu Asn Asp Val Tyr Leu Leu Asp Thr Trp Asp Gln Ile Phe Phe Trp    660

ATC GGG AAA GGT GCC AAT GAG TCA GAG AAG GAG GCG GCA GCA GAG ACA GCA CAG GAG TAC CTG CGG AGT CAC CCC GGC AGC CGT GAC CTC   2070
Ile Gly Lys Gly Ala Asn Glu Ser Glu Lys Glu Ala Ala Ala Glu Thr Ala Gln Glu Tyr Leu Arg Ser His Pro Gly Ser Arg Asp Leu    690

GAC ACC CCC ATC ATC GTG GTG AAG CAG GGC TTT GAA CCC CCC ACC TTC ACT GGC TGG TTC ATG GCC TGG GAT CCC CTC TGC TGG AGC GAC   2160
Asp Thr Pro Ile Ile Val Val Lys Gln Gly Phe Glu Pro Pro Thr Phe Thr Gly Trp Phe Met Ala Trp Asp Pro Leu Cys Trp Ser Asp    720

AGG AAA TCC TAC GAT GAG CTG AAA GCT GAG CTG GGG GAC AAT GCC AGC ATT GGG CAA CTT GTG TCA GGG CTC ACC TCC AAG AAT GAG GTC   2250
Arg Lys Ser Tyr Asp Glu Leu Lys Ala Glu Leu Gly Asp Asn Ala Ser Ile Gly Gln Leu Val Ser Gly Leu Thr Ser Lys Asn Glu VAL    750

TTC ACG GCC ACC ACC ACA CTT GTC CCC ACC AAG CTG GAG ACC TTC CCA CTG GAC GTG CTG GTG AAC ACT GCA GCT GAG GAC CTG CCC CGG   2340
PHE THR ALA THR THR THR LEU VAL PRO THR LYS LEU GLU THR PHE PRO LEU ASP VAL LEU VAL ASN THR ALA ALA GLU ASP LEU PRO ARG    780

GGT GTG GAT CCC AGC AGG AAG GAG AAC CAC CTC TCT GAC GAG GAC TTC AAG GCT GTT TTT GGC ATG ACC CGC TCT GCC TTT GCC AAC TTG   2430
GLY VAL ASP PRO SER ARG LYS GLU ASN HIS LEU SER ASP GLU ASP PHE LYS ALA VAL PHE GLY MET THR ARG SER ALA PHE ALA ASN LEU    810

CCC TTG TGG AAA CAG CAG AAC CTC AAG AAG GAG AAA GGA CTC TTC TAGGAGGGGAGGCTCCAGTACGGCCCCAAGCCCAAATAAACACATTATTTTTACTAAATA   2489
PRO LEU TRP LYS GLN GLN ASN LEU LYS LYS GLU LYS GLY LEU PHE                                                                825

AAATGGATTTGGAATGGAAAAAAAAAAAAAAAAAAAAAAAA-----3'                                                                          2529
```

FIG. 1. Nucleotide sequence of a single cDNA clone isolated from a library constructed from chicken intestine epithelial cells. The deduced amino acid sequence is shown below the nucleotide sequence. Amino acids indicated in uppercase letters were identified by automated Edman degradation of villin proteolytic fragments.

villin and in analogous regions of gelsolin, fragmin, and severin. The structural or functional significance of these sequences will become apparent when the three-dimensional structures of the proteins are determined.

V8 protease, trypsin, and chymotrypsin, which cleave specifically on the carboxyl side of acidic, basic, or large hydrophobic residues, respectively, were restricted from cleavage between domains 2 and 3 in the presence of Ca^{2+}

ONE	29	VP	VPTKSYGNFYEG	DCYVL	LSTRKTGS	GFSYNIHY	W	L	G	KNSSQD	E	QGA	A	77	VILLIN
	52	VP	VPTNLYGDFFTG	DAYVI	LKTVQLRNG	NLQYDLHY	W	L	G	NECSQD	E	SGA	A	101	GELSOLIN
		VP	VPKKHHSSFYTG	DSYIV	LTTYHPKTNPDKLAYDVHF		W	L	G	AFTTQD	E	AGT	A		FRAGMIN
	56	VP	VPESSYGKFYDG	DSYII	LHTFKEGN	SLKHDIHF	F	L	G	TFTTQD	E	AGT	A	104	SEVERIN
TWO	152	VE	MSWKSFNLG	DVFLL	DLGQLIIQ		W	N	G	PESNRA	E	RLR	A	189	
	176	VP	VSWESFNNG	DCFIL	DLGNNIHQ		W	C	G	SNSNRY	E	RLK	A	213	
		VP	KTYKSLNSG	DVFVL	DAGKTVIQ		W	N	G	AKAGLL	E	KVK	A		
	177	VP	LATSSLNSG	DCFLL	DAGLTIYQ		F	N	G	SKSSPQ	E	KNK	A	214	
THREE	272	EI	RPLTQDMLQHE	DCYIL	DQAGLKIFV		W	K	G	KNANKE	E	KQQ	A	310	
	292	NP	FAQGA LKSE	DCFIL	DHGKDGKIFV		W	K	G	KQANTE	E	RKA	A	331	
		RE	LKVKRNLLDSN	DVFVL	YTGAEVFA		W	V	G	KHASVG	E	KKK	A		
	288	GK	INKSS LKSE	DVFII	DLGNEIYT		W	I	G	SKSSPN	E	KKT	A	325	
FOUR	410	VP	VEKRWLGHFYGG	DCYLV	LYTYYVGPKVNRIIYI		W	Q	G	RHASTD	E	LAA	S	458	VILLIN
	431	VP	VDPATYGQFYGG	DSYII	LYNYRHGGRQGQIIYN		W	Q	G	AQSTQD	E	VAA	S	479	GELSOLIN
FIVE	532	VP	VRAASLNSN	DVFVL	KTPSSCYL		W	Y	G	KGCSGD	E	REM	G	569	
	554	VL	PKAGALNSN	DAFVL	KTPSAAYL		W	V	G	TGASEA	E	KTG	A	591	
SIX	636	EI	VDFTQDDLDEN	DVYLL	DTWDQIFF		W	I	G	KGANES	E	KEA	A	673	
	657	VP	GELMQEDLATD	DVMLL	DTWDQVFV		W	V	G	KDSQEE	E	KTE	A	696	

FIG. 2. Homologous sequences in villin, gelsolin, fragmin, and severin. Amino acids are represented by standard one-letter symbols. Sequences from analogous regions of each protein are compared; gaps have been introduced to improve the alignment. Boxes surround residues that are conserved in all four proteins. Numbers indicate the positions of residues in the villin, gelsolin, and severin sequences. The residue positions for fragmin are not depicted, since the complete sequence is not known.

but not in the presence of EGTA. In contrast with the 44T digests, proteolysis of 51T produced identical subfragments in Ca^{2+} and EGTA. We conclude from these results that the Ca^{2+}-dependent differences in proteolysis of 44T probably resulted from a change in 44T conformation and not from a change in protease activity or specificity. The Ca^{2+} dependence provides evidence in addition to the Ca^{2+} dependence of the 44T actin-severing activity that 44T contains the exchangeable-Ca^{2+}-binding site found in villin core (30, 32). Preliminary binding studies revealed that the amino-terminal domain alone displays a Ca^{2+}-dependent actin-monomer-binding activity (P. Janmey and P.M., unpublished data), which indicates that the Ca^{2+}-binding site is located in domain 1.

Sequence comparisons clearly show that villin and gelsolin are homologous with severin and fragmin. The 2-fold differences in molecular weights of villin and gelsolin compared to fragmin and severin are easily explained by a gene duplication sometime during the evolution of vertebrate species (17). This scheme does not explain, however, the presence of the headpiece domain, whose sequence is not in the other domains of villin. The finding that synapsin I (33), an actin-bundling protein purified from synaptic vesicles, has sequences in common with the villin headpiece (34) suggests that the villin headpiece domain might have evolved by some recombination event between the gelsolin gene and a gene encoding a different actin-binding protein. The addition of the headpiece domain then confers at least one of the actin-filament-binding sites necessary for crosslinking actin filaments into bundles. The location of the other bundling site is not known.

Our analysis of the villin sequence also identifies more completely a conserved sequence that divides each half of villin and gelsolin into three homologous regions. Our results establish a one-to-one correlation between the homologous sequences and structural domains and provides strong evidence that actin-severing proteins have similar structures.

We thank Drs. Y. Ben-Neriah, S. Rogelj, N. Simister, C. Brosius, and K. Mostov for help and advice on cDNA library construction and M13 sequencing. Work was supported by the National Institutes of Health, the Whitaker Fund, and the Pew Scholars Program (P.M.) and by a Charles A. King Trust Fellowship (W.L.B.).

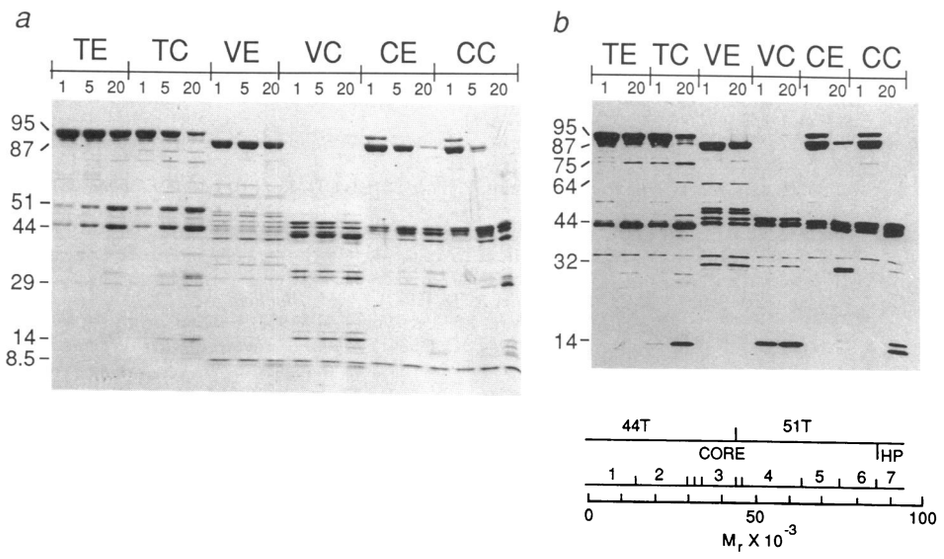

FIG. 3. Map of protease cleavage sites in villin. (*a*) Proteolysis of villin in the presence of 1 mM EGTA or 1 mM Ca^{2+} by trypsin (TE and TC), V8 protease (VE and VC), and chymotrypsin (CE and CC) for 1, 5, and 20 min produced peptides of M_r 87,000, 51,000, 44,000, 30,000, 14,000, and 8500 as detected by Coomassie blue stain. (*b*) Antiserum R211.3, specific for the amino terminus of villin, detected intact villin (M_r 95,000) and fragments of M_r 87,000, 75,000, 64,000, 44,000–50,000, 32,000, and 14,000. The map below the immunoblot shows that cleavage sites are clustered in six regions of the protein; locations of villin core, villin headpiece (HP), 44T, and 51T are indicated.

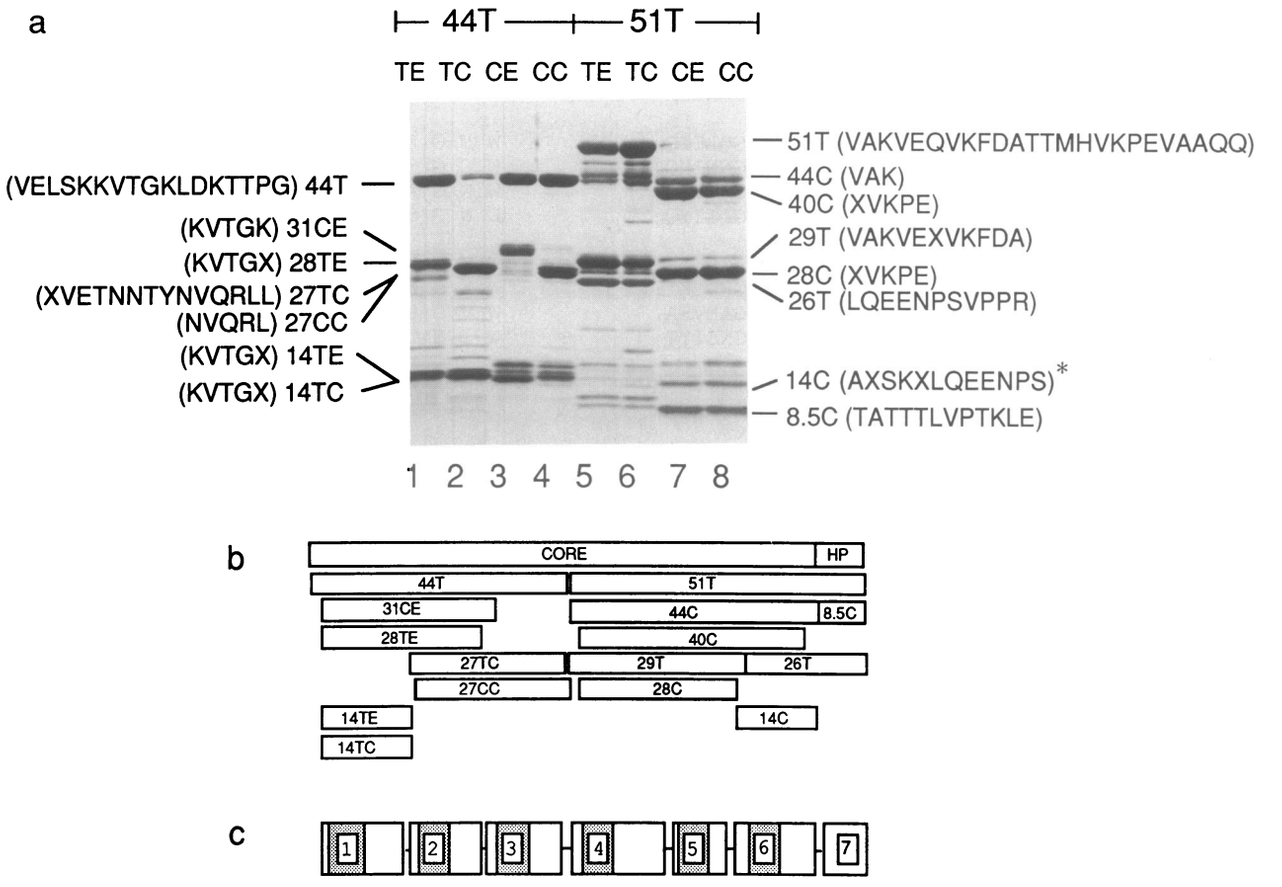

FIG. 4. Map of the protease-resistant domains of villin. (*a*) 44T and 51T (2 μg, 50 pmol) were digested with trypsin or chymotrypsin for 20 min. Digestion was terminated by addition of phenylmethylsulfonyl fluoride. Digests of 44T (lanes 1–4) and 51T (lanes 5–8) generated in 1 mM EGTA (lanes 1, 3, 5, and 7) and 1 mM Ca^{2+} (lanes 2, 4, 6, and 8) by trypsin (lanes 1, 2, 5, and 6) and chymotrypsin (lanes 3, 4, 7, and 8) were electrophoresed in $NaDodSO_4$/10–20% polyacrylamide gradient gels and electroblotted to Immobilon membranes, and the amino-terminal sequences of the fragments were determined by automated Edman methods. The sequences are displayed in one-letter code. Unidentified residues are denoted by X. Asterisk indicates a sequence obtained from a separate blot, derived from a gel on which 100 pmol of material was loaded. (*b*) Diagram shows the location of each fragment based on its amino-terminal sequence and size. (*c*) Villin consists of seven structural domains. The conserved sequences (stippled boxes) lie within the amino-terminal half of the six domains of the villin core.

1. Bretscher, A. & Weber, K. (1980) *Cell* **20,** 839–847.
2. Northrup, J., Weber, A., Mooseker, M. S., Franzini-Armstrong, C., Bishop, M. F., Dubyak, G. R., Tucker, M. & Walsh, T. P. (1986) *J. Biol. Chem.* **261,** 9274–9281.
3. Glenney, J. R. & Weber, K. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2810–2814.
4. Glenney, J. R., Jr., Geisler, N., Kaulfus, P. & Weber, K. (1981) *J. Biol. Chem.* **256,** 8156–8161.
5. Matsudaira, P., Jakes, R. & Walker, J. (1985) *Nature (London)* **315,** 248–250.
6. Matsudaira, P., Jakes, R., Cameron, L. & Atherton, E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6788–6792.
7. Stossel, T. P., Chaponnier, C., Ezzell, R. M., Hartwig, J. A., Janmey, P. A. & Yin, H. L. (1985) *Annu. Rev. Cell Biol.* **1,** 353–402.
8. Pollard, T. D. & Cooper, J. A. (1986) *Annu. Rev. Biochem.* **55,** 987–1035.
9. Kwiatkowski, D. J., Janmey, P. A., Mole, J. E. & Yin, H. L. (1985) *J. Biol. Chem.* **260,** 15232–15238.
10. Bryan, J. & Hwo, S. (1986) *J. Cell Biol.* **102,** 1439–1446.
11. Chaponnier, C., Janmey, P. A. & Yin, H. L. (1986) *J. Cell Biol.* **103,** 1473–1481.
12. Rouayrenc, J.-F., Fattoum, A., Mejean, C. & Kassab, R. (1986) *Biochemistry* **25,** 3859–3867.
13. Hasegawa, T., Takahashi, S., Hayashi, H. & Hatano, S. (1980) *Biochemistry* **19,** 2677–2683.
14. Hinssen, H. (1981) *Eur. J. Cell Biol.* **23,** 225–233.
15. Wang, L. L. & Spudich, J. A. (1984) *J. Cell Biol.* **99,** 844–851.
16. Brown, S. S., Yamamoto, K. & Spudich, J. A. (1982) *J. Cell Biol.* **93,** 205–210.
17. Kwiatkowski, D. J., Stossel, T. P., Orkin, S. H., Mole, J. E., Colten, H. R. & Yin, H. L. (1986) *Nature (London)* **323,** 455–458.
18. Ampe, C. & Vandekerckhove, J. (1987) *EMBO J.* **6,** 4149–4157.
19. Andre, E., Lottspeich, F., Schleicher, M. & Noegel, A. (1988) *J. Biol. Chem.* **263,** 722–738.
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
21. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
22. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning Techniques: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford).
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
25. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
26. Matsudaira, P. T. & Burgess, D. R. (1978) *Anal. Biochem.* **87,** 844–851.
27. Matsudaira, P. (1987) *J. Biol. Chem.* **262,** 10034–10038.
28. Pringault, E., Arpin, M., Garcia, A., Finidori, J. & Louvard, D. (1986) *EMBO J.* **5,** 3119–3124.
29. Burgoyne, R. D. (1987) *Trends Biochem. Sci.* **12,** 85–87.
30. Hesterberg, L. K. & Weber, K. (1983) *J. Biol. Chem.* **258,** 365–369.
31. Creighton, T. E. (1984) *Proteins, Structures and Molecular Principles* (Freeman, New York), p. 316.
32. Hesterberg, L. K. & Weber, K. (1986) *Eur. J. Biochem.* **154,** 135–140.
33. Bahler, M. & Greengard, P. (1987) *Nature (London)* **326,** 704–707.
34. McCafferty, C. A. & DeGennaro, L. J. (1986) *EMBO J.* **5,** 3167–3173.