Sequence of Human Villin: a Large Duplicated Domain Homologous with Other Actin-severing Proteins and a Unique Small Carboxy-terminal Domain Related to Villin Specificity

Monique Arpin, Eric Pringault, Joelle Finidori, Alphonse Garcia, Jean-Marc Jeltsch,* Joël Vandekerckhove,‡ and Daniel Louvard

Institut Pasteur, Département de Biologie Moléculaire, 75724 Paris Cedex 15, France; *U-184 Inserm. Laboratoire de Génétique Moléculaire du CNRS, Institut de Chimie Biologique, 67085 Strasbourg, France; and ‡Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium

Abstract. Villin is a calcium-regulated actin-binding protein that caps, severs, and bundles actin filaments in vitro. This 92,500-D protein is a major constituent of the actin bundles within the microvilli of the brush border surface of intestinal and kidney proximal tubule cells. Villin is a very early marker of cells involved in absorption and its expression is highly increased during intestinal cell differentiation.

The amino acid sequence deduced from the cDNA sequence revealed that human villin is composed of three domains. The first two domains appear as the result of a duplication: their structural organization is similar. We can then define a basic unit in which a slightly hydrophilic motif is followed by three hydrophobic motifs, similar between themselves and regularly spaced. The duplicated domain is highly homologous to three other actin-severing proteins and this basic structure represents the whole molecule in severin and fragmin, while two basic units compose gelsolin.

The third domain which is carboxy terminal is villin specific: it is unique among actin modulating proteins so far known. It could account for its actin-binding properties (dual regulation by calcium of severing and bundling activities). We propose that it may also be related to the subcellular localization of villin in different epithelial cell types.

VILLIN belongs to the large class of actin-modulating proteins regulated by calcium and present in nonmuscle cells. It was first isolated from the brush border of chicken intestinal epithelial cells (Bretscher and Weber, 1979, 1980; Mooseker et al., 1980; Craig and Powell, 1980). Among the proteins that regulate actin filament assembly by performing various functions such as sequestering actin monomers, blocking the end of actin filaments, severing or cross-linking these filaments, villin and gelsolin (Yin and Stossel, 1979) are best characterized in higher eucaryotes. (For review see Stossel et al., 1985.) Villin has two unique features. (*a*) It has a dual function in vitro regarding calcium concentration. At low Ca^{++} concentration, ($<10^{-7}$ M), villin acts as a bundling factor while at concentrations above 10^{-6} M, villin severs, nucleates, and blocks the end of actin filaments (Glenney et al., 1980, 1981*a*, *b*; Bretscher and Weber, 1980; for review see Pollard and Cooper, 1986; Mooseker, 1985). (*b*) Unlike most of the other actin-binding proteins so far characterized, villin presents a striking tissue-specific distribution (Bretscher et al., 1981; Robine et al., 1985). It is a major constituent of the brush border of cells specialized in absorption, namely intestinal and kidney proximal tubule cells. Villin is also present, although at a much lower concentration, at the luminal faces of a few cell types lacking an organized brush border; i.e., epithelial cells lining pancreatic and bile ducts (Robine et al., 1985). These cells are also involved in absorptive processes and share a common embryonic origin with intestinal cells.

We have previously shown that villin is already expressed in immature cells; i.e., embryonic intestinal cells and dividing cells of the adult intestinal mucosa. Villin expression is increased upon terminal differentiation (Robine et al., 1985). Analysis of villin mRNAs expressed in a human adenocarcinoma cell line (HT29) in which differentiation can be manipulated in vitro, indicates that this control occurs at the mRNA level (Pringault et al., 1986). Moreover, recent work has shown that villin can be detected early in embryogenesis of the gut in both chick (Shibayama et al., 1987) and mouse (Maunoury, 1988). Thus, in addition to its tissue-specific expression, villin represents an early marker to follow the differentiation of intestinal epithelial cells during embryogenesis and in the adult.

To better understand the cellular functions in vivo of this protein and to further analyze the regulation of its expression during development and terminal differentiation, we have isolated a full length cDNA clone coding for human villin. We present an analysis of the villin amino acid sequence with regard to the functional domains previously characterized by

biochemical approaches (Glenney and Weber, 1981; Glenney et al., 1981*b*; Matsudaira et al., 1985). Sequence homologies with other actin-binding proteins, which share similar functions and particularly with gelsolin, are discussed.

Materials and Methods

cDNA Library Construction

Total RNA was isolated from a subclone of the HT-29 cell line (HT29-18-C1; Huet et al., 1987) by the guanidium isothiocyanate method (Chirgwin et al., 1979) and enriched for polyA$^+$ RNA by passage over oligo dT cellulose (Aviv and Leder, 1972). The polyA$^+$ RNA was size fractionated on a 5–20% sucrose gradient and the fractions containing villin mRNA were identified by Northern blot analysis using a cRNA probe corresponding to the carboxy-terminal end of human villin (Pringault et al., 1986).

cDNA was synthesized from 1 μg of polyA$^+$ RNA according to the method described by Gubler and Hoffman (1983). After methylation of the cDNA with Eco RI methylase, Eco RI linkers were ligated to the blunt-ended cDNA, digested with Eco RI, and separated on an Ultrogel column (model AcA34; LKB Instruments, Inc., Bromma, Sweden) in TE (20 mM Tris, 1 mM EDTA) buffer. 30 ng of cDNA were then ligated with 1 μg of Eco RI-digested λgt10 vector DNA and the ligated particles were packaged in vitro to generate a cDNA library containing 10^6 independent recombinants.

To isolate the 5′ end of villin cDNA it was necessary to perform three successive primer extensions using oligonucleotide sequences complementary to villin mRNA. These oligonucleotide probes are underlined in Fig. 3. The oligonucleotides were hybridized to 2 μg of size-fractionated polyA$^+$ mRNA in a molar ratio of 10:1. The synthesis of cDNA was then carried out according to the conditions previously described. The λgt10 libraries were screened with cRNA probes generated by in vitro transcription (Melton et al., 1984) from the cDNA clones already characterized.

Northern Blot Analysis

PolyA$^+$ RNA (1.5 μg) isolated from the HT29-18-C1 cell line was fractionated by electrophoresis on 1% agarose gels in the presence of 1 M formaldehyde (Lehrach et al., 1977) and transferred to nitrocellulose. Blots were prehybridized at 55°C for 16 h in 50% formamide, 4× SSC, 0.05 M Na_2HPO4, pH 7.4, 1× Denhardt's solution, 250 μg/ml denatured salmon sperm DNA, and 500 μg/ml tRNA. Hybridization with ^{32}P-labeled RNA probe (2×10^6 cpm/ml) was carried out for 24 h at 55°C in the same solution except that the tRNA was omitted. Blots were washed twice in 1× SSC, 0.1% SDS, once in 0.2× SSC, 0.1% SDS at 65°C, and finally once in 0.1× SSC, 0.1% SDS for 30 min at 70°C.

Sequence Analysis

Restriction fragments of cDNA were subcloned in M13mp18-mp19 derivatives and sequenced by the dideoxy-chain termination procedure described by Sanger et al. (1977). Overlapping sequence data were obtained for both strands.

Protein Sequence Analysis

Human villin was purified by B. West, L. West, and M. Mooseker (Department of Biology, Yale University) from isolated human brush borders (Carboni et al., 1987) by methods described by Coleman and Mooseker (1985). Before sequencing, 25 μg of villin were subjected to SDS-PAGE on a 10% gel and electroblotted onto a glass-fiber membrane (model GF/C; Whatman Inc., Clifton, NJ) coated with poly(4-vinyl-N-methylpyridine). For details see Bauw et al., (1987). The immobilized protein was detected by a dilute fluorescamine stain (1 mg/liter of acetone), excised from the membrane, and mounted in the reaction chamber of a gas-phase protein sequenator (model 470 A; Applied Biosystems, Inc., Foster City, CA) operated following the instructions of the manufacturer. Initial sequencing yields, calculated as a percentage of protein loaded on the gel, are ~10%. It is known that protein transfer is seldom quantitative (usually between 60 and 90%) and that artifactual NH_2-terminal blocking occurs during gel electrophoresis (in many cases, >50% of the protein is blocked; see also Moos et al., 1988). We therefore assume that this value is not abnormally low and is probably reflecting the NH_2-terminal sequence of the majority of the villin molecules rather than being the result of a proteolytic cleavage close to the NH_2 terminus of a fraction of the molecules of villin, which would be otherwise NH_2-terminally blocked.

Computer Analysis of the Amino Acid Sequences

A hydropathy profile for human villin was determined using the program devised by Kyte and Doolittle (1982). Amino acid sequence comparisons were carried out according to the program described by Wilbur and Lipman (1983). The protein libraries from both the National Biochemical Research Foundation (Washington DC) and of Newat were searched for homologous protein sequences.

Results

Isolation of Human Villin cDNA Clones

We previously reported the characterization of a cDNA clone (515 bp) encoding the carboxy-terminal end of human villin (330 bp) and a 3′ untranslated region (185 bp). Northern blot analysis carried out with this cDNA probe revealed that in human intestine and kidney, two different sizes of mRNA hybridized with this probe whereas in the same tissues from rat and chicken only one mRNA species was observed (Pringault et al., 1986).

A λgt10 library constructed with size-fractionated polyA$^+$ RNA was screened with this probe. Four positive clones were isolated from ~5 × 10^5 recombinants. The two largest clones with insertions of 1.75 (V1) and 1.33 kb (V2) were characterized by restriction map analysis and by sequencing (Fig. 1). V1 contained 1.6 kb of coding sequence and 185 bp of the noncoding sequence present in the probe. V2 contained 330 bp of coding sequence identical to that of the probe and a noncoding region of 1 kb.

Furthermore, Northern blot analysis was carried out with cRNA probes corresponding to (*a*) the cDNA clone previously described (Pringault et al., 1986) and used for screening (Fig. 2 *A*), (*b*) the complete V2 clone (Fig. 2 *B*), and (*c*) the 3′ end of the V2 clone (180 bp; Fig. 2 *C*). The cRNA probe corresponding to the complete V2 clone hybridized with both mRNA species (Fig. 2 *B*), while the cRNA probe of 180 bp hybridized only with the larger mRNA (Fig. 2 *C*). Altogether, these results demonstrated that human villin is encoded by two different sizes of mRNA (~2.8 and 3.6 kb), and that the two bands observed by Northern blot analysis are due to a difference in the 3′ noncoding region.

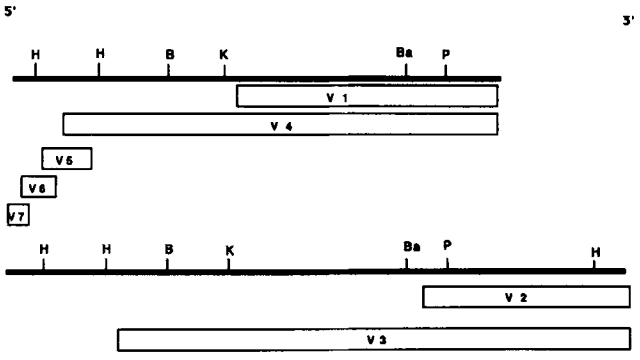

Figure 1. Cloning strategy for human villin cDNA. The dark lines represent the smaller (*upper line*) and larger (*lower line*) villin mRNAs. A partial restriction map is indicated: *Ba*, Bam HI; *B*, Bgl_2; *H*, Hind III; *P*, Pvu II. The open boxes correspond to the partial villin cDNA clones isolated from cDNA libraries.

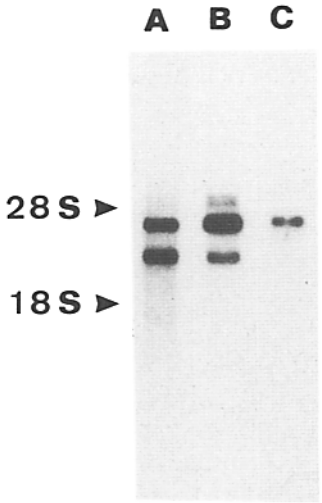

Figure 2. Hybridization analysis of polyA$^+$ RNA from the HT29-18-C_1 cell line. 1.5 μg of polyA$^+$ RNA was fractionated on 1% agarose/formaldehyde gel, transferred to nitrocellulose, and hybridized with the ^{32}P-labeled RNA probes. Lane *A*, cRNA probe corresponding to the cDNA clone encoding the 3′ end of the smaller mRNA (515 bp). Lane *B*, cRNA probe corresponding to the cDNA clone V2 (1.33 kb). Lane *C*, cRNA probe corresponding to the 3′ end of the cDNA clone V2 (180 bp).

To isolate longer villin cDNAs, a cRNA probe (350 bases in length) corresponding to the 5′ end of V1 was generated and used for screening 5×10^5 recombinants. Two additional villin cDNA clones were isolated by hybridization: the largest one with an insertion of 3 kb (V3) corresponded to the mRNA with the longer 3′ noncoding region. The other cDNA with an insertion of 2.3 kb (V4) corresponded to the smaller mRNA (Fig. 1).

The small number of villin cDNAs isolated from our library, given the relative abundance of villin mRNA in the HT29-18-C1 cell line, together with the lack of larger cDNA clones suggested to us that secondary structures present in the villin mRNA might be interfering with the synthesis of villin cDNA. To overcome this problem and to obtain cDNA encoding the 5′ sequence of the villin mRNA, we used an oligonucleotide (19 mers) complementary to nucleotides 546–564 of villin mRNA as a primer for construction of another cDNA library. A cRNA probe (180 bp), derived from the cDNA clones already characterized and located upstream of the primer, was used to screen the library. The largest clone isolated contained 470 nucleotides (V5). This clone did not encode the complete 5′ region of villin mRNA, and to obtain this region we had to repeat this cloning strategy twice. We constructed two other libraries using a 13-mer and then a 20-mer oligonucleotide complementary to the nucleotides 174–186 and 79–98, respectively, of the villin mRNA. cRNA probes located upstream of the primers and, respectively, 60 and 20 bp long were used to screen the libraries and isolate the clones V6 and V7 (Fig. 1). According to the amino-terminal sequence derived from protein sequence (12 amino acids) of human villin (Fig. 3), the V7 clone encoded the 5′ coding region of villin terminating at the second nucleotide of the codon after the initiation codon ATG. The sequence of the 5′ untranslated region was obtained from fragment isolated from a genomic library (generous gift from H. Lehrach; Frischauf et al., 1983), which also contained part of the 5′ coding sequence overlapping with the cDNA clones V6 and V7.

Nucleotide and Deduced Amino Acid Sequence of Human Villin

The complete nucleotide sequence of the overlapping cDNAs and the deduced amino acid sequence of human villin are shown in Fig. 3. Comparison of this primary structure with the NH_2-terminal sequence of the human villin determined by protein sequencing indicated that the ATG found at position 25 corresponded to the initiation codon and that the methionine is removed in the mature protein. Moreover, the ATG codon is preceded by a sequence (CACC) that is in good agreement with the consensus sequence defined for the initiation of translation of most eukaryotic mRNAs (Kozak, 1986). The first ATG codon is followed by an open reading frame of 2,478 nucleotides. This open reading frame encodes a polypeptide of 826 amino acids that gives a calculated molecular mass of 92,459 D and calculated pH of 5.94 in agreement with the pH measured by two-dimensional gel electrophoresis (Sahuquillo-Mérino, C., unpublished data). The 3′ untranslated region of the shorter cDNA is 185 bp long with a polyadenylation site located 30 bp upstream of the poly A region. The cDNA coding for the larger mRNA contains a 3′ untranslated region ~1 kb long with two potential polyadenylation sites: one corresponding to that of the smaller mRNA and the second one located 18 nucleotides upstream of the poly A stretch in the larger cDNA (data not shown). Automated gas-phase sequencing revealed the presence of two residues ALA/THR (ratio 2:1) in the first position, while in the further twelve positions unambiguous sequences were found in agreement with the proposed cDNA sequence. While the assignment of THR as the amino-terminal residue fits with the DNA sequence, it cannot be excluded that the additional amino-terminal ALA may reflect the presence of a villin isoform possibly derived by allelic microheterogeneity. On the other hand, it is also very well known that ambiguities, particularly arising in the first step of the Edman degradation, may often be the result of simple contamination by free amino acids.

Computer analysis of the human villin sequence using the program of Wilbur and Lipman (1983) indicated that this protein is composed of three large domains. The first two domains are repeated in tandem with a single gap of 17 residues. The third domain corresponding to the "head-piece" region of villin (Glenney and Weber, 1981) is distinct. The two repeated domains cover residues 1–360 and residues 380–720. Within each of them, we can distinguish two groups of different motifs. The first one is repeated once (a, a′) while the second one is repeated three times (bb′, cc′, dd′). The hydropathy plot shows that the motifs aa′ display a slightly hydrophilic profile while the three repeated sequences (bb′, cc′, dd′) correspond to rather hydrophobic areas of the molecule (Fig. 4). Sequence comparison of the motifs a and a′ (residues 16–49 and 397–430) shows that they are highly homologous since 16 amino acids out of 33 are strictly identical (see Fig. 6 *A*). The three identical groups of 17 residues (bb′, cc′, dd′), which are repeated in each domain, also display, between themselves, a strong homology but appear different from the first group aa′. Alignment of these triplicate sequences shows a characteristic motif of highly conserved amino acids (I/L – W_{+3} – G_{+6} – S/T/N_{+10} – E_{+13} – A_{+17}) separated by a constant number of amino acids that are often conservative (see Fig. 6 *B*). These repeated motifs are separated by regular intervals of 93–106 residues along the molecule and confer to the primary structure of human villin several elements of symmetry.

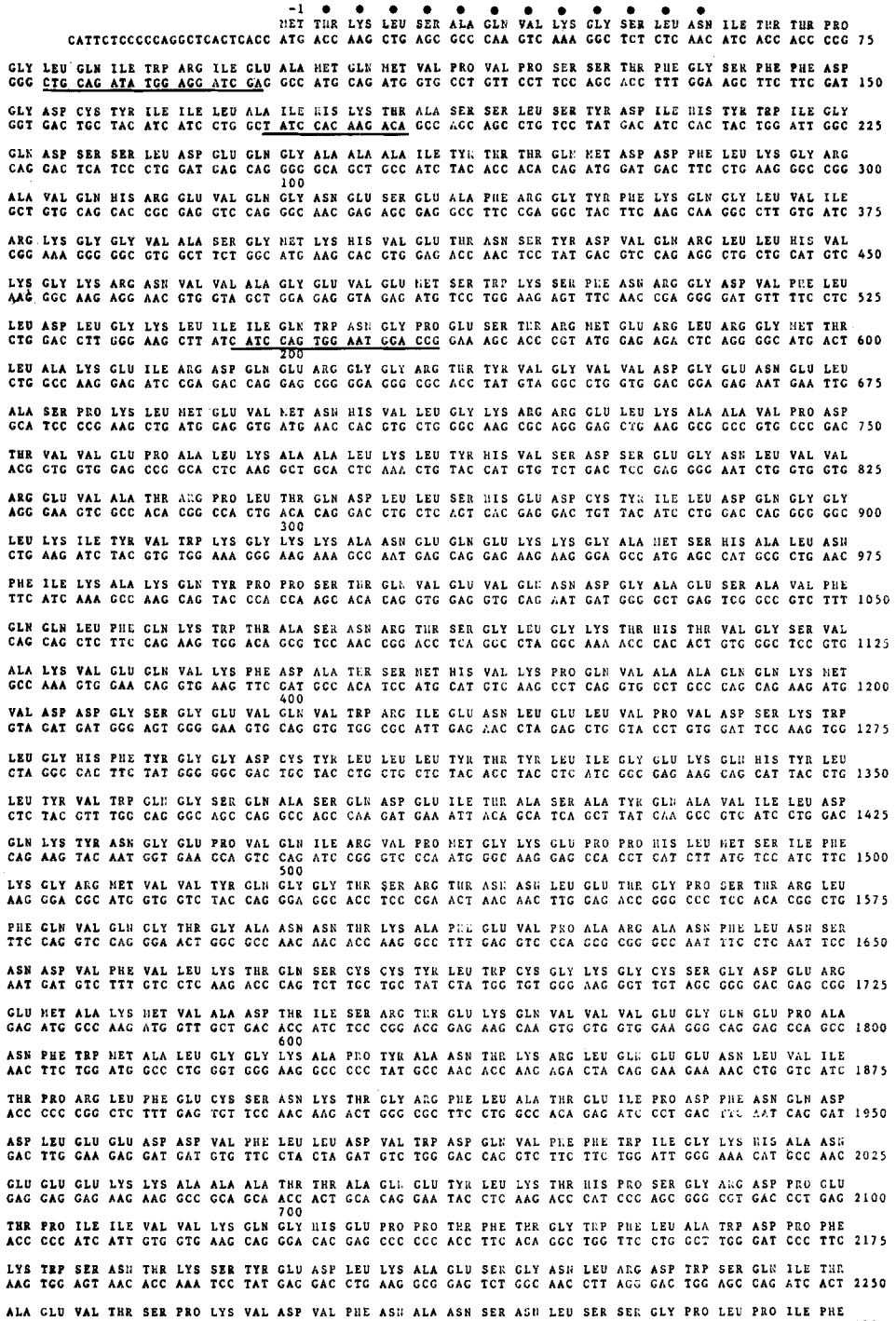

Figure 3. Nucleotide and predicted amino acid sequence of human villin cDNA. Numbers above the lines refer to amino acid position, numbering starting at position THR 1 of the NH_2-terminal sequence found in mature human villin. Numbers at the end of each line refer to nucleotide position. Residues with an asterisk were also obtained by amino-terminal sequence analysis. The oligonucleotides (546–564, 174–187, 79–99) used as primers for cDNA synthesis and the polyadenylation site are underlined. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number X12901.

Structural analysis of human villin shows the presence of a third domain in the molecule. This domain codes for the carboxy-terminal part of villin and corresponds to the chicken villin headpiece isolated and characterized by Glenney et al. (1981*a*). Interestingly, no significant homology is found between the headpiece region and the rest of the molecule nor with other proteins listed in the protein data banks.

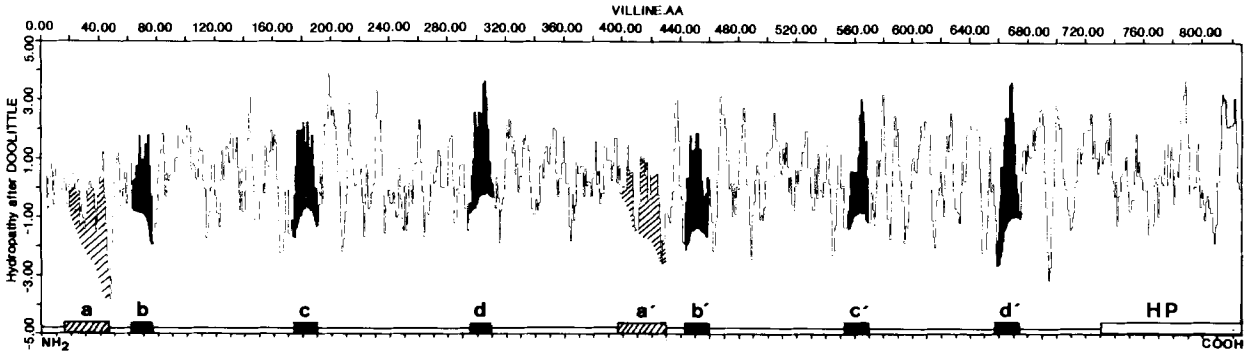

Figure 4. Hydropathy profile of human villin. The plot was determined using a window of five residues (Kyte and Doolittle, 1982). Motifs *a* and *a'* which display homologous sequences are represented by dashed lines. Repeated motifs (*bb'*, *cc'*, *dd'*) are represented by a dark area. *HP* corresponds to the head piece region.

Homology between Human Villin and Human Gelsolin

The sequence of human villin was compared at the sequence of human gelsolin determined by Kwiatkowski et al. (1986). Gelsolin is a Ca^{2+}-regulated actin-binding protein that has been originally purified from rabbit macrophages (Yin and Stossel, 1979) and whose function is to sever, nucleate, and block actin filaments. Unlike villin, however, gelsolin has no bundling activity. Comparison of the primary structure of villin and gelsolin shows a very high degree of homology (57%) in the sequences covering the two duplicated domains with only 6 short gaps <7 residues (Fig. 5). The same internal motifs (aa', bb', cc', dd') are found in each domain of both proteins and inside each repeat identical characteristic sequences are present (Fig. 6, *A* and *B*). The overall structure of gelsolin (Kwiatkowski et al., 1986) is restricted to the two large domains repeated in tandem present in the villin molecule. Thus the structural specificity of villin resides in the carboxy-terminal part of the molecule which appears as an additional domain present only in villin. Since only villin displays a bundling activity, it is tempting to correlate the presence of this headpiece to the specific function of villin as was already shown by Glenney et al. (1981*a*).

Villin Shares Sequence Homology with Other Actin-binding Proteins

The primary structure of other actin-binding proteins showing similar F-actin severing and capping activities have been recently reported (Kwiatkowski et al., 1986; Ampe and Vandekerckhove, 1987; André et al., 1988). On the basis of their molecular mass they can be divided into two groups: villin (92,500) and gelsolin (90,000) have been found only in higher eucaryotes while two proteins with similar binding activity but half-size molecular mass have been isolated from lower eucaryotes: fragmin (cap 42a; 42 kD) purified from *Physarum polycephalum* (Hasegawa et al., 1980; Hinssen, 1981) and severin (40 kD) from *Dictyostelium discoideum* (Brown et al., 1982). Fig. 6 *A* compares the motifs a and a' of villin and gelsolin with the amino-terminal sequence of severin and fragmin. A large motif with a strict homology (W_6 - R_7 - V_{14} - P_{15} - V_{16} - G_{22} - F_{24} - G_{27} - D_{28} - Y_{30} - L_{33}) is observed in these four proteins (Fig. 6 *A*). Moreover Fig. 6 *B* indicates that the three motifs bb', cc', dd' found in villin and gelsolin are also present as highly conserved sequences in severin and fragmin (I/L/V - W_{+4} - G_{+6} - S/T/N/Q_{+10} - E_{+13} - A_{+17}). It should be noted that fragmin and severin constitute only half of the villin-gelsolin molecules and therefore contain only the domains a, b, c, d.

Discussion

To obtain the complete protein sequence of human villin and for future studies on the villin gene, our aim was to isolate a complete cDNA sequence derived from the villin mRNA. Since our library did not contain a full length cDNA coding for human villin three successive primer extensions were performed along the 5' end of villin mRNA. Analysis a posteriori of the secondary structures of villin mRNA using the program devised by Zuker and Stiegler (1981) revealed that several hairpin loops are indeed present in this mRNA which may have prevented the synthesis of full length cDNA by the reverse transcriptase.

Two large villin cDNA clones were isolated corresponding to the two sizes of human villin mRNAs. The nucleotide sequence of the cDNA encoding the entire length of the smaller mRNA was determined and the amino acid sequence deduced. No differences in the coding region were found between this cDNA and a partial villin cDNA clone coding for the larger mRNA and encompassing the amino acid residues 200–826. It is likely that these two mRNAs encode the same protein although a small difference in the amino acid sequence at the amino terminus of villin cannot be excluded.

The presence of two mRNAs coding for villin is a feature characteristic of the human species. So far, no differences in the ratio of these two mRNAs could be detected in tissues expressing villin or upon differentiation of intestinal cells in culture (Pringault et al., 1986).

Sequence comparison with other actin-binding proteins isolated from various species indicates that the same structural organization is present in these proteins. Indeed, fragmin and severin isolated from lower eucaryotes, and villin and gelsolin purified from vertebrates, all contain one related domain with four motifs, three of which are homologous. In villin and gelsolin however, this domain is duplicated. The structural organization reported here allows us to define a basic unit composed of four motifs. This observation suggests that these proteins have evolved from a common ancestor gene by duplication in gelsolin and by duplication and addition of a specific domain in villin (Fig. 7).

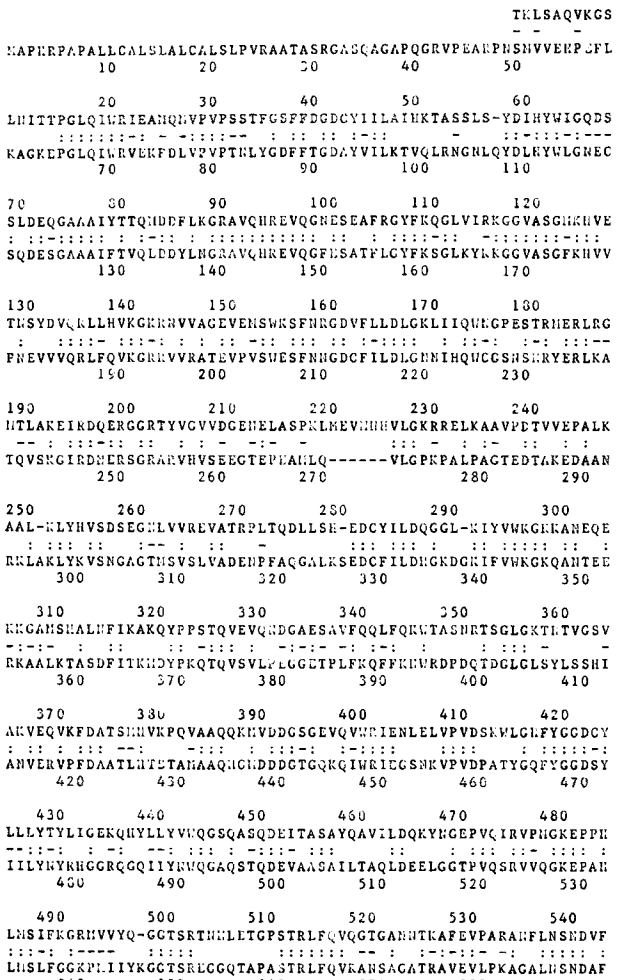

Figure 5. Comparison of the predicted amino acid sequence of human villin and human plasma gelsolin. Villin amino acid residues (*upper line*) were aligned with human plasma gelsolin amino acid residues (*lower line*) using the program devised by Wilbur and Lipman (1983). The following parameters were defined to maximize the match: K-tuple size (2); window size (20); gap penalty (3). Identical residues. Conservative residues. The conservative replacements have been defined according to the following amino acid grouping. 1: R, K. 2: L, F, P, M, V, I. 3: S, T, Q, N, C. 4: A, G, W. 5: H. 6: E, D. 7: Y.

This overall organization may have important implications for understanding how these proteins regulate the actin filament assembly. Proteolytic cleavages of the proteins have been used to localize the actin- and calcium-binding sites and more precisely to identify their different activities. Indeed, proteolysis of intact chicken villin with V8 protease generates two fragments: a 8.5-kD fragment described as the "head piece" located at the carboxy-terminal end of villin and the core (85 kD) which retains the Ca^{2+}-dependent actin blocking, severing, and nucleating activity but loses the bundling function (Glenney et al., 1981*a*: Glenney and Weber, 1981). Moreover, trypsin cleavage of chicken villin generates two fragments 44T and 51T (Matsudaira et al., 1985). The Ca^{++}-regulated actin-binding site was assigned to the amino-terminal fragment (44T). The severing activity of villin and gelsolin has been localized in the amino-terminal part of these proteins (Glenney et al., 1981*a*, *b*; Matsudaira et al., 1985; Kwiatkowski et al., 1985; Yin et al., 1988), while the nucleating and capping activities have not been precisely localized. It has been suggested that the actin-binding sites correspond to regions that present sequence homology between villin and gelsolin (Matsudaira et al., 1985; Kwiatkowski et al., 1986). In this respect, it is important to note that, although these proteins have similar functions such as the severing activity, these functions are not identical. Indeed the binding of villin to G-actin and actin filaments is different from gelsolin, fragmin, and severin since it can be completely and rapidly reversed by EGTA (Bretscher and Weber, 1980; Walsh et al., 1984*a*, *b*; for review see Mooseker, 1985). Moreover, if the two domains that compose the "core" of the protein have a symmetrical structure, the attempts made to localize the activities of this protein indicate that the two domains are not functionally identical. One alternative is that these homologous domains play a role in the architecture of the molecule and the functional sites may or may not lie in these regions of the molecule.

Finally, if these proteins have in common the calcium-regulated control of actin organization they probably perform different physiological functions. Important features differentiate villin from gelsolin with which it shares extensive homology. Villin has a bundling activity that can be partially assigned to a domain, the head piece, present only in this protein. This fragment (8.5 kD) together with the core are able to bundle actin filaments at a Ca^{2+} concentration $<10^{-6}$ M (Glenney and Weber, 1981; Glenney et al., 1981*a*). This raises the question of the functional significance of this domain. One possibility is that the head piece can associate only with actin filaments that have a specific organization. For instance, it is conceivable that villin specifically associates with the ordered actin microfilaments with a single polarity such as those found in the intestinal microvilli of the brush border. This hypothesis would account for the striking tissue-specific expression of this protein and its subcellular localization. Villin is mainly localized at the apical pole of a few epithelial cell types whether these cells display an organized brush border such as intestinal cells or do not as is the case for pancreatic and liver duct cells. The availability of a complete villin cDNA clone should allow us to test this hypothesis by injecting, for instance, appropriately modified cDNA sequences in an expression vector into polarized and nonpolarized epithelial cells.

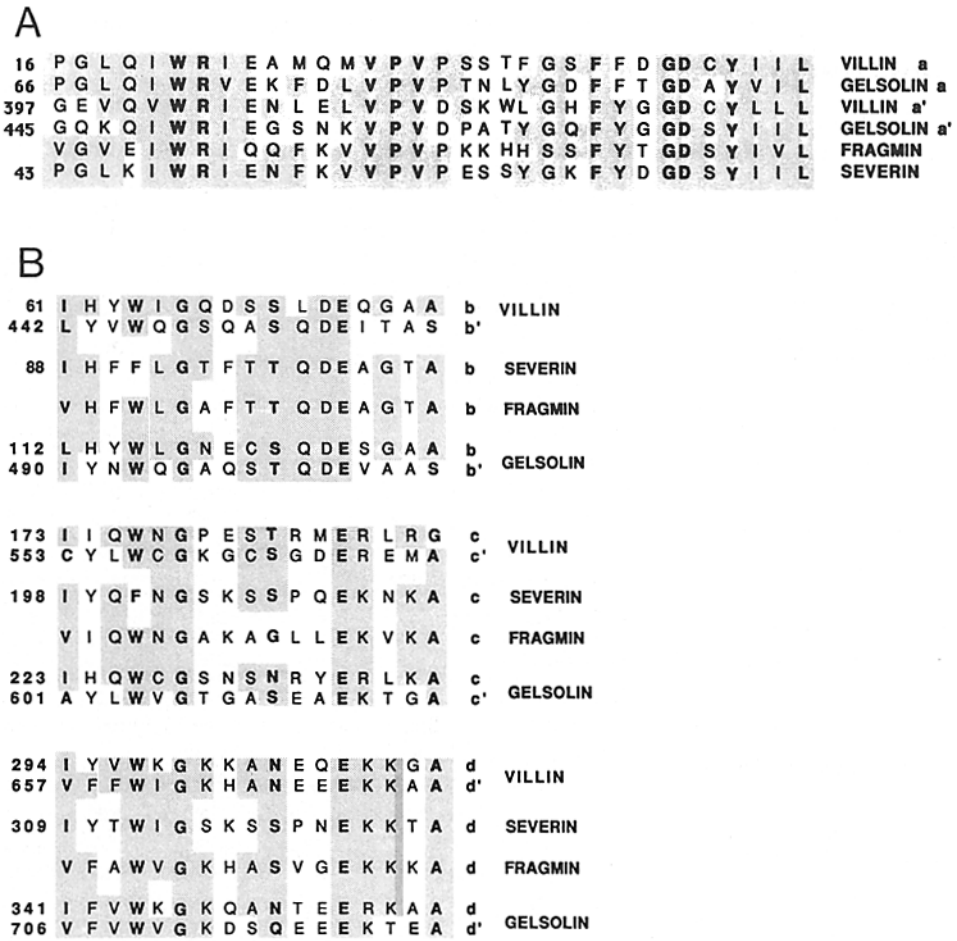

Figure 6. Sequence comparison between villin, gelsolin, fragmin, and severin. (*A*) Amino acid sequences of the regions *a* and *a'* of villin were compared with homologous domains of gelsolin (Kwiatkowski et al., 1986), fragmin (Ampe and Vandekerckhove, 1987), and severin (André et al., 1988). (*B*) The three repeated motifs of villin (*bb'*, *cc'*, *dd'*) are aligned with homologous motifs of gelsolin, fragmin, and severin. The common amino acids in the motifs are shown in bold type while conservative residues are shown in lightly shaded boxes.

We wish to thank Dr. Mark Mooseker's laboratory (Yale University, New Haven) for providing us with purified human villin. We thank Annie Lamouroux (Gif-sur Yvette) for helpful suggestions in constructing cDNA libraries; Thierry Lerouge (Strasbourg) for his help in cDNA sequencing; and Christian Sahuquillo-Mérino for the cell culture. We are grateful to Dr. P. Chambon (Faculté de Médecine, Strasbourg) for accepting one of us (M. Arpin) in his laboratory. We thank Dr. M. Buckingham (Institut Pasteur) and Dr. M. Mooseker (Yale University) for critically reading the manuscript.

This work was supported by grants from the Institut National de la Santé et de la Recherche Médicale (n° 86-7008), the Fondation pour la Recherche Médicale, the Association pour la Recherche sur le Cancer (n° 6379), the CNRS UA-1149, and the Ligue Nationale Française contre le Cancer. Dr. J. Vandekerckhove is a Research Associate of the Belgian National Fund for Scientific Research.

Received for publication 3 May 1988, and in revised form 7 July 1988.

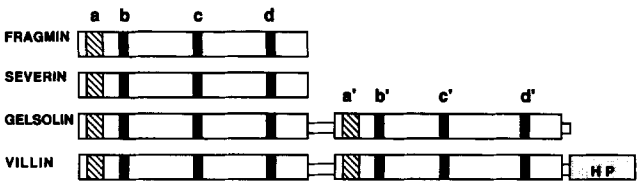

Figure 7. Schematic representation of the structural organization of four actin-binding proteins. In each large domain, dashed areas (*a* and *a'*) represent two homologous motifs while the dark areas correspond to the three motifs identical with each other and repeated. The carboxy-terminal end of villin which is unique among these proteins is indicated by *HP* (head piece).

References

Ampe, C., and J. Vandekerckhove. 1987. The F-actin capping proteins of *Physarum polycephalum*: cap 42a is very similar, if not identical, to fragmin and is structurally and functionally very homologous to gelsolin; cap 42(b) is *Physarum* actin. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:4149–4157.

André, E., F. Lottspeich, M. Schleicher, and A. Noegel. 1988. Severin, gelsolin, and villin share a homologous sequence in regions presumed to contain F-actin severing domains. *J. Biol. Chem.* 263:722–727.

Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

Bauw, G., M. De Loose, D. Inzé, M. Van Montagu, and J. Vandekerckhove. 1987. Alterations in the phenotype of plant cells studied by NH_2-terminal amino acid-sequence analysis of proteins electroblotted from two-dimensional gel-separated total extracts. *Proc. Natl. Acad. Sci. USA.* 84:4806–4810.

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

Bretscher, A., and K. Weber. 1980. Villin is a major protein of the microvillus cytoskeleton which binds both G and F actin in a calcium-dependent manner. *Cell.* 20:839–847.

Bretscher, A., M. Osborn, J. Wehland, and K. Weber. 1981. Villin associates with specific microfilamentous structures as seen by immunofluorescence microscopy on tissue sections and cells microinjected with villin. *Exp. Cell Res.* 135:213–219.

Brown, S. S., K. Yamamoto, and J. A. Spudish. 1982. A 40,000-Dalton protein

from *Dictyostelium discoideum* affects assembly properties of actin in a Ca^{2+}-dependent manner. *J. Cell Biol.* 93:205–210.
Carboni, J. M., C. L. Howe, A. B. West, K. W. Barwick, M. S. Mooseker, and J. S. Morrow. 1987. Characterization of intestinal brush border cytoskeletal proteins of normal and neoplastic human epithelial cells: a comparison with the avian brush border. *Am. J. Pathol.* 129:589–600.
Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5299.
Coleman, T. R., and M. S. Mooseker. 1985. Effects of actin filament cross-linking and filament length on actin-myosin interaction. *J. Cell Biol.* 101: 1850–1857.
Craig, S. W., and L. D. Powell. 1980. Regulation of actin polymerization by villin, a 95,000 dalton cytoskeletal component of intestinal brush borders. *Cell.* 22:739–746.
Frischauf, A.-M., H. Lehrach, A. M. Poustka, and N. Murray. 1983. Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170:827–842.
Glenney, Jr., J. R., and K. Weber. 1981. Calcium control of microfilaments: uncoupling of the F-actin-severing and -bundling activity of villin by limited proteolysis *in vitro*. *Proc. Natl. Acad. Sci. USA.* 78:2810–2814.
Glenney, Jr., J. R., A. Bretscher and K. Weber. 1980. Calcium control of the intestinal microvillus cytoskeleton: its implications for the regulation of microfilament organizations. *Proc. Natl. Acad. Sci. USA.* 77:6458–6462.
Glenney, Jr., J. R., N. Geisler, P. Kaulfus, and K. Weber. 1981*a*. 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.
Glenney, J. R., P. Kaulfus, and K. Weber. 1981*b*. F-actin assembly modulated by villin: Ca^{++}-dependent nucleation and capping of the barbed end. *Cell.* 24:471–480.
Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene.* 25:263–269.
Hasegawa, T., S. Takahashi, H. Hayashi, and S. Hatano. 1980. Fragmin: a calcium ion sensitive regulatory factor on the formation of actin filaments. *Biochemistry.* 19:2677–2683.
Hinssen, H. 1981. An actin-modulating protein from *Physarum polycephalum.* I. Isolation and purification. *Eur. J. Cell Biol.* 23:225–233.
Huet C., C. Sahuquillo-Mérino, E. Coudrier, and D. Louvard. 1987. Absorptive and mucus-secreting subclones isolated from a multipotent intestinal cell line (HT29) provide new models for cell polarity and terminal differentiation. *J. Cell Biol.* 105:345–357.
Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283–292.
Kwiatkowski, D. J., P. A. Jammey, J. E. Mole, and H. L. Yin. 1985. Isolation and properties of two actin-binding domains in gelsolin. *J. Biol. Chem.* 260: 15232–15238.
Kwiatkowski, D. J., T. P. Stossel, S. H. Orkin, J. E. Mole, H. R. Colten, and H. L. Yin. 1986. Plasma and cytoplasmic gelsolins are encoded by a single gene and contain a duplicated actin-binding domain. *Nature (Lond.).* 323: 455–458.
Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.
Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker. 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry.* 16:4743–4751.
Matsudaira, P., R. Jakes, and J. E. Walker. 1985. A gelsolin-like Ca^{++}-dependent actin-binding domain in villin. *Nature (Lond.).* 315:248–250.
Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acis Res. 12:7035–7056.
Moos, Jr., M., N. Y. Nguyen, and T.-Y. Liu. 1988. Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. *J. Biol. Chem.* 263:6005–6008.
Mooseker, M. S. 1985. Organization, chemistry, and assembly of the cytoskeletal apparatus of the intestinal brush border. *Annu. Rev. Cell Biol.* 1: 209–241.
Maunoury, R., S. Robine, E. Pringault, C. Huet, J. L. Guénet, J. A. Gaillard, and D. Louvard. 1988. Villin expression in the visceral endoderm and in the gut anlage during early mouse embryogenesis. *EMBO (Eur. Mol. Biol. Organ.) J.* In press.
Mooseker, M. S., T. A. Graves, K. A. Wharton, N. Falco, and C. L. Howe. 1980. Regulation of microvillus structure: calcium-dependent solation and cross-linking of actin filaments in the microvilli of intestinal epithelial cells. *J. Cell Biol.* 87:809–822.
Pollard, T. D., and J. A. Cooper. 1986. Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. *Annu. Rev. Biochem.* 55: 987–1035.
Pringault, E., M. Arpin, A. Garcia, J. Finidori, and D. Louvard. 1986. A human villin cDNA clone to investigate the differentiation of intestinal and kidney cells *in vivo* and in culture. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:3119–3124.
Robine, S., C. Huet, R. Moll, C. Sahuquillo-Mérino, E. Coudrier, A. Zweibaum, and D. Louvard. 1985. Can villin be used to identify malignant and undifferentiated normal digestive epithelial cells? *Proc. Natl. Acad. Sci. USA.* 82:8488–8492.
Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
Shibayama, T., J. M. Carboni, and M. S. Mooseker. 1987. Assembly of the intestinal brush border: appearance and redistribution of microvillar core proteins in developing chick enterocytes. *J. Cell Biol.* 105:335–344.
Stossel, T. P., C. Chaponnier, R. M. Ezzell, J. H. Hartwig, P. A. Janmey, D. J. Kwiatkowski, S. E. Lind, D. B. Smith, F. S. Southwick, H. L. Yin, and K. S. Zaner. 1985. Nonmuscle actin-binding proteins. *Annu. Rev. Cell Biol.* 1:353–402.
Walsh, T. P., A. Weber, K. Davis, E. Bonder, and M. S. Mooseker. 1984*a*. Calcium dependence of villin-induced actin depolymerization. *Biochemistry.* 23:6099–6102.
Walsh, T. P., A. Weber, J. Higgins, E. M. Bonder, and M. S. Mooseker. 1984*b*. Effect of villin on the kinetics of actin polymerization. *Biochemistry.* 23:2613–2621.
Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA.* 80:726–730.
Yin, H. L., and T. P. Stossel. 1979. Control of cytoplasmic actin gel-sol transformation by gelsolin, a calcium-dependent regulatory protein. *Nature (Lond.).* 281:583–586.
Yin, H. L., K. Lida, and P. A. Janmey. 1988. Identification of a polyphosphoinositide-modulated domain in gelsolin which binds to the sides of actin filaments. *J. Cell Biol.* 106:805–812.
Zuker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9:133–148.