Volume 328, number 1,2, 71–76 FEBS 12798 August 1993
© 1993 Federation of European Biochemical Societies 00145793/93/$6.00

The 100 kDa F-actin capping protein of *Dictyostelium* amoebae is a villin prototype ('protovillin')

A. Hofmann[a], A.A. Noegel[b], L. Bomblies[a], F. Lottspeich[b] and M. Schleicher[a]

[a]*Institute for Cell Biology, Ludwig-Maximilians-University Munich, Schillerstraße 42, 80336 München, Germany* and
[b]*Max-Planck-Institute for Biochemistry, 82152 Martinsried, Germany*

Received 11 June 1993

The 100 kDa actin-binding protein from *Dictyostelium* amoebae is an F-actin capping protein that displays neither severing nor crosslinking nor nucleating activities [Hofmann et al. (1992) Cell Motil. Cytoskel. 23,133–144]. Cloning and sequencing of the gene revealed that the protein is highly homologous to vertebrate villin, a unique component of brush border microvilli and contains six domains fused to a villin-like headpiece domain via a threonine/proline rich neck region. The functional differences and similarities between the 100 kDa protein and villin are reflected in the amino acid sequences. We draw from the data the following conclusions. (i) The presence of a six domain protein in *Dictyostelium* suggests that in contrast to the current view gene duplications must have happened before *Dictyostelium* branched off during evolution. (ii) The villin-like molecule in *Dictyostelium* appears to be a premature villin ('protovillin') which is able to cap actin filaments but still lacks the other villin-type actin-binding activities. This renders capping of actin filaments as the evolutionarily oldest function of an F-actin binding protein.

Actin-binding protein; Cytoskeleton, Evolution; Villin; *Dictyostelium discoideum*

1. INTRODUCTION

Capping proteins form a major group of actin-binding proteins (for reviews see [1,2]) and are characterized by their ability to regulate the length of actin filaments by binding to their barbed ends. In a recent review article [3] the capping proteins have been subdivided into two classes. Class I includes capping proteins that exhibit an additional severing activity. Well-characterized members of this class are severin, gelsolin or villin. Severin (40 kDa) as a prototype of these proteins [4] consists of three homologous domains. The capping activity is localized in the first domain whereas domains two and three carry F-actin binding sites which in combination with the capping activity enable severin to fragment filaments and to promote actin polymerization [5,6]. Gelsolin (83 kDa) which is very similar to severin [7] with respect to its interaction with actin consists of six domains that are thought to have evolved from a severin-like precursor via gene-duplication. Villin (92 kDa) resembles gelsolin with its six domains but harbors in addition a short headpiece domain (8 kDa) [8–10]. It is a major component of microvilli and was shown to be able to trigger microvilli formation in CV-1 cells in transfection experiments [11]. The headpiece contains an additional F-actin binding site and is responsible for villin's crosslinking activity in the absence of Ca^{2+} [12,13]. Class II capping proteins are structurally more divergent than the members of class I and in their active form they are either monomers or heterodimers. They lack an additional severing activity as has been shown for cap32/34 [14,15], radixin [16], and gCap39 [17].

Cap100 which we isolated from vegetative and developing *Dictyostelium discoideum* amoebae [18] caps actin filaments but shows no F-actin severing activity. These features would identify cap100 as a member of class II capping proteins. However, cloning and sequencing of the gene revealed that cap100 consists of six domains and a headpiece, and is highly homologous to the class I protein villin. The *Dictyostelium* 100 kDa capping protein is even more related to vertebrate villin than to *Dictyostelium* severin.

We draw from these data two major conclusions. (i) The presence of a six-domain protein in slime molds suggests that gene-duplications from one domain proteins (profilin) to three domains (severin, fragmin) and finally six domains (gelsolin, villin) [19] must have happened before *Dictyostelium* branched off. (ii) Different from villin the 100 kDa protein from *D. discoideum* neither fragments nor crosslinks actin filaments; its only major function appears to be the inhibition of elongation by capping barbed filament ends [18]. This indicates that capping of actin filaments is the evolutionarily 'oldest' function of an F-actin binding protein. Cap100 the villin-like molecule in *Dictyostelium* seems to be a true precursor of villin, and therefore we would like to designate this protein as 'protovillin'.

Correspondence address. M. Schleicher, Ludwig-Maximilians-University Munich, Schillerstraße 42, 80336 München, Germany. Fax: (49) (89) 5996-882.

2. MATERIALS AND METHODS

After digestion of protovillin with trypsin two partially sequenced peptides were used to synthesize highly degenerated oligonucleotide primers for polymerase chain reactions (PCR). Genomic *D. discoideum* DNA (1–3 μg) was mixed with 20 nmol dNTP, 500 pmol primers and 4 U Taq-polymerase (Amersham International, Amersham, UK) and subjected to 35 cycles with 1 min denaturation at 92°C, 1 min annealing at 50°C, and 3 min extension at 72°C. The denaturation step in the first cycle was extended to 3 min and the elongation step in the last cycle to 6 min. All cloning procedures were carried out essentially as described [20]. The PCR with the degenerated primers yielded a 944 bp fragment which was used for Southern blot analysis of genomic DNA. The probe hybridized to a 2.6 kb *Cla*I band; the corresponding fragment that turned out to carry 3′-coding and non-coding regions of the protovillin gene was cloned into pIC19R. Further Southern blot analysis employing a 5′-probe of this clone led to the isolation of a 4.7 kb *Eco*RI-*Nde*I-fragment harboring the missing 5′-sequences. The corresponding cDNA clone was isolated from a λgt11 cDNA library kindly provided by Drs. R. Kessin and M.-L. Lacombe, Columbia University [21] essentially as described [22]. DNA sequencing of both strands was done by the chain termination method [23,24] using uni and reverse primers as well as sequence specific oligonucleotide primers (19-mers). The resulting fragments were separated in a buffer gradient gel [25]. The sequences were analyzed with the programs from the University of Wisconsin Genetics Computer Group (UWGCG, [26]).

DNA and RNA were isolated as described [22]. Nuclear DNA was digested with restriction endonucleases, separated on 0.7% agarose gels in Tris-borate/EDTA buffer, pH 8.3, transferred to nitrocellulose (BA85, Schleicher & Schuell), and probed with nick-translated DNA under different conditions. For probing of RNA under high stringency conditions we used hybridization- and wash-buffer containing 50% formamide and 2 × SSC at 37°C; hybridization at intermediate stringency conditions was done with 30% formamide buffers at 37°C.

3. RESULTS AND DISCUSSION

3.1. *Isolation and analysis of genomic and cDNA clones coding for* Dictyostelium *protovillin*

The PCR with degenerated oligonucleotide primers yielded a stretch of DNA that hybridized to a genomic *Cla*I fragment of 2.6 kb. This fragment encoded about 2/3 of the C-terminus of protovillin and carried in addition an extended 3′-noncoding region. The 5′-sequence and the promoter region were located on a 4.6 kb *Eco*RI-*Nde*I fragment. Fig. 1 shows the complete nucleotide and deduced amino acid sequences of protovillin. The total protovillin gene is 2970 bp in length and codes for 959 amino acids with a calculated molecular mass of 109 kDa. The ATG start codon is preceded by three adenosines a feature of the majority of sequenced *Dictyostelium* genes [22]. The open reading frame is interrupted by an intron with a length of 90 bp at position #262. Splicing follows the 5′-GT...AG-3′ rule and the exact splicing sites were determined by sequencing a protovillin cDNA clone. The reading frame ends with a TAA codon, the most often used stop codon in *D. discoideum* [27]. Sequence analysis identified this protein as a *Dictyostelium* homologue of vertebrate villins. In contrast to all other villins described so far, the core-domain (amino acids 55–832) and headpiece (amino acids 881–959) are separated by a proline- and threonine-rich 'neck' (amino acids 833–880). In addition to its high proline and threonine content the neck is characterized by the internal sequence repeat TPKPITTPTV (amino acids 840–849 and 851–860).

3.2. *Genomic organization and Northern blot analysis of* D. discoideum protovillin

Genomic *Dictyostelium* DNA was probed with a cDNA fragment coding for the core domains (Fig. 2B, left panel) of protovillin or with a PCR fragment that represented the headpiece (Fig. 2B, middle panel). In *Eco*RI and *Eco*RI+*Nde*I digested DNA one band, in a *Cla*I digested DNA two bands were detected with a probe that coded for the gelsolin-like domains of protovillin (Fig. 2A, probe A). The appearance of two bands in the *Cla*I digestion is in agreement with an internal *Cla*I restriction site in the DNA probe. After incubation with a headpiece-specific DNA probe from protovillin (Fig. 2B, probe B) only one band could be detected in all cases. This suggested that (a) protovillin is encoded by a single gene, and (b) that at the DNA level the protovillin core and headpiece are unique sequences in *Dictyostelium*. Crosshybridization with DNA sequences coding for severin was not detected.

Northern blot analysis of growth phase and starved *Dictyostelium* cells (Fig. 2B, right panel) showed only one mRNA band with a size of approximately 3.0 kb. Comparable amounts of protovillin mRNA were present in vegetative (t_0) and in developing (t_5) cells. To check whether equal amounts of total RNA were loaded, blots were reprobed with a gelation factor specific probe [28] (Fig 2B, bottom right panel).

3.3. *Distinct sequence differences between protovillin and villin*

The sequence homology between protovillin and villin is highly significant (Fig. 3) but there are deviations that might reflect the difference in function. Protovillin was found to be a capping and a G-actin binding protein, it did not exhibit a severing activity nor a nucleating function under physiological salt conditions, a crosslinking activity could not be detected [18]. The F-actin capping activity was Ca^{2+}-independent but could be inhibited by PIP$_2$. Thus, in its interaction with actin filaments vertebrate villin shares with *Dictyostelium* protovillin only a PIP$_2$-inhibitable F-actin capping activity. Thorough studies on the domain structure of villin [29], gelsolin [30], and severin [5,6] localized the capping function to domain I and characterized the severing function as a cooperation between an F-actin binding site of domain II and the capping activity of domain I [3]. In analogy we would expect that domain I in protovillin exhibits capping activity but fails to sever actin filaments due to a still immature F-actin binding activity in domain II. In villin as well as in

```
 -52                                        aatttatttatttaaaaaaattgtaaatttataaatattgtacttttaaa
   1 ATGGAACCACCACTTGAACTACCAACACAAAGAAAAAGAGTTATCCCATCAAAATTTGGTATCCTCAAAAGAAATGCAGAAATTGAAGCA
     M  E  P  P  L  E  L  P  T  Q  R  K  R  V  I  P  S  K  F  G  I  L  K  R  N  A  E  I  E  A   30
  91 GAAAAAAACAGAGAAAATTTACAACAATCTTCATGTTTTAGTCATATAAATGAAATTGGTAAAGAAATTGGTTTAGAAATTTGGAAAATT
     E  K  N  R  E  N  L  Q  Q  S  S  C  F  S  H  I  N  E  I  G  K  E  I  G  L  E  I  W  K  I   60
 181 ATTGATGATTCAACAATTCAAAAAGTTCCAAAAGTGAATCATTCAACTTTTGAAACTAATAAATCTTATTTATTATTAATGgtttgtaat
     I  D  D  S  T  I  Q  K  V  P  K  V  N  H  S  T  F  E  T  N  K  S  Y  L  L  L  M  ---------  87
 271 taattaatttttttttttttttttaaaattaattatacatttattaattattttttttaaaaaaaaaaaaaaatttatagGGACAATTT
     ---------------------------------------------------------------------------------G  Q  F   90
 361 TATGATGGTAATATGAATATTAAAACTTATAATATTCATTTTTGGATTGGTGAATTATTAATAAATTCACAAGAAACAATTAATTTTTGT
     Y  D  G  N  M  N  I  K  T  Y  N  I  H  F  W  I  G  E  L  L  I  N  S  Q  E  T  I  N  F  C  120
 451 AATGATAGAATTGAAGAACTTGAAAGAATTATTAAATATAATCAAAAACAATTTGATTCAGAACAATTTTATCCAGAACCAATTCTTTAT
     N  D  R  I  E  E  L  E  R  I  I  K  Y  N  Q  K  Q  F  D  S  E  Q  F  Y  P  E  P  I  L  Y  150
 541 AGAGAATTTCAAGGTAAAGAAGGTGATATTTTTATGTCATACTTTAAATCTTATGGTGGTCCAAGATATGTTGCACCATTGAAATTAACA
     R  E  F  Q  G  K  E  G  D  I  F  M  S  Y  F  K  S  Y  G  G  P  R  Y  V  A  P  L  K  L  T  180
 631 TCAGCATCAGCAGCAATTGCAACAGCAGCAAAACAATATAAATTATTCCACCTTAAGGGTAGACGTAATATTAGAGTTAAACAAGTTGAT
     S  A  S  A  A  I  A  T  A  A  K  Q  Y  K  L  F  H  L  K  G  R  R  N  I  R  V  K  Q  V  D  210
 721 ATTTCATCAAAGTCATTAAATAGTGGCGATGTTTTCGTATTGGATTGTGAGGATTTCATTTATCAATGGAATGGAAGCGAATCAAGTCGT
     I  S  S  K  S  L  N  S  G  D  V  F  V  L  D  C  E  D  F  I  Y  Q  W  N  G  S  E  S  S  R  240
 811 TTAGAAAAAGGTAAAGGTTTAGATCTTACCATTAGATTACGTGATGAAAAATCAGCAAAAGCAAAGATTATCGTTATGGATGAAAATGAT
     L  E  K  G  K  G  L  D  L  T  I  R  L  R  D  E  K  S  A  K  A  K  I  I  V  M  D  E  N  D  270
 901 ACCGATAAGGATCATCCAGAATTTTGGAAGAGATTAGGTGGTTGCAAAGATGATGTTCAAAAAGCTGAACAAGGTGGTGATGATTTCGCT
     T  D  K  D  H  P  E  F  W  K  R  L  G  G  C  K  D  D  V  Q  K  A  E  Q  G  G  D  D  F  A  300
 991 TATGAAAAGAAATCGGTTGAACAAATTAAATTATATCAAGTTGAAAATTTAAATTATGAAGTACATCTTCATCTAATCGATCCAATTGGT
     Y  E  K  K  S  V  E  Q  I  K  L  Y  Q  V  E  N  L  N  Y  E  V  H  L  H  L  I  D  P  I  G  330
1081 GATGTTTATTCAACCACTCAATTGAATGCAGAGTTTTGTTATATTTTAGATTGTGAAACCGAATTATACGTTTGGTTAGGTAAAGCATCA
     D  V  Y  S  T  T  Q  L  N  A  E  F  C  Y  I  L  D  C  E  T  E  L  Y  V  W  L  G  K  A  S  360
1171 GCAAATGATCAAAGAACAGTTGCAATGGCAAATGCAATGGATTTATTACATGAAGATAATAGACCAAGTTGGACACCAATTATAAAGATG
     A  N  D  Q  R  T  V  A  M  A  N  A  M  D  L  L  H  E  D  N  R  P  S  W  T  P  I  I  K  M  390
1261 ACTCAAGGCTCTGAGAATACTCTCTTTAAAGATAAATTTAAAAAAGGTAGTTGGGGTGAATATGTTAATGATAACTTTGAAAAGAAACCA
     T  Q  G  S  E  N  T  L  F  K  D  K  F  K  K  G  S  W  G  E  Y  V  N  D  N  F  E  K  K  P  420
1351 ATCACAGGTAAAGGTGTTGCTGCCAAAGCAGTTCAAGAGAAAATCAATGTAGATGCTCTTCATAATCCTGAAAAATATCAACTTTCAAAG
     I  T  G  K  G  V  A  A  K  A  V  Q  E  K  I  N  V  D  A  L  H  N  P  E  K  Y  Q  L  S  K  450
1441 GAAGAAAGAAAATCAACTATTCCAACTCTTCATCATGTCGATGATAAACATAGAGGTGAATTAAAGATTTGGCATGTTAGAAATCGTAAT
     E  E  R  K  S  T  I  P  T  L  H  H  V  D  D  K  H  R  G  E  L  K  I  W  H  V  R  N  R  N  480
1531 AAATTTGAAATTTCTCAATCTGAATTTGGTTTATTCTATAATCAATCTTGTTATTTAGTATTGTTTACTTTATTTGCTGCCGATGGCTCA
     K  F  E  I  S  Q  S  E  F  G  L  F  Y  N  Q  S  C  Y  L  V  L  F  T  L  F  A  A  D  G  S  510
1621 AATAATTCAATATTATACTATTGGCAAGGTAGATTCAGTTCAAGTGAAGATAAAGGTGCCGCTGCTCTTTTGGCTAAAGATGTTGGTAAA
     N  N  S  I  L  Y  Y  W  Q  G  R  F  S  S  S  E  D  K  G  A  A  A  L  L  A  K  D  V  G  K  540
1711 GAATTGCATCGTTCTTGTATTCATGTCAGAACCGTTCAAAATAAAGAACCAAACCACTTTTTAGAACATTTCCAAGGTCGTATGGTTGTT
     E  L  H  R  S  C  I  H  V  R  T  V  Q  N  K  E  P  N  H  F  L  E  H  F  Q  G  R  M  V  V  570
1801 TTTAAAGGTTCAAGACCAAATGCAACCACTGAAGTTTCATTGGAAAATTTATCTTCATCACTTCAAGGCTTATATCATGTTAGAGGTACT
     F  K  G  S  R  P  N  A  T  T  E  V  S  L  E  N  L  S  S  S  L  Q  G  L  Y  H  V  R  G  T  600
1891 GAGCCAATTAATATTCATTCAATTCAAGTTGAAAAAGCAATTTCATCATTGGATTCAAATGATTCTTTCATTTTAGTTAATTTTAAAAAT
     E  P  I  N  I  H  S  I  Q  V  E  K  A  I  S  S  L  D  S  N  D  S  F  I  L  V  N  F  K  N  630
1981 ACAATTTCCTATATTTGGGTTGGTAAATACTCTGATGAAAAAGAAGCTGCCCTTCAAATTTCTTCAAATGTTTTCACTGGTTATAATTTC
     T  I  S  Y  I  W  V  G  K  Y  S  D  E  K  E  A  A  L  Q  I  S  S  N  V  F  T  G  Y  N  F  660
2071 CAATTGATTGATGAAGGTGATGAAACTTCTGAATTTTGGGAATCTTTAGAAACAAACTCATCCTTATCACTCTTAAAAGATTATTACACT
     Q  L  I  D  E  G  D  E  T  S  E  F  W  E  S  L  E  T  N  S  S  L  S  L  L  K  D  Y  Y  T  690
2161 CAATTGAGAACCGTTGAACAAGAGAAGAAAACTCGTTTATTCCAATGTTCAAATAATAGTGGTGTATTCAAAGTATTTGAAATTCATGAT
     Q  L  R  T  V  E  Q  E  K  K  T  R  L  F  Q  C  S  N  N  S  G  V  F  K  V  F  E  I  H  D  720
2251 TTCTCTCAAGATGATTTAGATTCTGATGATGTAATGATCTTGGATAATCAAAAACAAATTTTCGTTTGGGTTGGTAAAGAAAGTAGTGAT
     F  S  Q  D  D  L  D  S  D  D  V  M  I  L  D  N  Q  K  Q  I  F  V  W  V  G  K  E  S  S  D  750
2341 ACTGAAAAATTAATGGCAAATGAAACTGCTCTCGAGTATATTATGAATGCTCCAACTCATAGAAGAGATGATCCAATCTTTACCATTCAA
     T  E  K  L  M  A  N  E  T  A  L  E  Y  I  M  N  A  P  T  H  R  R  D  D  P  I  F  T  I  Q  780
2431 GATGGTTTTGAACCACATGAATTCACTTTTAATTTCCATGCTTGGCAAGTTAATAAAACTCAACAAGATTCTTATAAATCAAAATTAAGT
     D  G  F  E  P  H  E  F  T  F  N  F  H  A  W  Q  V  N  K  T  Q  Q  D  S  Y  K  S  K  L  S  810
2521 GCTATCCTTGGTTCAAACAATAGTGGTCCAGCTTCTCCAATTATGTTACCTACTTCTGGTGTTACTTTAAAACCAACAACAGCAGCTACT
     A  I  L  G  S  N  N  S  G  P  A  S  P  I  M  L  P  T  S  G  V  T  L  K  P  T  T  A  A  T  840
2611 CCAAAACCAATAACTACACCAACAGTAACTACTCCTAAACCAATAACTACACCGACAGTAGCTACTCTAAAAACAGTTACACCTGCTGTA
     P  K  P  I  T  T  P  T  V  T  T  P  K  P  I  T  T  P  T  V  A  T  L  K  T  V  T  P  A  V  870
2701 ACTTTAAAGCCAACAACTGTAACTACTCCATCTAAAGTTGCTACCACTACAAATACCTCTACTCCATCACCAACTACAATTACAACATTC
     T  L  K  P  T  T  V  T  T  P  S  K  V  A  T  T  T  N  T  S  T  P  S  P  T  T  I  T  T  F  900
2791 TATCCATTATCAGTTTTAAAACAAAAAACTAATTTACCAAATGATATTGATAAAAGTTGTCTTCACTTGTACCTTTCTGATGAAGAATTT
     Y  P  L  S  V  L  K  Q  K  T  N  L  P  N  D  I  D  K  S  C  L  H  L  Y  L  S  D  E  E  F  930
2881 TTATCAACCTTTAAAATGACAAAAGAAATCTTCCAAAAAACTCCTGCTTGGAAAACTAAACAATTACGTGTTGATAATGGATTATTTTAA
     L  S  T  F  K  M  T  K  E  I  F  Q  K  T  P  A  W  K  T  K  Q  L  R  V  D  N  G  L  F  *  959
2971 attaaaaacaattgataaaaaaaattaaaaataaaattaatt
```

Fig. 1. Genomic DNA and deduced amino acid sequence of *D. discoideum* protovillin. Small letters represent noncoding DNA stretches, the intron is marked with a dashed line. The two underlined polypeptides indicate the peptide sequences obtained by Edman degradation. The putative polyadenylation site is written with bold letters.

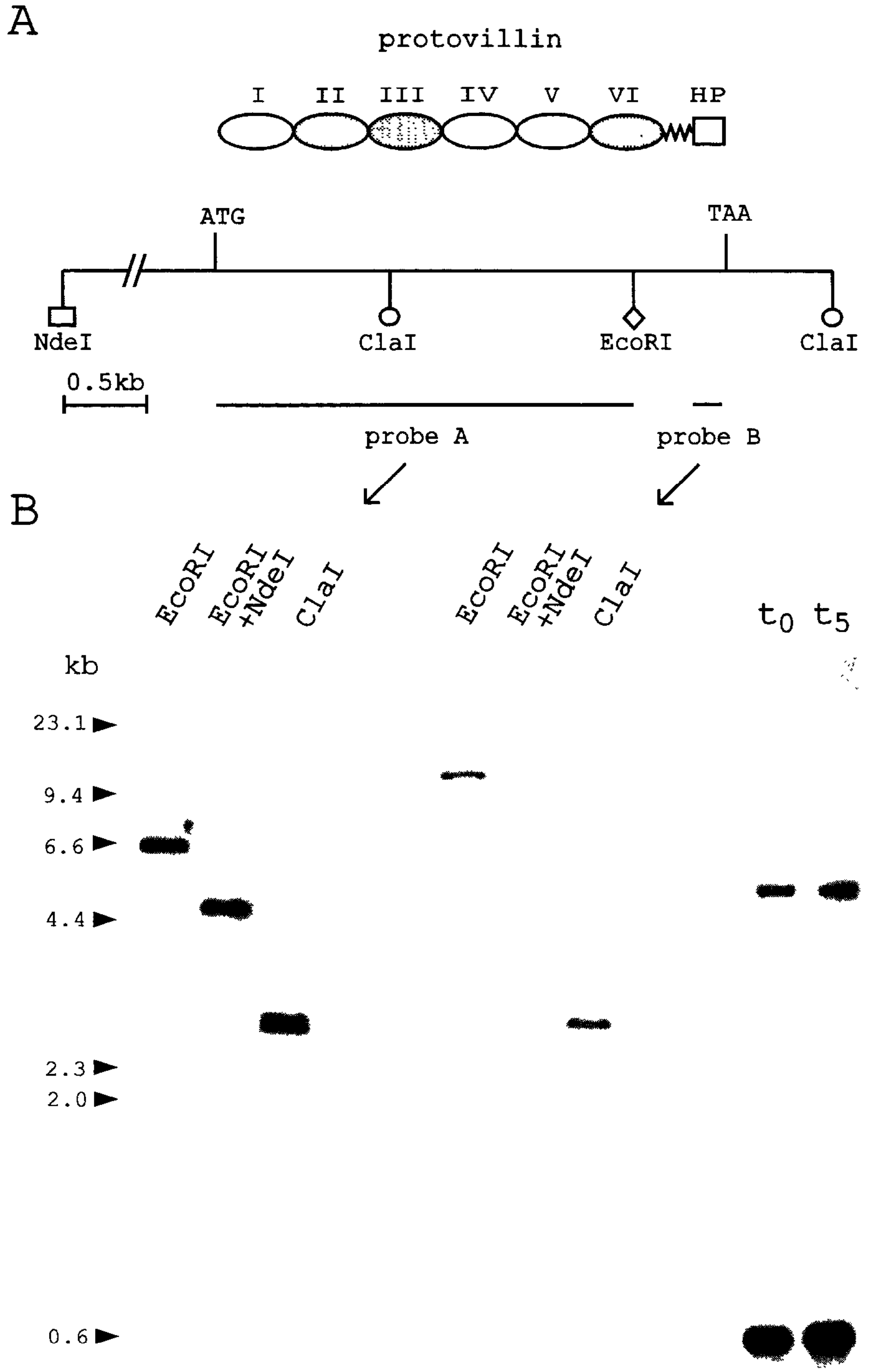

Fig. 2. Identification of a single copy gene that codes for *D. discoideum protovillin Genomic Dictyostelium* DNA was digested with *Eco*RI, *Eco*RI+*Nde*I, or *Cla*I and the resulting Southern blots were probed under medium stringency conditions with probe A, representing the protovillin core domain (B, left panel), or with the headpiece-specific fragment probe B (middle panel). Fig. 2B (right panel) shows a Northern blot of vegetative (t_0) and developing (t_5) *Dictyostelium* cells. The blot was probed under high stringency conditions and reprobed with a gelation factor specific probe (B, bottom right panel).

gelsolin and severin the first amino acids of domain II seem to be crucial for binding along an actin filament. In domain II of protovillin these amino acids (Fig. 3, marked !!!) are missing, whereas the PIP_2-binding motif [31] is present (+++). This could explain the absence of a severing activity as well as the inhibitory effect of PIP_2 for

```
           TAIL           →←  ---------------------------------------------
 30 AEKNRENLQQSSCFSHINEIGKEIGLEIWKIIDDSTIQKVPKVNHSTFETNKSYLLLMGQFYDGNMNIKTY
  4 SKKVTGKLDKTTP...........GIQIWR.IENMEMVPVPTKSYGNFYEGDCYVLLSTRKTGSGF...SY
    * o    o* **            o**oo* o**  *  oo    * o     o*oo *    *     *o

    ----------------------- DOMAIN I --------------------------------------
101 NIHFWIG.ELLINSQETINFCNDRIEELERIIKYNQKQFDSEQFYPEPILYREFQGKEGDIFMSYFKSYGG
 60 NITYWLGKNSSQDEQGAAAIYTTQMDEY...............LGSVAVQHREVQGHESETFRAYFK....
    oo *o*o *   * o *       **o                     *  oo oo*o** o * oo

    ----------------- →← !!!!!!!!!!+++++++++
171 PRYVAPLKLTSASAAIATAAKQY........KLFHLKGRRNIRVKQVDISSKSLNSGDVFVLDCEDFIYQW
112 ......QGLIYKQGGVASGMKHVETNTYNVQRLLHVKGKKNVVAAEVEMSWKSFNLGDVFLLDLGQLIIQW
            o    ***o** o          *o o*oo**o*   *o**o oo o oooo*oo  * o oo

                             DOMAIN II
234 NGSESSRLEKGKGLDLTIRLRD.EKSAKAKIIVMDENDTDKDHPEFWKRLGGCKDDVQKAEQGGDDFAYEK
177 NGPESNRAERLRAMTLAKDIRDRERAGRAKGGVV.EGENEAASPELMQALTHVLGEKKNIKAATPDEQVHQ
    oo oo o o* *** o*  *oo o****oo  o* o * *   oo    o*    *      ** o

    →←  -------------------------------------------------- DOMAIN III ---
303 KSVEQIKLYQVENL..NYEVHLHLIDPIGDVYSTTQLNAEFCYILD.CETELYVWLGKASANDQRTVAMAN
247 ALNSALKLYHVSDASGNLVIQEVAIRPLTQ....DMLQHEDCYILDQAGLKIFVWKGKNANKEEKQQAMSR
         *ooo o     o  *    o o***      o* o ooooo     * oo oo *  ***  oo*

    ------------------------------------------------------→←      LINKER
372 AMDLLHEDNRPSWTPIIKMTQGSENTLFKDKFKKGSWGEYVNDNFEKKPITGKGVAAKAVQEKINVDALHN
314 ALGFIKAKNYLASTSVETENDGSESAVFRQLFQK.WTVPNQTSGLGKTHTVGKVAKVEQV...........
    o*  **  o  * o *    *ooo **o** o o            o    oo      o

                 →←  --------------------------------------------------
443 PEKYQLSKEERKSTIPTLHHVDDKHRGELKIWHVRNRNKFEISQSEFGLFYNQSCYLVLFTLFAADGSNNS
373 ..KFDATTSHVKPEVAAQQKMVDDGSGEAEVWRVENQELVPVEKRWSGHFYGGDCYLVLYTYY.VGPKVNR
      o** *    o  * *  **     oo  *o*o o *   *     o oo   ooooo** *      o

    -------------------- DOMAIN IV --------------------------------- →←
514 ILYYWQGRFSSSEDKGAAALLAKDVGKELHRSCIHVRTVQNKEPNHFLEHFQGRMVVF.KGSRPNATTEVS
414 IIYIWQGRHASTDELAASAYQAVFLDQKYNNEPVHVRVTMGKEPAHLMAIFKGKMVVYENGSSRAGGTE..
    o*o oooo *o*** *o*o  o  *        *ooo    ooo o *  o o*ooo*  oo   **oo

                                               DOMAIN V
584 LENLSSSLQGLYHVRGTEPINIHSIQVEKAISSLDSNDSFILVNFKNTISYIWVGKYSDEKEAALQISSNV
510 ....PASSTRLFHVHGTNEYNTKAFEVPVRAASLNSNDVFVLKTPSSCYSWYGKGCSGDEREMG.KMVADI
         *o   o*oo*oo*  o ** *o    *oo*ooo o*o       o*   o  *oo*o *  * ***

                                            →←  ---------------------------
655 FTGYNFQLIDEGDETSEFWESLETNSSLSLLKDYYTQLRTVEQEKKTRLFQCSNNSGVFKVFEIHDFSQDD
576 ISKTEKPVVAEGQEPPEFWVALGGKTSYANSK....RLQEENPSVPPRLFEVSNKTGRFLATEIVDFTQDD
     *  *  ** oo*o  ooo *o * *o *  o     o   *     ooo* oo *o o   oo oo*ooo

    --------------------------- DOMAIN VI --------------------------------
726 LDSDDVMILDNQKQIFVWVGKESSDTEKLMANATALEYIMNAPTHRR.DDPIFTIQDGFEPHEFTFNFHAW
643 LDENDVYRRDTWDQIFFWIGKGANESEKEAAAETAQEYLRSHPGSRDLDTPIIVVKQGFEPPTFTGWFMAW
    oo *oo   o   ooo o*oo * * oo  o  oo oo*    * o  o oo  * *oooo  oo  o oo

    ----------------------------------- →←‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡ NECK ‡‡‡‡‡‡‡‡‡
796 Q.VNKTQQDSYKSKLSAILGSNNSGPASPIMLPTSGVTLKPTTAATPKPITTPTVTTPKPITTPTVATLKT
714 DPLCWSDRKSYD.ELKAELGDNAS.....IGQLVSGLT.................................
    * *  **  oo   o o oo o o     o    oo*o

    ‡‡‡‡‡‡‡‡‡‡‡‡‡‡→←                     HEADPIECE
866 VTPAVTLKPTTVTTPSKVATTTNTSTPSPTTITTFYPLSVLKQKT..NLPNDIDKSCLHLYLSDEEFLSTF
746 ...............SKNEVFTATTTLVPTKLETF.PLDVLVNTAAEDLPRGVDPSRKENHLSDEDFKAVF
                   oo    o o*o  oo * oo oo oo * *  *oo  *o o     oooo*o * o

                      >>>> →
935 KMTKEIFQKTPAWKTKQLRVDNGLF
801 GMTRSAFANLPLWKQQNLKKEKGLF
     oo*  o   o oo   *o* * ooo
```

Fig. 3. Amino acid comparison between protovillin and villin. Identical amino acids are marked with ○, conserved residues with * (conserved residues were defined according to the permutation matrix [PIR] as: D,E,Q,N; F,Y,W; I,L,V,M; S,T,A,G; or H,R,K). The borders of domains [35] are indicated with arrows; !, +, > and † mark peptides discussed in the text.

capping of actin filaments. The extend of similarity of domain I in protovillin as compared with all domains of villin (BESTFIT, [26]) is best in villin's domain I followed by domain IV, whereas the other domains show a moderate homology. The presence of the headpiece which is an unique and functionally very important part of vertebrate villin characterized the *Dictyostelium* protein clearly as a villin-like molecule. Point mutations in this region of villin and expression of the mutated protein in CV-1 cells showed that a short cluster of the charged amino acids KKEK (amino acids 819–822) is absolutely required for activity [13]. In protovillin the same stretch of amino acids reads RVDN (Fig. 3, >>>; amino acids 953–956): this divergence is in good agreement with the lack of an F-actin bundling acitvity.

3.4. *Mix-and-match and ripe-for-action*

The principle of modular isoforms is well established among actin-binding proteins. Prominent examples are the F-actin crosslinking molecules α-actinin [32], gelation factor [28], spectrin [33], filamin [34]. The rods of these elongated molecules are either shaped by α-helical (α-actinin, spectrin) or strictly β-sheet sub-domains (gelation factor, filamin). Although the rods are different in structure and origin, all of these proteins share the same actin-binding site. This led to the speculation that in a mix-and-match fashion rearrangements in the genome generated proteins with new sets of functional domains. Vertebrate villins so far fitted very well to this hypothesis and was believed to be generated from a gelsolin-like six-domain protein which had acquired a headpiece region and thus became an F-actin crosslinking molecule, essential for the formation of microvilli. The presence of protovillin in *D. discoideum* suggests a modification of this mix-and-match hypothesis. Domain swapping might have happened but if one considers the surprisingly low similarity between severin and protovillin, it occurred long before *Dictyostelium* branched off in evolution. As of now protovillin looks like a villin-precursor that performs only an archaic capping activity and is not yet ripe-for-action. It still lacks essential changes in the sequence that allow bundling of actin filaments and consequently the appearance of microvilli. It remains to be shown whether a protovillin exists in cells of other organisms as well and only the specialized epithelial cells with microvilli harbor in addition a mature villin.

Acknowledgements: We thank Dr. Günther Gerisch and Dr. Paul R. Fisher for helpful discussion; we are grateful to Gerhard Rahn for providing oligonucleotide primers and to Daniela Rieger for help in protein purification The work was supported by grants from the Deutsche Forschungsgemeinschaft to A.A N. and M.S.

REFERENCES

[1] Vandekerckhove, J. (1990) Curr. Opin. Cell Biol. 2, 41–50.

[2] Schleicher, M. and Noegel, A.A. (1992) New Biol 4, 461–472

[3] Weeds, A. and Maciver, S. (1993) Curr. Opin. Cell Biol. 5, 63–69.

[4] Yin, H.L., Janmey, P.A. and Schleicher, M. (1990) FEBS Lett 264, 78–80.

[5] Eichinger, L., Noegel, A.A. and Schleicher, M. (1991) J. Cell Biol. 112, 665–676

[6] Eichinger, L. and Schleicher, M. (1992) Biochemistry 31, 4779–4787.

[7] André, E., Lottspeich, F., Schleicher, M. and Noegel, A.A. (1988) J Biol. Chem. 263, 722–727.

[8] Bretscher, A. and Weber, K. (1979) Proc. Natl. Acad. Sci. USA 76, 2321–2325.

[9] Pringault, E., Arpin, M., Garcia, A., Finidori, J and Louvard, D. (1986) EMBO J. 5, 3119–3124.

[10] Bazari, W.L., Matsudaira, P, Wallek, M , Smeal, T , Jakes, R. and Ahmed, Y. (1988) Proc Natl. Acad. Sci. USA 85, 4986–4990.

[11] Friederich, E., Huet, C , Arpin, M. and Louvard, D (1989) Cell 59, 461–475.

[12] Bretscher, A. and Weber, K. (1980) Cell 20, 839–847.

[13] Friederich, E., Vancompernolle, K., Huet, C., Goethals, M , Finidori, J , Vandekerckhove, J. and Louvard, D. (1992) Cell 70, 81–92.

[14] Hartmann, H., Noegel, A.A., Eckerskorn, C., Rapp, S. and Schleicher, M. (1989) J Biol. Chem. 264·12639–12647

[15] Haus, U., Hartmann, H., Trommler, P., Noegel, A.A. and Schleicher, M. (1991) Biochem. Biophys Res. Commun. 181, 833–839

[16] Tsukita, S., Hieda, Y. and Tsukita, S. (1989) J. Cell Biol. 108, 2369–2382.

[17] Yu, F.-X., Johnston, PA , Sudhof, T.C. and Yin, H.L. (1990) Science 250, 1413–1415.

[18] Hofmann, A , Eichinger, L , André, E., Rieger, D. and Schleicher, M (1992) Cell Motil. Cytoskel 23, 133–144.

[19] Matsudaira, P. and Janmey, P. (1988) Cell 54, 139–140.

[20] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in: Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

[21] Lacombe, M -L., Podgorski, G.J., Franke, J and Kessin, R.H. (1986) J. Biol. Chem 261, 6645–6653

[22] Doering, V., Schleicher, M. and Noegel, A A. (1991) J. Biol. Chem. 266, 17509–17515.

[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc Natl. Acad. Sci. USA 74, 5463–5468.

[24] Chen, E.Y and Seeburg, P.H. (1985) DNA 4, 165–170.

[25] Biggin, M.D., Gibbson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.

[26] Devereux, J , Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.

[27] Sharp, P.M. and Devine, K.M. (1989) Nucleic Acids Res 17, 5029–5038

[28] Noegel, A.A., Rapp, S., Lottspeich, F., Schleicher, M. and Stewart, M. (1989) J Cell Biol 109, 607–618.

[29] de Arruda, M V., Bazari, H., Wallek, M and Matsudaira, P. (1992) J Biol. Chem 267, 13079–13085.

[30] Way, M., Pope, B., Gooch, J., Hawkins, M. and Weeds, A.G (1990) EMBO J. 9, 4103–4109.

[31] Yu, F.-X., Sun, H.-Q , Janmey, P.A. and Yin, H.L (1992) J. Biol. Chem. 267, 14616-14621.

[32] Witke, W., Schleicher, M and Noegel, A.A. (1992) Cell 68, 53–62

[33] Beyers, T.J., Husain-Chishti, A., Dubreuil, R.R., Branton, D. and Goldstein, L.S B. (1989) J. Cell Biol. 109, 1633–1641.

[34] Gorlin, J.B., Yamin, R., Egan, S., Stewart, M., Stossel, T.P., Kwiatkowski, D.J. and Hartwig, J.H. (1990) J. Cell Biol 111, 1089–1105.

[35] Way, M and Weeds, A. (1988) J. Mol. Biol. 203, 1127–1133.