Sequence and Domain Organization of Scruin, an Actin-Cross-linking Protein in the Acrosomal Process of *Limulus* Sperm

Michael Way,* Mitchell Sanders,* Celia Garcia,§ Jun Sakai,* and Paul Matsudaira‡

*Whitehead Institute, Cambridge, Massachusetts 02142; ‡Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; and §Department of Parasitology, University of São Paulo, São Paulo 05508-900, Brazil

Abstract. The acrosomal process of *Limulus* sperm is an 80-μm long finger of membrane supported by a crystalline bundle of actin filaments. The filaments in this bundle are crosslinked by a 102-kD protein, scruin present in a 1:1 molar ratio with actin. Recent image reconstruction of scruin decorated actin filaments at 13-Å resolution shows that scruin is organized into two equally sized domains bound to separate actin subunits in the same filament. We have cloned and sequenced the gene for scruin from a *Limulus* testes cDNA library. The deduced amino acid sequence of scruin reflects the domain organization of scruin: it consists of a tandem pair of homologous domains joined by a linker region. The domain organization of scruin is confirmed by limited proteolysis of the purified acrosomal process. Three different proteases cleave the native protein in a 5-kD Protease-sensitive region in the middle of the molecule to generate an NH_2-terminal 47-kD and a COOH-terminal 56-kD protease-resistant domains. Although the protein sequence of scruin has no homology to any known actin-binding protein, it has similarities to several proteins, including four open reading frames of unknown function in poxviruses, as well as kelch, a *Drosophila* protein localized to actin-rich ring canals. All proteins that show homologies to scruin are characterized by the presence of an ~50-amino acid residue motif that is repeated between two and seven times. Crystallographic studies reveal this motif represents a four β-stranded fold that is characteristic of the "superbarrel" structural fold found in the sialidase family of proteins. These results suggest that the two domains of scruin seen in EM reconstructions are superbarrel folds, and they present the possibility that other members of this family may also bind actin.

Atomic resolution structures of the actin monomer have been obtained from cocrystals of monomeric actin complexed with either DNAse I (Kabsch et al., 1990), profilin (Schutt et al., 1993), or segment 1 of gelsolin (McLaughlin et al., 1993). In contrast, actin filaments have not been crystalized in vitro. However, it has been possible to obtain a model for the structure of the actin filament through the comparison of data obtained from x-ray fiber diffraction patterns of oriented actin filaments and calculated patterns derived from the G-actin structure, together with the known helical parameters of the actin filament (Holmes et al., 1990; Lorenz et al., 1993).

An alternative approach to structural analysis of actin filaments has come from image reconstructions of electron micrographs of both actin filaments, as well as filaments decorated with actin-binding domains (Lehman et al., 1994; McGough et al., 1994; Milligan et al., 1990; Rayment et al., 1993; Schroder et al., 1993). These studies have been typically limited to a maximum resolution of ~20 Å. However, one system, the acrosomal process from *Limulus*, has the potential to reveal the structure of an actin filament at 7-Å resolution because the filaments are organized into a crystalline bundle that diffracts to high resolution in the 400-kV cryoelectron microscope (Schmid et al., 1993).

The actin bundle found in the *Limulus* acrosomal process is a model system for understanding the structure of an actin bundle, the cell biology of the acrosome reaction, and the biochemistry of actin cross-linking proteins. In unactivated *Limulus* sperm, actin is present as a preformed bundle of filaments that is coiled around the base of the nucleus (DeRosier et al., 1982; Tilney, 1975). During the acrosome reaction, the bundle uncoils into a straight 80-μm long acrosomal process that projects from the anterior of the sperm (Tilney, 1975). It has been shown that the twist of the actin filaments is changed during the transition from a coiled to straight bundle, but the mechanism by which this occurs is unknown (DeRosier et al., 1982).

The actin filaments in the *Limulus* acrosomal process are

Address all correspondence to Paul Matsudaira, Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02142. Tel.: (617) 258-5188. Fax: (617) 258-7663.

The current address for Michael Way is European Molecular Biology Laboratory, 69012 Heidelberg, Germany.

cross-linked by equimolar amounts of a 102-kD protein, scruin (Schmid et al., 1991). Schmidt et al. (1993) have shown that the bundle of actin filaments isolated from activated sperm is crystalline and diffracts to at least 7 Å in the cryoelectron microscope. Consequently, the *Limulus* acrosomal process has emerged as a model system to understand the interactions of a cross-linking protein with actin using image reconstruction techniques (Bullitt et al., 1988; Owen and DeRosier, 1993; Schmid et al., 1994). Two 13-Å resolution helical reconstructions of scruin-decorated actin filaments reveal that scruin is organized into two globular domains; one domain binds actin subdomain 1 and the other domain binds subdomain 3 of the adjacent actin subunit in the same filament (Owen and DeRosier, 1993; Schmid et al., 1994). Based on this localization, both domains of scruin are proposed to interact with a conserved helix-loop-β strand motif found in these two ancient duplicated subdomains in actin (Schmid et al., 1994). In addition, because each scruin molecule is bound to a single actin filament then, cross-links between filaments must occur through interactions between scruin molecules on adjacent filaments.

To facilitate structural analysis of the *Limulus* acrosomal bundle at higher resolution, and to investigate the relationship between scruin and other actin-binding proteins, we have determined the sequences of scruin and the actin isoform found in the bundle. Our sequence and biochemical analysis of scruin suggests that the protein is organized into two domains, each composed of sixfold repeat of a 50-amino acid residue motif. This motif places scruin in a large and diverse family of proteins found in poxviruses, yeast, *Drosophila*, *Arabidopsis*, and mammals.

Materials and Methods

Protein Isolation and NH_2-Terminal Sequencing

The true discharge form of the acrosomal process was purified as described previously (Schmid et al., 1991) in the presence of a cocktail of protease inhibitors (0.1 mM PMSF, 1 KIU/ml aprotinin, and 0.2 μg/ml leupeptin) at all stages. Final protein concentrations were determined either by the BCA (Pierce Chemical Co., Rockford, IL) or the Bio Rad Laboratories (Richmond, CA) assays and the purity of acrosome bundle preparations assessed by SDS-PAGE.

The scruin and actin in acrosomal bundles were separated by denaturing the bundles in 6 M guanidine HCl, followed by modification on their cysteine residues with 4-vinyl pyridine, and chromatographed through a size exclusion column (TSK300SWXL; ToyoSoda, Tokyo, Japan). The scruin and actin containing peaks were collected, the guanidine was removed by dialysis, and the protein was dried by vacuum centrifugation. Alternatively, intact bundles were dissolved in SDS sample buffer and electrophoresed through SDS polyacrylamide gels, and were then electroblotted onto Immobilon P (Millipore Corp., Bedford, MA) as previously described by Matsudaira (1987).

The purified protein or band was incubated overnight with cyanogen bromide in 90% formic acid or Achromobacter protease I, and the fragments were purified by reverse phase HPLC. NH_2-terminal peptide sequences were obtained using a gas phase sequencer, (model 2090E; Beckman Instruments, Inc., Fullerton, CA, or model 475; Applied Biosystems, Inc., Foster City, CA) both equipped for on-line identification of phenylthiohydantoin-amino acids. In addition, 2-μl aliquots of peaks were analyzed by matrix-assisted laser desorption mass spectrometry using a LASERMAT (Finnigan MAT, San Jose, CA). Masses were calculated using insulin as an internal calibration.

Limited Proteolysis of Acrosomal Bundles

Purified acrosomal bundles were resuspended to a final concentration of 1 mg/ml in acrosome buffer (150 mM NaCl, 1 mM $MgCl_2$, 20 mM Tris-HCl, pH 8.0) containing either 2.0 mM EGTA or 0.5 mM $CaCl_2$. 30–50 μg of acrosome bundles were incubated at room temperature with sequencing grade trypsin (1:1,450; wt/wt), chymotrypsin (1:100; wt/wt), Endo-Asp protease (1:50; wt/wt), or V-8 protease (1:100; wt/wt) (Boehringer Mannheim, Indianapolis, IN). A 5-μl aliquot from the reaction was quenched after 1, 5, 20, and 60 min with 0.5–1.0 μl of PMSF (100 mM in isopropanol), solubilized with 4× SDS sample buffer, boiled for 2 min, and electrophoresed through 7.5–20% gradient or straight 7.5% SDS-polyacrylamide gels (Matsudaira and Burgess, 1978). For NH_2-terminal sequence analysis, individual fragments were blotted to Immobilon P membranes (Millipore) (Matsudaira, 1987).

Protease-treated acrosomal processes and untreated controls were negatively stained with 1% uranyl acetate and examined in an electron microscope (model 410; Philips Technologies, Cheshire, CT).

Library Construction and Clone Isolation

Total RNA was isolated from *Limulus* testes and poly A+ RNA was selected using total RNA and Poly(A) quik mRNA isolation kits (Stratagene, La Jolla, CA). Random and oligo dT primed cDNA libraries were synthesized from 5 μg of twice selected poly A+ RNA using a cDNA synthesis kit (Timesaver; Pharmacia Fine Chemicals, Piscataway, NJ). The resulting cDNA libraries were cloned into the Eco RI site of λ Zap II (Stratagene) and packaged in vitro (Gigapack II Plus; Stratagene).

Based on two scruin-derived protein sequences, YMVDNDTI and MIRHVTM, a pair of degenerate oligonucleotide primers, S4 (TAC/T ATG GTX GAC/T AAC/T GAC/T ACX ATA/C/T) and S2 (ATG ATA/C/T A/CGX CAC/T AAA/G GTX ACX ATG), were synthesized. These primers were used to amplify scruin cDNA probe, S42, from the randomly primed *Limulus* testes cDNA library using the following PCR cycling conditions (94°C 3.0 min followed by 15 cycles of 94°C 1.0 min, 45°C 1.0 min, 72°C 1.5 min, and then 25 cycles of 94°C 1.0 min, 50°C 1.0 min, and 72°C 1.5 min). Using standard hybridization conditions, the S42 probe identified four positives in 600,000 plaques of the oligo dT primed *Limulus* testes cDNA library whose inserts had been size selected >1.3 kb. All four positives were rescreened to homogenity and excised from λ Zap II in vivo into Bluescript according to the manufacturer's instructions. Similar methods were used to isolate actin clones from the same library using an actin DNA probe corresponding amino acid residues 56–374 of chicken β actin generated by PCR. Northern analysis was performed using standard procedures.

DNA Sequencing and Analysis

The cDNA sequences of all scruin and actin clones were derived from double-strand sequencing of random clones generated by sonication (Bankier et al., 1987), using Sequenase II (U.S. Biochemical Corp., Cleveland, OH) and the Bluescript SK/KS primers. Assembly and analysis of scruin and actin sequences was achieved using the DNASTAR software package (DNASTAR Inc., Madison, WI), while database sequence searches were run using BLAST (Altschul et al., 1990).

Construction and Expression of Scruin Domains

To express scruin and scruin domains in *Escherichia coli*, the scruin cDNA was modified for the T7-based expression vector pMW172 (Way et al., 1990) by inserting an NdeI site adjacent to the first codon and a TAA-TAG double-stop HindIII site after the last codon. All changes were achieved using standard PCR techniques, and the complete sequence of all expression constructs was verified by double strand sequencing with Sequenase II (U.S. Biochemical Corp.). All pMW172 constructs were expressed in the *E. coli* strains BL21(DE3) or BL21(DE3) pLysS. In all cases, cultures were induced to express by the addition of isopropyl-β-thiogalactopyranoside to a final concentration of 0.6 mM once freshly inoculated cultures had reached an OD_{600} ~0.7–1.0.

Expressed protein was isolated from inclusion bodies as described for the methods developed for the purification of actin-binding domains of gelsolin and α-actinin (Way et al., 1992). Briefly, inclusion bodies were solubilized in 10 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.2 mM EGTA, 1 mM NaN_3, and 1 mM DTT (buffer A) containing 8 M urea. After solubilization, inclusion bodies were diluted to a final concentration of 6 M urea with buffer A and clarified at 40,000 *g* before filtration through a 0.2-μm filter. The resultant filtrate was applied to a DEAE-cellulose column equilbrated in buffer A containing 6 M urea. All expressed scruin mutants bound the column and were eluted with an NaCl gradient from 20 to 300 mM. Column fractions containing the desired protein were concentrated by centrifugation on Centriprep 30 units (Amicon, Beverly, MA) and then chro-

matographed through Sephacryl S200 (Pharmacia) in buffer A plus 6 M urea. The protein eluted in a single peak that was exhaustively dialyzed against 10 mM Tris-HCl, pH 8.0, 0.2 mM EGTA, 1 mM DTT, and 1 mM NaN_3 to remove the urea and stored in the same buffer at 4°C.

Results

Isolation of Scruin cDNA

Using the degenerate oligonucleotide primers S4 and S2 initially, we amplified a 144-bp cDNA probe (S42) from the randomly primed cDNA. The sequence of S42 contained a single open reading frame (ORF[1]) and the sequences of the original primers used in the amplification. On Northern blots of *Limulus* testes, the S42 probe detected a message of ~3.3 kb consistent with a predicted message of at least 2.7–2.8 kb for a 102-kD protein (Fig. 1). Using S42 as a probe, we isolated four potential scruin clones (L1-4). Furthermore, Northern analysis with the largest clone L1 under very stringent conditions detects the same 3.2-kb message as S42 (data not shown). Fig. 2 shows the cDNA and derived protein sequences of L1 (EMBL accession No. Z38132). The identity of L1 as scruin is confirmed by several peptide sequences, obtained from native scruin isolated from the acrosomal process, that lie outside the S42 probe sequence (Fig. 2).

Scruin Contains a Duplicated Domain

Analysis of scruin cDNA shows it encodes a 918–amino acid residue protein with a predicted molecular mass of 103 kD, consistent with the mass of scruin determined by SDS-PAGE. Dot plot analysis of scruin indicates that there is a strong tandem repeat between the two halves of the molecule: alignments show the NH_2- and COOH-terminal halves are 32% identical over 390 amino acids. The presence of multiple equally spaced diagonal lines in dot plots indicates the homology within the molecule is largely derived from tandem repeats of a smaller ~50 residue sequence (data not shown). A closer examination of the scruin sequence reveals this 50-residue motif is present 12 times (repeats 1–12), which in turn are grouped in two blocks of six repeats (Figs. 3 and 4). Repeats 1–6 and 7–12 comprise most of the NH_2- and COOH-terminal halves of the molecule, respectively. Although the repeats are quite diverged, multiple alignment of all 12 repeat sequences identifies widely separated but conserved residues (Fig. 3). The pairwise identity between any two repeats varies between 11 and 42%. However, repeats from corresponding positions in the two halves of the molecule always show the highest degree of identity (23–42%). For example, repeats 2 and 8 are 42% identical, but the level of identity with any other repeat varies between 11 and 23%. The high similarity between repeats at corresponding positions in the two halves of the molecule is consistent with the idea that scruin has evolved from at least one gene duplication event.

1. *Abbreviations used in this paper*: MIPP, mouse IAP-promoted placental protein; ORF, open reading frame.

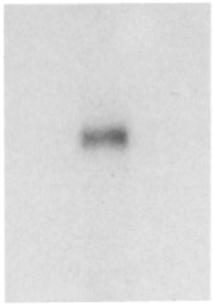

Figure 1. Northern blot analysis on 5 μg of poly A+ RNA with the S42 scruin probe detects a single relatively abundant message of ~3.3 kb in testes tissue. The full-length scruin clone L1 also detects the same message. The positions of RNA standards are indicated in kilobases.

The Repeat Motif in Scruin Is Widespread

Database searches show that the scruin has no sequence similarity to any known actin binding protein. However, the search identifies significant homology between the 50–amino acid residue repeat motif in scruin and a number of proteins including: the first open reading frame of the kelch gene in *Drosophila* (Xue and Cooley, 1993), mouse intracisternal A-particle–promoted placental protein (MIPP) (Chang-Yeh et al., 1991), a homologue of MIPP in *Caenorhabditis elegans* (Wilson et al., 1994), expressed sequence tags for kelch and MIPP in human brain (Adams et al., 1993*a*,*b*), a protein in *Arabidopsis* (GenBank accession No. X71915), and galactose oxidase (McPherson et al., 1992). However, the greatest number of protein sequences that contain the repeat motif are found several open reading frames of the poxvirus family, including A55R, F3L, C2L, and B10R in vaccinia virus (Goebel et al., 1990); D16L, C7L, J6R, and B20R in smallpox variola major virus (Massung et al., 1994); T6, T8, and T9R in shope fibroma virus (Upton et al., 1990); C4L and C13L in swinepox virus (Massung et al., 1993); and P65 in ectromelia virus (Senkevich et al., 1993) (Fig. 4). The strong conservation of these ORFs between the different genera of the poxvirus family would suggest an important role in virus function, for instance, in reproduction and/or virus host interactions. The number of individual repeat motifs is highly variable among this group of proteins, ranging from two repeats in B10R to seven repeats in galactose oxidase. However, in all cases that we have studied, the repeats are always grouped together consecutively. Indeed, scruin is unique in this respect, because it is the only example that contains two distinct sets of repeats (Fig. 4).

The repeat sequences identified in database searches are highly divergent from each other; thus, a meaningful multiple alignment of all available repeats is difficult to construct. However, an alignment between the repeats in scruin and kelch, whose repeats are more similar to each other than any other protein, shows a conserved pattern of identical or similar residues at fixed but widely spaced positions in the repeat. For example, a Gly-Gly sequence (shown in bold in Fig. 3) in the first third of each repeat is always preceded by a highly hydrophobic sequence that is often centered around a conserved tyrosine residue. The strict presence of this tandem glycine sequence together with at least 50% identity with the consensus sequence at other positions were used to define the number of repeats and their location in the schematic shown in Fig. 4.

Proteolysis Confirms the Domain Organization of Scruin

To investigate the domain organization of scruin biochemically, we treated purified acrosomal preparations with a vari-

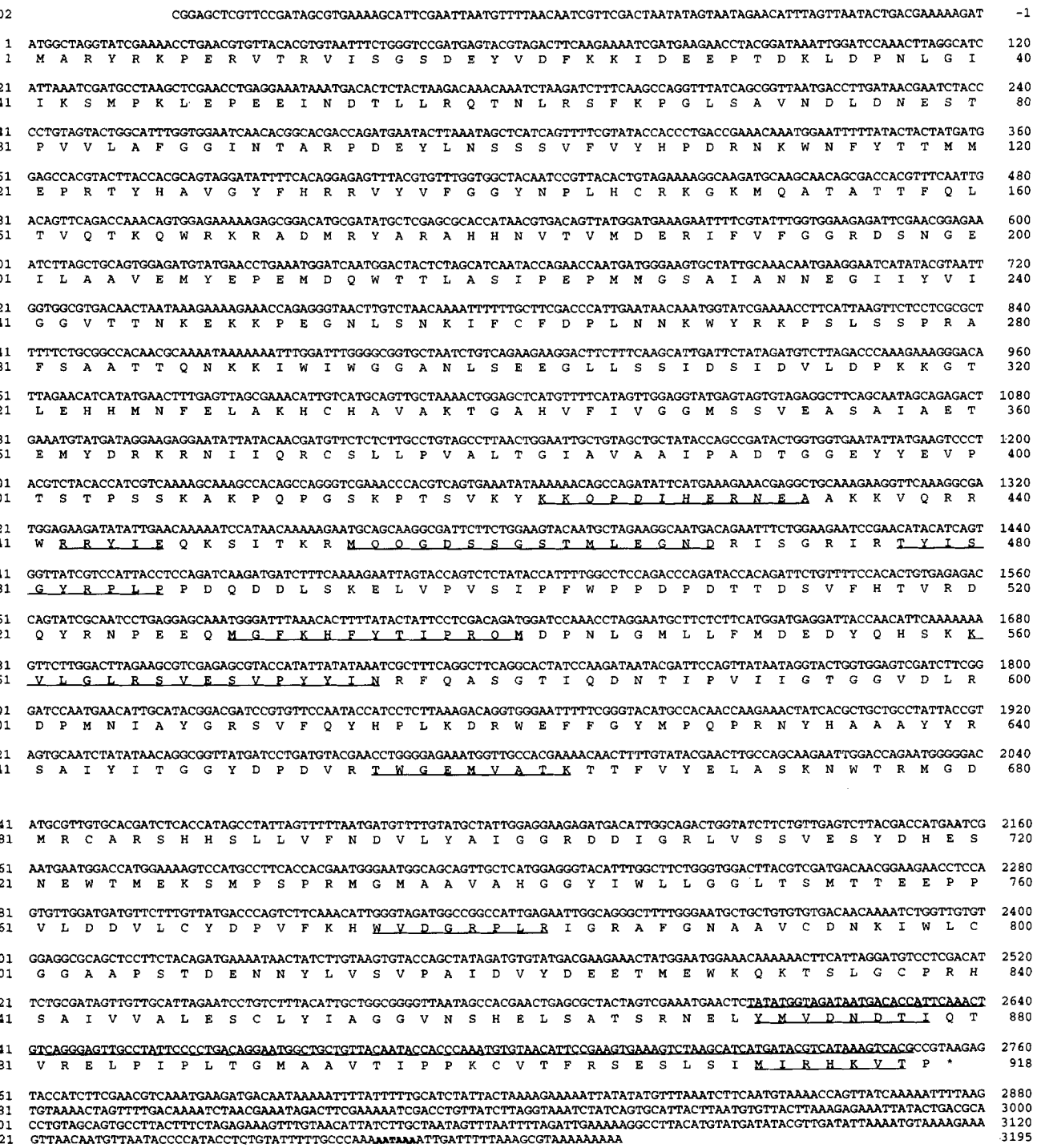

Figure 2. The nucleotide sequence of L1 together with the deduced amino acid sequence of scruin. Amino acid sequence obtained from NH_2-terminal sequencing of scruin peptides and the sequence at the extreme COOH terminus corresponding to the S42 probe are underlined. The polyadenylation signal at the 3′ end is shown in bold.

ety of proteases. A time course of a tryptic digest performed in EGTA or calcium shows scruin is cleaved into 50–60-kD fragments, while actin remains intact (Fig. 5 *A*). Similarly sized protease resistant domains were also observed when acrosomal bundles were digested with chymotrypsin, V8 protease, or Endoproteinase Asp (Endo-Asp) (data not shown). In all instances, scruin was more rapidly cleaved in EGTA than in calcium. We were able to obtain NH_2-terminal sequences from all the larger proteolytic scruin fragments but not from the smaller polypeptides (Fig. 5 *A*). Because the NH_2-terminus of intact scruin is blocked from sequencing by Edman degradation, we would assume the smaller polypeptides are derived from NH_2-terminal half of scruin. All NH_2-terminal sequences from the larger group of polypeptides mapped to a 5-kD region between the NH_2- and COOH-terminal repeats (Fig. 5 *B*). In addition, scruin

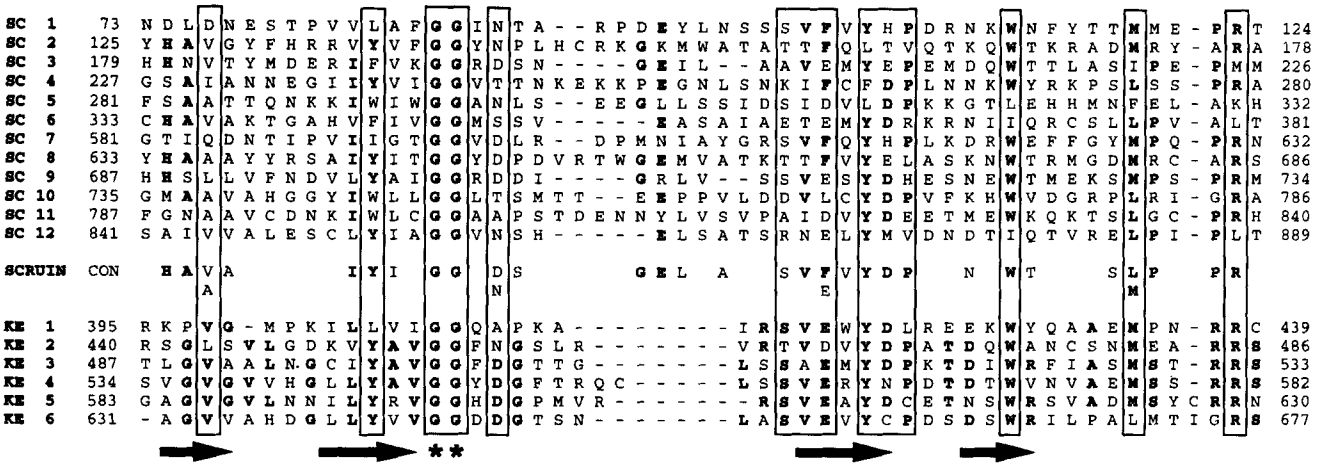

Figure 3. Alignment of the repeat sequences in scruin (*SC*) and kelch (*KE*) generated by the program MEGALIGN. The residue positions at the start and stop of each repeat are indicated for both proteins. The start position of the first repeat is based on the structural analysis of the repeat motif by Bork and Dolittle (1994). In the case of scruin, where at least 5 out of the 12 repeats have an identical residue, it is shown in bold, whereas the kelch residues are shown in bold when at least 3 of the 6 repeats have identity. In addition, the scruin consensus residues not shown in bold correspond to positions where 4 out of the 12 repeats show identity. Boxes around the residues in both scruin and kelch highlight positions in repeat sequences where there is strong conservation between the two proteins. The double asterisk indicates the double-glycine residue motif, and the four bold arrows, underneath the alignment, indicate the positions of the β strands in the putative structural fold of the repeat.

sometimes becomes degraded during preparation or storage of acrosomal bundles, especially when protease inhibitors are not included. NH_2-terminal sequencing of the larger product of this natural breakdown fragment of scruin also maps to the protease-sensitive region in the middle of the molecule (Fig. 5 *B*). To examine the effect of proteases on the integrity of the acrosomal bundle, we performed low speed sedimentation assays with undigested and digested preparations. In all cases, the digestion products remained associated with the actin, suggesting the bundle remained intact even though scruin had been cleaved in half (data not shown). Electron micrographs of negatively stained samples showed that protease-treated bundles were indistinguishable from untreated controls (Fig. 6).

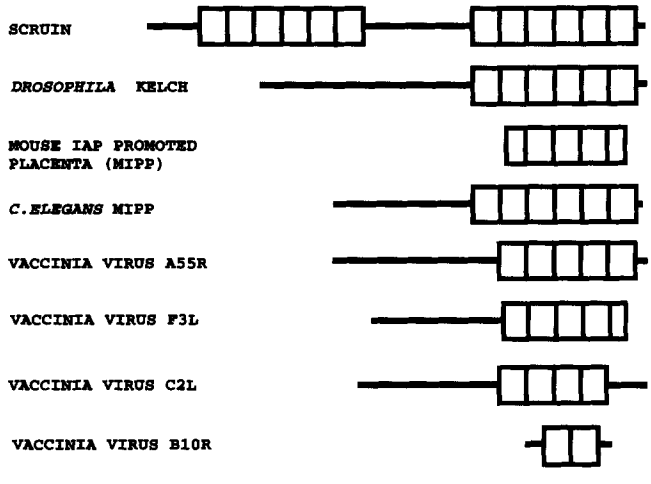

Figure 4. A schematic representation and names of proteins that contain multiple tandem repeats that show homology to scruin. For clarity, only the ORFs from vaccinia virus have been shown. All homologies are restricted to the ~50–amino acid residue repeat sequences that are represented by the open-boxed regions. The number and location of repeats was based on the strict presence of the double-glycine motif and the presence of at least half the other conserved residues in the scruin consensus. Outside the boxed regions there is no similarity between scruin and any of the proteins shown.

Expression of Scruin Domains

We tried to analyze the actin-binding activity of scruin through expression in *E. coli*. However, we were unable to engineer a pMW172 construct expressing full-length scruin because removal of the 5′ untranslated region of L1 appears to make the cDNA unstable. In all the strategies tried, we were only able to detect in transformed cells the truncated vector with no insert. Because clones corresponding to the NH_2 terminus of scruin could not be constructed, we expressed two COOH-terminal domains of scruin, 454C, and 590C. 454C corresponds to the natural breakdown product of scruin (Fig. 5), while 590C is a smaller COOH-terminal fragment lacking nine residues of the first repeat but continuing to the COOH terminus of scruin.

Both constructs expressed well, were easily purified in urea, and remained fully soluble after removal of the urea. Although neither protein showed signs of precipitation, both proteins became viscous like F-actin upon removal of urea. Electron microscopy showed that both recombinant proteins had assembled into short worm-like filaments (Fig. 7). The filaments were not similar in appearance to F-actin (Fig. 7). Further studies showed that filament formation was more pronounced in the presence of salt and once formed, they could not be dissociated by ionic or nonionic detergents. At higher protein concentrations or in the presence of calcium, the filaments tended to form amorphous aggregates rather than to elongate. We were unable to prevent the formation of filaments, even at extremely low protein concentrations, as judged by electron microscopy or from silver staining of

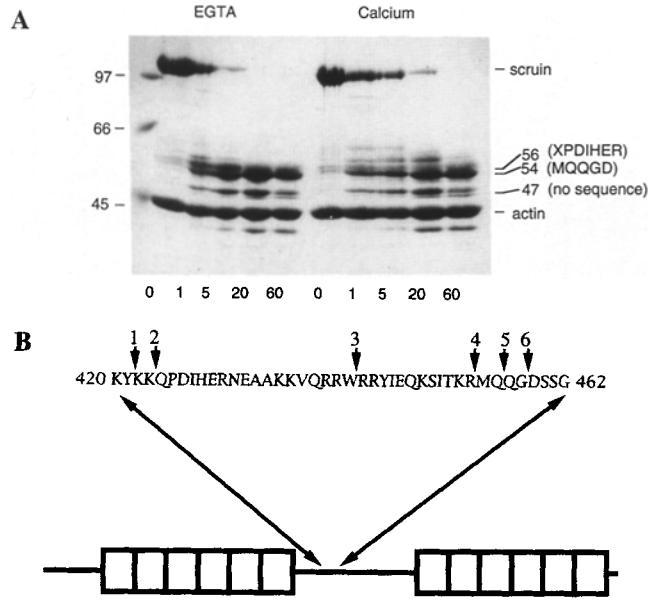

Figure 5. (*A*) A representative trypsin proteolysis of purified acrosomal bundles in calcium or EGTA. While actin remains intact, scruin is cleaved into two to four bands with molecular masses of ~47–57 kD. In calcium, although the final proteolytic pattern appears identical to that in EGTA, the rate of proteolysis is slower. Similar patterns and rates in calcium and EGTA were seen with chymotrypsin, V-8 protease, and Endo-Asp (not shown). The NH_2-terminal sequences and sizes of the higher molecular mass proteolytic fragments generated by trypsin are indicated. No sequence was obtained from the smaller 47-kD trypsin fragment, suggesting that this corresponds to the extreme NH_2-terminal half of the protein because scruin is blocked for sequencing. A similar pattern of larger sequenceable bands and smaller blocked bands was found with chymotrypsin, V-8 protease, and Endo-Asp. (*B*) A schematic representation of the scruin sequence together with the sequence of the protease sensitive region in the middle of the molecule. The numbered arrows indicate the positions of the cleavage sites for (*1*) chymotrypsin calcium; (*2*) trypsin calcium; (*3*) chymotrypsin EGTA; (*4*) trypsin EGTA; and (*55*) natural breakdown; (*6*) Endo-Asp based on NH_2-terminal sequencing of larger proteolytic fragments.

sedimentation assays. Because in control experiments 454C and 590C sedimented in the absence of actin, we were unable to study actin binding using sedimentation assays. We could not confidently determine by electron microscopy whether these domains decorated F-actin because of the appearance of tangled filaments.

Identification of the Actin Isoform in the Acrosomal Process

Northern analysis with a chicken β-actin probe detects three distinct messages in the range of 1.4–1.8 kb in *Limulus* testes tissue (data not shown). Using the same actin probe, we isolated and subsequently sequenced full-length clones (ACT3, ACT5, and ACT11) corresponding to these three actin messages (their EMBL accession numbers are Z38130, Z39131, and Z38129, respectively). Since the NH_2 terminus of actin in the acrosomal process is blocked, we obtained internal sequences from peptides generated by protease or cyanogen bromide cleavage. All six peptide sequences analyzed could only be derived from ACT5, suggesting this sequence represents the actin in the acrosomal process. Furthermore, the difference in mass of the peptide corresponding to residues 70–85 of the ACT5 sequence measured by MALD MS and the predicted mass from the cDNA sequence suggest histidine 74 is methylated.

Discussion

First identified in 1975 by Tilney, scruin is an unusual case in which we know more about its three-dimensional structure than about its biochemical properties (Tilney, 1975). This situation has been largely caused by the high degree of crystalline order of the actin bundle and our inability to purify scruin as a soluble protein. In helical reconstructions, scruin appears as two globular domains, one spherical and one elongated, joined by linker or neck region (Owen and DeRosier, 1993; Schmid et al., 1994). Now, based on the scruin sequence, we can confidently assume that each of these two globular domains is composed of a sixfold repeat domain. In addition, our map of protease cleavage sites show the sixfold repeat-containing domains are protease resistant, as would be expected of a compactly folded domain, while the neck region between the repeat domains is extremely protease sensitive. Thus, both the sequence and protease map of scruin, together with the three-dimensional helical reconstructions, provide a more detailed model of scruin.

In the three-dimensional reconstruction of the acrosomal filament, the spherical and elongated domains of scruin overlay a helix-loop-β strand motif on the surface of subdomain 1 of one monomer and subdomain 3 of the monomer across the filament (Schmid et al., 1994). Because the helix-loop-β strand motif contributes most of the residues available for contact on the surface of subdomains 1 and 3 of actin, Schmid and colleagues concluded that this motif must be a scruin binding site on each actin subdomain. If this conclusion is correct, then the two homologous domains in scruin bind structurally homologous but not identical sites on separate actin subunits.

This finding immediately challenges a common assumption about actin-binding sites for families of actin-binding proteins. Normally, we would assume that homologous actin-binding domains bind identical positions on actin. For example, the homologous pair of actin-binding domains in fimbrin or the actin-binding domains of α-actinin and dystrophin bind identical sites or subdomains of actin. Clearly this assumption is not true for scruin. The position of the actin-binding domain of α-actinin determined by image reconstruction agrees with genetic studies of yeast fimbrin (Sac6p) and both studies map the actin-binding sites to residues in actin subdomains 1 and 2 (Holtzman et al., 1994; Honts et al., 1994; McGough et al., 1994). However, it has not been proven that both actin-binding domains of Sac6p bind to this site on actin. Indeed, the two actin-binding domains of Sac6p are only 26% identical (Adams et al., 1991) compared to the 32% identity that we see between the NH_2- and COOH-terminal halves of scruin.

Inspection of the helix-loop-β strand motif in the three-dimensional structure of vertebrate actin reveals a subset of hydrophobic residues that are oriented toward the interior and a group of hydrophilic residues that are exposed at the surface, and thus capable of interacting with scruin (Schmid

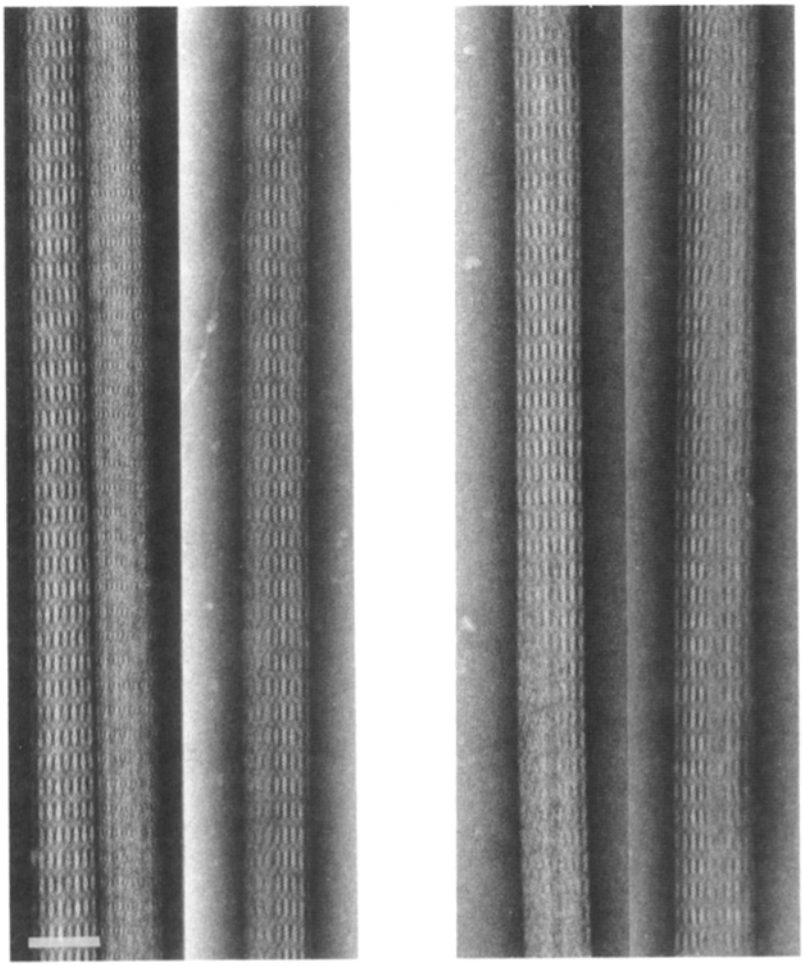

Figure 6. Electron micrographs of negatively stained purified acrosomal bundles (*CONTROL*) and bundles digested with trypsin in the presence of calcium (*DIGESTED*). The samples shown correspond to the 0- and 60-min time points in Fig. 5. Bar, 100 nm.

et al., 1994). Although actin sequences are highly conserved, the sequence of ACT5 confirms that the actin in the acrosomal process of *Limulus* has the helix-loop-β strand motif. Comparison of the protein sequence of ACT5 and rabbit actin shows they are 92.1% identical and that there is only a single substitution in each of the two helix-loop-β strand motifs. This high level of sequence conservation validates the use of available structures of rabbit skeletal actin in reconstructions of the scruin actin filament complex.

The ability of scruin to bind structurally homologous motifs in two different actin monomers immediately suggests the existence of two complementary and structurally homologous actin-binding sites in scruin. The overall organization of scruin is consistent with this, the protein is divided into homologous NH_2- and COOH-terminal domains, residues 18–390 and 526–898, respectively. However, the internal sixfold repeats, residues 73–381 and 581–889, make it difficult to predict the location of the actin-binding sites by inspection of the scruin sequence alone. We attempted to address this question through expression of scruin domains in *E. coli* as a first step in the identification of the actin-binding sites in scruin. However, DNA instability prevented us from engineering any expression construct containing the NH_2-terminus of scruin. By contrast, we were able to construct and express the COOH-terminal half of scruin (454C). However, while 454C was soluble, it readily assembled into filaments in the absence of actin, even at low protein concentrations and prevented analysis of actin-binding by sedimentation or electron microscopy. Similar results were seen with 590C, which essentially corresponds to the sixfold repeat domain in the COOH-terminal half of the molecule.

We feel that the strong self-association property of the scruin domains may reflect the extreme stability of scruin cross-links in the actin bundle. Unlike the situation with other actin cross-linking proteins where the functional cross-linking unit contacts two different actin filaments, reconstructions show each scruin molecule contacts a single filament. Thus, scruin must form a dimer to cross-link filaments. The stability of the acrosomal process must then reflect the extensive scruin–scruin contacts in the hexagonally packed bundle, as well as the scruin–actin contacts. The large number of stabilizing protein contacts within the bundle accounts for why both scruin domains generated by proteolysis remain associated with actin during low speed sedimentation assays (data not shown) and why protease-treated acrosomal processes appear identical to untreated samples in the electron microscope (Fig. 6). It may be that the self assembly of expressed scruin domains might be explained in part by the scruin–scruin interactions that naturally occur in the acrosomal process.

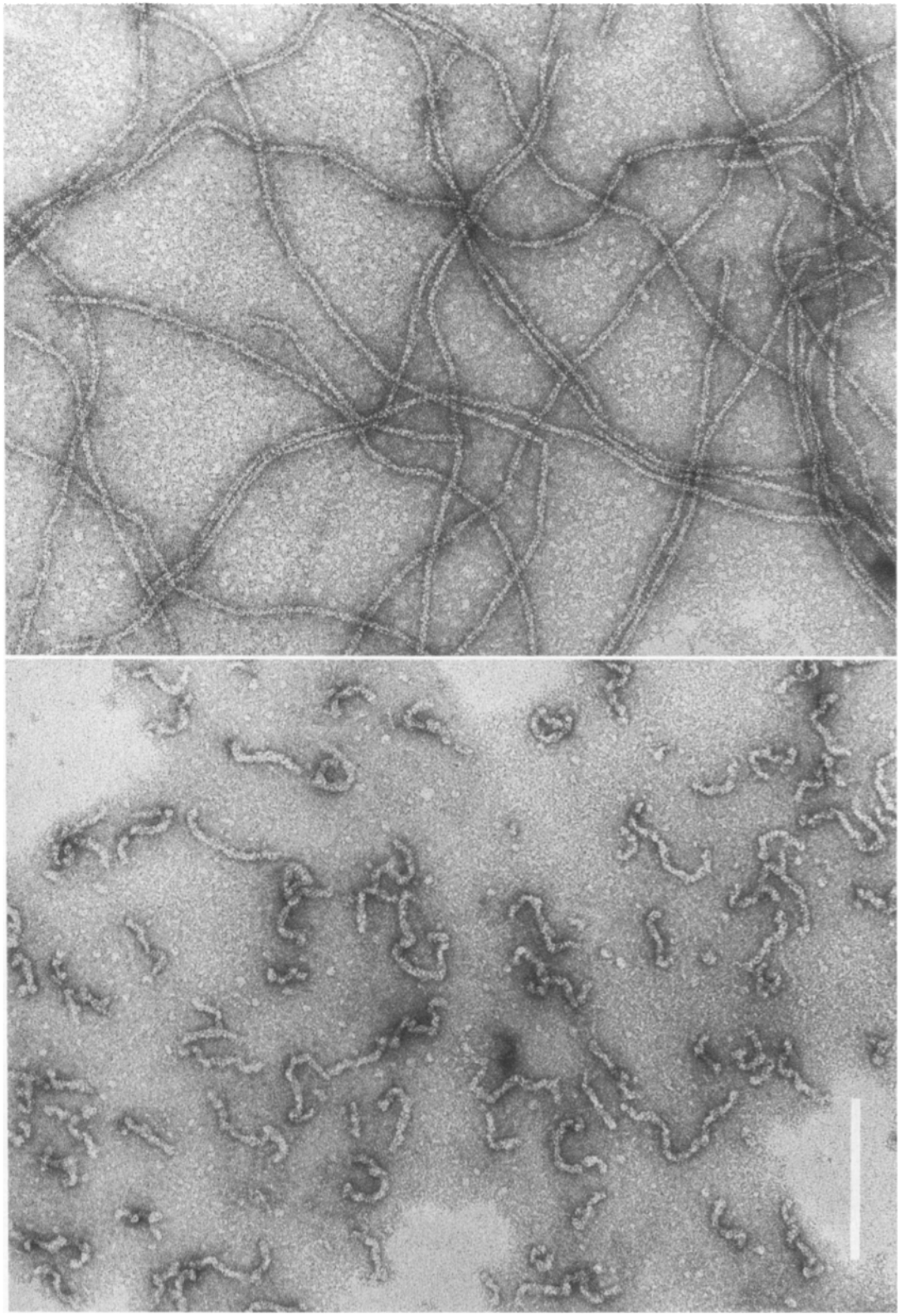

Figure 7. A typical field of negatively stained 590C "worms" (590C) and control actin filaments for comparison (*ACTIN*). Bar, 40 nm.

The Repeat Motif in Scruin Is a Common Structural Fold

A consensus sequence derived from all 12 repeats of scruin shows 36% identity with a consensus sequence derived from the repeats in kelch, suggesting the repeat in both proteins will share a similar fold. Recent sequence analysis by Bork and Doolittle (1994) has revealed that the six repeats in kelch represent a conserved structural fold found in the crystal structure of a number of enzymes, including the catalytic domain of galactose oxidase from *Dactylium dendroides* (Ito et al., 1994). This domain is composed of a single structural motif, consisting of four antiparallel β strands joined by loops of variable sizes, repeated seven times like the petals of a flower around a central channel. The conserved residues in the consensus repeat sequences of kelch and scruin map within the β strands, suggesting that the repeats in both these proteins are also organized around a similar four-stranded antiparallel β-sheet motif.

The circular arrangement of the β-strand motif is a well-known structure called a β propeller or "superbarrel" (Murzin, 1992; Chothia and Murzin, 1993). This structure is found in several proteins that have no sequence similarity including a number of fungal, bacterial, and viral enzymes, including influenza neuraminidase (sialidase) (Bork and Doolittle, 1994; Crennell et al., 1993; Ito et al., 1994; Varghese et al., 1983). The number of repeat motifs in the superbarrel structure varies between 6 and 8, depending on the protein. We assume that the three-dimensional structures of the sixfold repeat domains in kelch and scruin most closely resemble the sialidase family because the superbarrel domain in this class of proteins contains six repeats (Crennell et al., 1993; Varghese et al., 1983).

Is the Sixfold Repeat Sequence an Actin-binding Domain?

Kelch is closely associated with actin in the ring canals of *Drosophila* (Xue and Cooley, 1993). Ring canals are cytoplasmic bridges that connect the 15 nurse cells to the developing oocyte. Formed by incomplete cytokinesis during cell division and lined by a ring of actin filaments, ring canals are a common feature in germ cell development in a variety of species (Xue and Cooley, 1993). Although there is no direct evidence that kelch binds actin, it is tempting to speculate, given the homology with scruin, that the sixfold repeat domain in kelch is responsible for localizing the protein to the actin in the ring canal.

The role of the large number of ORFs containing the repeat motif in the poxviruses is currently unknown. However, a number of studies have shown that poxviruses have several effects on the actin cytoskeleton during their life cycle (Hiller et al., 1979, 1981; Krempien et al., 1981; Meyer et al., 1981). The most spectacular of which is the virally induced assembly of large microvilli at the cell surface in the later stages of the infection cycle (Hiller et al., 1979; Krempien et al., 1981). Furthermore, electron microscope examination shows virions at the tips of actin bundles in these large microvillar structures (Stokes, 1976). The relationship between the vaccinia life cycle and the actin cytoskeleton suggests that the virus encodes proteins that modify the assembly state of actin. The presence of a viral profilin, an actin-binding protein known to modulate f-actin assembly, in the genome of vaccinia offers one means to alter the actin cytoskeleton during infection (Blasco et al., 1991). However, recombinant virus lacking the profilin gene has identical effects on the actin cytoskeleton as wild-type virus (Blasco et al., 1991), suggesting that other as yet unidentified actin proteins may be encoded in the virus genome.

The sequence of scruin presented here shows that it is a member of a widespread family of proteins that contain a common repeat motif domain. However, scruin is unique in this family, being the only example that is known to bind directly to actin. Further characterization of the other members of the family will be required to understand the function of this structural domain and to confirm whether they are also capable of binding actin.

We would like to thank Dr. Dan Gibson for his expertise on *Limulus* physiology; Drs. Brian Pope, Alan Weeds, and Steve Kron for discussions concerning the properties of expressed scruin domains; Dr. Sharon Bickel for the great autoclave filter denaturing method; Mark Chafel, Ya-Huei Tu, Sean Britt, and Mike Burgess for technical assistance and oligonucleotide synthesis; and finally Navin Pokala, Anne Williamson, as well as Drs. Walter Gratzer, Amy McGough, David DeRosier, Simon Atkinson, and Michael Schmid for comments and critically reading this manuscript. M. Way was supported by a North Atlantic Treaty Organization–Science Engineering Research Council fellowship. P. Matsudaira was supported by the National Institutes of Health (DK35306 and CA 44704).

Received for publication 6 September 1994 and 13 October 1994.

References

Adams, A. E., D. Botstein, and D. G. Drubin. 1991. Requirement of yeast fimbrin for actin organization and morphogenesis in vivo. *Nature (Lond.).* 354:404–408.

Adams, M. D., A. R. Kerlavage, C. Fields, and J. C. Venter. 1993*a*. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* 4:256–267.

Adams, M. D., M. B. Soares, A. R. Kerlavage, C. Fields, and J. C. Venter. 1993*b*. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genet.* 4:373–380.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

Bankier, A. T., K. M. Weston, and B. G. Barrell. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol.* 155:51–93.

Blasco, R., N. B. Cole, and B. Moss. 1991. Sequence analysis, expression, and deletion of a vaccinia virus gene encoding a homolog of profilin, a eukaryotic actin-binding protein. *J. Virol.* 65:4598–4608.

Bork, P., and R. F. Doolittle. 1994. *Drosophila kelch* motif is derived from a common enzyme fold. *J. Mol. Biol.* 236:1277–1282.

Bullitt, E. S., D. J. DeRosier, L. M. Coluccio, and L. G. Tilney. 1988. Three-dimensional reconstruction of an actin bundle. *J. Cell Biol.* 107:597–611.

Chang-Yeh, A., D. E. Mold, and R. C. Huang. 1991. Identification of a novel murine IAP-promoted placenta-expressed gene. *Nucleic Acids Res.* 19: 3667–3672.

Chothia, C., and A. G. Murzin. 1993. New folds for all β proteins. *Structure.* 1:217–222.

Crennell, S. J., E. F. Garman, W. G. Laver, E. R. Vimr, and G. L. Taylor. 1993. Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. USA.* 90:9852–9856.

DeRosier, D. J., L. G. Tilney, E. M. Bonder, and P. Frankl. 1982. A change in twist of actin provides the force for the extension of the acrosomal process in *Limulus* sperm: the false-discharge reaction. *J. Cell Biol.* 93:324–337.

Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, and E. Paoletti. 1990. The complete DNA sequence of vaccinia virus. *Virology.* 179:247–266.

Hiller, G., C. Jungwirth, and K. Weber. 1981. Fluorescence microscopical analysis of the life cycle of vaccinia virus in chick embryo fibroblasts. Virus-cytoskeleton interactions. *Exp. Cell Res.* 132:81–87.

Hiller, G., K. Weber, L. Schneider, C. Parajsz, and C. Jungwirth. 1979. Interaction of assembled progeny pox viruses with the cellular cytoskeleton. *Virology.* 98:142–153.

Holmes, K. C., D. Popp, W. Gebhard, and W. Kabsch. 1990. Atomic model of the actin filament. *Nature (Lond.).* 347:448–449.

Holtzman, D. A., K. F. Wertman, and D. G. Drubin. 1994. Mapping actin surfaces required for functional interactions in vivo. *J. Cell Biol.* 126:423–432.

Honts, J. E., T. S. Sandrock, S. M. Brower, J. L. O'Dell, and A. E. M. Adams. 1994. Actin mutations that suppress and are suppressed by fimbrin mutations identify a likely fimbrin-binding site on actin. *J. Cell Biol.* 126:413–422.

Ito, N., S. E. V. Phillips, K. D. S. Yadav, and P. F. Knowles. 1994. Crystal structure of a free radical enzyme, galactose oxidase. *J. Mol. Biol.* 238:794–814.

Kabsch, W., H. G. Mannherz, D. Suck, E. F. Pai, and K. C. Holmes. 1990. Atomic structure of the actin:DNase I complex. *Nature (Lond.).* 347:37–44.

Krempien, U., L. Schneider, G. Hiller, K. Weber, E. Katz, and C. Jungwirth. 1981. Conditions for pox virus-specific microvilli formation studied during synchronized virus assembly. *Virology.* 113:556–564.

Lehman, W., R. Craig, and P. Vibert. 1994. $Ca^{(2+)}$-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction. *Nature (Lond.).* 368:65–67.

Lorenz, M., D. Popp, and K. C. Holmes. 1993. Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J. Mol. Biol.* 234:826–836.

Massung, R. F., V. Jayarama, and R. W. Moyer. 1993. DNA sequence analysis of conserved and unique regions of swinepox virus: identification of genetic elements supporting phenotypic observations including a novel G protein-coupled receptor homologue. *Virology.* 197:511–528.

Massung, R. F., L. I. Liu, J. Qi, J. C. Knight, T. E. Yuran, A. R. Kerlavage, J. M. Parsons, J. C. Venter, and J. J. Esposito. 1994. Analysis of the complete genome of smallpox variola major virus strain Bangladesh-1975. *Virology.* 201:215–240.

Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262: 10035–10038.

Matsudaira, P. T., and D. R. Burgess. 1978. SDS microslab linear gradient polyacrylamide gel electrophoresis. *Anal. Biochem.* 87:386–396.

McGough, A., M. Way, and D. DeRosier. 1994. Determination of the alpha-actinin binding site on actin filaments by cryo-electron microscopy and image analysis. *J. Cell Biol.* 126:433–443.

McLaughlin, P. J., J. T. Gooch, H. G. Mannherz, and A. G. Weeds. 1993. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing [see comments]. *Nature (Lond.).* 364:685–692.

McPherson, M. J., Z. B. Ogel, C. Stevens, K. D. Yadav, J. Keen, and P. F. Knowles. 1992. Galactose oxidase of *Dactylium dendroides*. Gene cloning and sequence analysis. *J. Biol. Chem.* 267:8146–8152.

Meyer, R. K., M. M. Burger, R. Tschannen, and R. Schafer. 1981. Actin filament bundles in vaccinia virus infected fibroblasts. *Arch. Virol.* 67:11–18.

Milligan, R. A., M. Whittaker, and D. Safer. 1990. Molecular structure of F-actin and location of surface binding sites. *Nature (Lond.).* 348:217–221.

Murzin, A. G. 1992. Structural principles for the propeller assembly of beta-sheets: the preference for seven-fold symmetry. *Proteins.* 14:191–201.

Owen, C., and D. De Rosier. 1993. A 13-A map of the actin-scruin filament from the limulus acrosomal process. *J. Cell Biol.* 123:337–344.

Rayment, I., H. M. Holden, M. Whittaker, C. B. Yohn, M. Lorenz, K. C. Holmes, and R. A. Milligan. 1993. Structure of the actin-myosin complex and its implications for muscle contraction. *Science (Wash. DC).* 261:58–65.

Schmid, M. F., J. M. Agris, J. Jakana, P. Matsudaira, and W. Chiu. 1994. Three-dimensional structure of a single filament in the *Limulus* acrosomal bundle: scruin binds to homologous helix-loop-beta motifs in actin. *J. Cell Biol.* 124:341–350.

Schmid, M. F., J. Jakana, P. Matsudaira, and W. Chiu. 1993. Imaging frozen, hydrated acrosomal bundle from *Limulus* sperm at 7 Å resolution with a 400 kV electron cryomicroscope. *J. Mol. Biol.* 203:384–386.

Schmid, M. F., P. Matsudaira, T. W. Jeng, J. Jakana, E. Towns-Andrews, J. Bordas, and W. Chiu. 1991. Crystallographic analysis of acrosomal bundle from *Limulus* sperm. *J. Mol. Biol.* 221:711–725.

Schroder, R. R., D. J. Manstein, W. Jahn, H. Holden, I. Rayment, K. C. Holmes, and J. A. Spudich. 1993. Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1. *Nature (Lond.).* 364:171–174.

Schutt, C. E., J. C. Myslik, M. D. Rozycki, N. C. Goonesekere, and U. Lindberg. 1993. The structure of crystalline profilin-beta-actin. *Nature (Lond.).* 365:810–816.

Senkevich, T. G., G. L. Muravnik, S. G. Pozdnyakov, V. E. Chizhikov, O. I. Ryazankina, S. N. Shchelkunov, E. V. Koonin, and V. I. Chernos. 1993. Nucleotide sequence of XhoI O fragment of ectromelia virus DNA reveals significant differences from vaccinia virus. *Virus Res.* 30:73–88.

Stokes, G. V. 1976. High-voltage electron microscope study of the release of vaccinia virus from whole cells. *J. Virol.* 18:636–643.

Tilney, L. G. 1975. Actin filaments in the acrosomal reaction of *Limulus* sperm. Motion generated by alterations in the packing of the filaments. *J. Cell Biol.* 64:289–310.

Upton, C., J. L. Macen, D. S. Wishart, and G. McFadden. 1990. Myxoma virus and malignant rabbit fibroma virus encode a serpin-like protein important for virus virulence. *Virology.* 179:618–631.

Varghese, J. N., W. G. Laver, and P. M. Colman. 1983. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature (Lond.).* 303:35–40.

Way, M., B. Pope, J. Gooch, M. Hawkins, and A. G. Weeds. 1990. Identification of a region in segment 1 of gelsolin critical for actin binding. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:4103–4109.

Way, M., B. Pope, and A. G. Weeds. 1992. Evidence for functional homology in the F-actin binding domains of gelsolin and alpha-actinin: implications for the requirements of severing and capping. *J. Cell Biol.* 119:835–842.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature (Lond.).* 368:32–38.

Xue, F., and L. Cooley. 1993. Kelch encodes a component of intercellular bridges in *Drosophila* egg chambers. *Cell.* 72:681–693.