

# SF-Assemblin, the Structural Protein of the 2-nm Filaments from Striated Microtubule Associated Fibers of Algal Flagellar Roots, Forms a Segmented Coiled Coil

Klaus Weber,\* Norbert Geisler,\* Uwe Plessmann,\* Angela Bremerich,‡ Karl-Ferdinand Lechtreck,‡ and Michael Melkonian‡

\*Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, D-3400 Goettingen, Federal Republic of Germany; and ‡Universität zu Köln, Botanisches Institut, Gyrhofstrasse 15, D-5000 Köln 41, Federal Republic of Germany

**Abstract.** The microtubule associated system I fibers of the basal apparatus of the flagellate green alga *Spermatozopsis similis* are noncontractile and display a 28-nm periodicity. Paracrystals with similar periodicities are formed in vitro by SF-assemblin, which is the major protein component of system I fibers. We have determined the amino acid sequence of SF-assemblin and show that it contains two structural domains. The NH<sub>2</sub>-terminal 31 residues form a nonhelical domain rich in proline. The rod domain of 253 residues is  $\alpha$ -helical and seems to form a segmented coiled coil with a 29-residue repeat pattern based on four heptads followed by a skip residue. The distinct cluster of acidic residues at the COOH-terminal end of the motifs (periodicity about 4 nm) may be related to tubulin binding of SF-assemblin and/or its self assembly. A

similar structure has been predicted from cDNA cloning of  $\beta$ -giardin, a protein of the complex microtubular apparatus of the sucking disc in the protozoan flagellate *Giardia lamblia*. Although the rod domains of SF-assemblin and  $\beta$ -giardin share only 20% sequence identity, they have exactly the same length and display 42% sequence similarity. These results predict that system I fibers and related microtubule associated structures arise from molecules able to form a special segmented coiled coil which can pack into 2-nm filaments. Such molecules seem subject to a strong evolutionary drift in sequence but not in sequence principles and length. This conservation of molecular architecture may have important implications for microtubule binding.

THE basal apparatus of flagellate green algae is the functional homologue of the centriolar complex of animal cells (for review see Wheatley, 1982; Kalnins, 1992). It contains among other structures the basal bodies, usually four sets of attached flagellar root microtubules and two additional types of filamentous structures which have been called system I and system II fibers (for reviews see Melkonian, 1980; Lechtreck and Melkonian, 1991b). System II fibers consist of 4–8-nm filaments and contain centrin, a Ca<sup>2+</sup>-binding protein highly homologous to calmodulin (Salisbury et al., 1984, 1988; Huang et al., 1988). Centrin filaments connect the basal bodies with each other and with the nucleus and can also occur within the flagellar transition region (for review see Salisbury, 1989). These centrin based fibers seem responsible for various motile functions of the basal apparatus (Melkonian et al., 1992). In contrast, the system I fibers are not contractile. They are associated with

the flagellar root microtubules and thus probably form a stabilizing element of the basal apparatus. The system I fibers consist of layers of fine 2-nm filaments and show a complex cross-striation pattern with a 28-nm periodicity (Lechtreck and Melkonian, 1991a). Recent electron microscopical and image reconstruction studies favor a model in which system I fibers are built from rod-like molecules, 48 nm in length, in a staggered arrangement with identical polarities (Patel et al., 1992).

The flagellate green alga *Spermatozopsis similis* has been used to obtain homogeneous preparations of basal apparatuses. These yield by mechanical disintegration and high salt extraction the system I fibers. The major structural protein of the microtubule associated fibers has an apparent polypeptide molecular weight of 34,000 and has been called SF-assemblin (SF for striated fiber).<sup>1</sup> SF-assemblin solubilized by 2 M urea forms upon dialysis paracrystals with a staining pattern resembling that of the original system I fibers (Lechtreck and Melkonian, 1991a; Patel et al., 1992). To un-

Please address correspondence to Dr. Klaus Weber, Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, P.O. Box 2841, D-3400 Goettingen, FRG.

1. *Abbreviation used in this paper:* SF-assemblin, striated fiber-assemblin.

derstand this self-assembly process and to gain insight into the structure of the system I fiber, we have determined the amino acid sequence of SF-assemblin and analyzed its secondary structure by circular dichroism spectroscopy. The protein is rich in  $\alpha$ -helix and can give rise to a segmented coiled coil which shows a 29-residue repeat arising from a skip residue which is inserted after every four heptads. We also relate our results to recent electron microscopical and image reconstruction studies on system I fibers (Patel et al., 1992).

## Materials and Methods

### Purification of SF-Assemblin

The procedures for culturing of *Spermatozopsis similis* have been described (Lechtreck and Melkonian, 1991a). Cells were concentrated with a tangential flow filtration system (Millipore Corp., Bedford, MA) followed by centrifugation at 600 g (Sorvall, Dupont RC 28S; rotor GS3). Cells were washed once in MT buffer (30 mM Hepes, 5 mM EGTA, 15 mM KCl, 0.1 mM DTT, pH 7) and lysed by addition of an equal volume of MT buffer containing 2–3% Triton X-100. The resulting cytoskeletons were harvested at 1,500 g (15 min) and washed several times with MT buffer with decreasing concentrations of Triton X-100 (1, 0.5, 0.25%) and twice with MT buffer. Pellets were stored at  $-20^{\circ}\text{C}$ . Isolated cytoskeletons were extracted with 2 M NaCl in MT buffer for 3–5 h at  $4^{\circ}\text{C}$  with agitation, and then centrifuged at 48,500 g for 30 min at  $4^{\circ}\text{C}$ . The resulting pellet was extracted with MT buffer containing 2.5 M urea (disassembly buffer) for 3–5 h at  $4^{\circ}\text{C}$  and centrifuged at 200,000 g for 60 min at  $4^{\circ}\text{C}$ . The harvested supernatant fraction was dialyzed for 8–16 h at  $4^{\circ}\text{C}$  against reassembly buffer (150 mM KCl, 10 mM MES, 2 mM EDTA, 0.1–1 mM DTT, pH 6.25). The paracrystals formed during this reassembly step were harvested by centrifugation (48,500 g, 30 min,  $4^{\circ}\text{C}$ ). For further purification, once reconstituted paracrystals were dissolved in a small amount of disassembly buffer, centrifuged, and again dialyzed against reassembly buffer. Twice recycled paracrystals (for reviews see Lechtreck and Melkonian, 1991a) were used for the protein sequence work.

Alternative final purification steps of SF-assemblin started with the reconstituted paracrystals. These were dissolved in 8 M urea, 20 mM Tris-HCl, 1 mM DTT, pH 7.8, and the solution applied to a Mono Q column (HR 5/5, Pharmacia-LKB, Uppsala, Sweden) equilibrated with the urea buffer. The column was eluted with a linear salt gradient (0–300 mM KCl) in the same solvent. Fractions containing SF-assemblin were identified by SDS-PAGE. Dialysis against reassembly buffer yielded again paracrystals. In some experiments, once reconstituted paracrystals were dissolved in 8 M urea, 20 mM Tris-HCl, 50 mM KCl, 1 mM DTT, pH 7.8, and SF-assemblin chromatographed by gel filtration in the same solvent on a Hiload 26/60 Superdex 200 prep grade column (Pharmacia-LKB). SF-assemblin eluted at a position corresponding to a fully denatured protein of molecular weight 30,000 when the column was calibrated with standard proteins in the 8-M urea buffer. Fractions containing SF-assemblin were dialyzed against reassembly buffer to remove the urea. Paracrystals were harvested by centrifugation (48,500 g for 30 min at  $4^{\circ}\text{C}$ ). When desired, the twice reconstituted paracrystals were dissolved in 8 M urea and subjected to anion exchange chromatography on Mono Q as described. Purity of SF-assemblin was monitored by SDS-PAGE.

### Peptide Mapping of SF-Assemblin by Limited Proteolysis

Peptide mapping with endoproteinase Glu-C (Sigma Chemie, Deisenhofen, FRG) followed the procedure of Cleveland et al. (1977) with minor modifications. Bands of interest were cut with a razor blade from the gels stained with Coomassie brilliant blue. They were washed for 1 h with several changes of distilled water and incubated in SDS sample buffer with or without reducing agent (see also below for oxidative cross-linking). Gel pieces were put on top of the stacking gel and overlaid with 2.5–20 ng of endoproteinase Glu-C in SDS sample buffer. Subsequent electrophoresis was interrupted for 10–20 min once the running front had reached the separation gel, and then continued normally (Cleveland et al., 1977). For  $\text{NH}_2$ -terminal sequencing of proteolytic products, the resulting fragments were blotted on PVDF membrane.

### Sequence Analysis

Preliminary experiments on SF-assemblin transferred from SDS-PAGE to a PVDF membrane for  $\text{NH}_2$ -terminal sequencing or for digestion with trypsin or endoproteinase Asp-N used the procedures of Bauw et al. (1989). The peptides released by digestion were subjected to HPLC and the profiles screened by sequencing. Gas phase sequencing was on an Applied Biosystems sequenator (model A470) and a Knauer model 810. Both instruments operated with an on-line PTH amino acid analyzer. In addition, SF-assemblin present in a gel band was treated with endoproteinase Glu-C, yielding two fragments of molecular weights 20,000 and 12,000, respectively (Fig. 1). The fragments were blotted on PVDF and subjected to sequencing.

SF-assemblin was dissolved in 6 M guanidine-HCl and reduced with 2-mercaptoethanol. An excess of 4-vinylpyridine was used to alkylate cysteine residues. Alkylated SF-assemblin was recovered by chloroform/methanol precipitation (Wessel and Flügge, 1984). Two thirds of the material ( $\sim 70\ \mu\text{g}$ ) were used for CNBr cleavage. Fragments obtained by HPLC were characterized by sequencing. Larger fragments and mixtures were subjected to proteolytic cleavage. Endoproteinase Asp-N, Lys-C, and Arg-C were from Boehringer-Mannheim, FRG. Secondary fragments obtained by HPLC were subjected to extensive sequencing. One third of the alkylated SF-assemblin ( $\sim 35\ \mu\text{g}$ ) was subjected to digestion with endoproteinase Lys-C in 4 M urea followed by HPLC separation and sequencing of interesting fragments. Fourier analysis of periodicities in the SF-assemblin sequence was done as described (Wiche et al., 1991).

### Circular Dichroism

Measurements were performed with a Jasco 720 spectropolarimeter (Japan Spectroscopic Co., Ltd., Tokyo, Japan) using a cuvette of 0.1 cm pathlength. SF-assemblin, purified by Mono Q-chromatography in 8 M urea, was dialyzed against reassembly buffer to remove the urea. Paracrystals were harvested (see above) and dissolved in 50 mM sodium phosphate buffer, pH 7.8, supplemented with 2 M urea (ultra pure; Life Technologies, Inc., Gaithersburg, MD), and 0.2 mM DTT at  $\sim 0.13\ \text{mg/ml}$ . Final protein concentrations were obtained by quantitative amino acid analysis. Due to the presence of urea data below 200 nm could not be obtained.

### Electron Microscopy

Paracrystals (4  $\mu\text{l}$  in reassembly buffer) were applied to pioloform-coated copper grids and allowed to adhere for about 2 min. Grids were stained with 1% uranyl acetate in distilled water for 90 s to 3 min, and then observed with a Philips CM12 electron microscope. Protofilaments were obtained by dissolving the paracrystals in 50 mM sodium phosphate buffer, pH 7.8, containing 2 M urea. They were mixed with 30% glycerol by volume, sprayed on mica, and shadowed with tantalum/tungsten at an angle of  $9^{\circ}\text{C}$ .

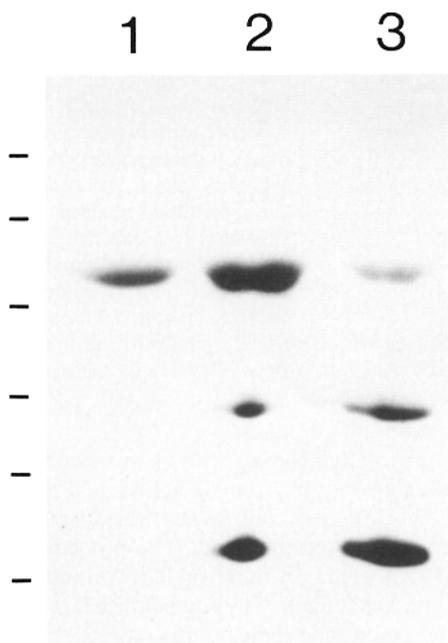
### Oxidative Crosslinking of SF-Assemblin

Paracrystals formed by SF-assemblin were dissolved in MT buffer, 2 or 4 M in urea, which lacked a reducing agent. After extensive dialysis against the same solvents, SDS-PAGE was performed with a sample buffer lacking 2-mercaptoethanol. In some experiments when samples with or without reducing agent were analyzed in parallel on the same gel, samples containing reducing agent were treated before electrophoresis with an excess of iodoacetamide to alkylate all free sulfhydryl groups. Peptide mapping by limited proteolysis with endoproteinase Glu-C of oxidized SF-assemblin was performed essentially as described above, except that all solutions lacked reducing agents. In some experiments oxidized SF-assemblin or its fragments containing the disulfide bond were reduced before SDS-PAGE. This treatment with 2-mercaptoethanol was performed on stained gel pieces. Due to the acidic pH of the samples, only a partial reduction of the disulfide bond is expected.

## Results

### Amino Acid Sequence of SF-Assemblin

SF-assemblin blotted from SDS-PAGE on to a PVDF membrane lacked an  $\text{NH}_2$ -terminal blocking group and 21 residues were obtained by automated sequencing. When a gel slice containing SF-assemblin was treated with en-



**Figure 1.** SDS-PAGE of purified SF-assemblin and the two fragments produced by treatment with endoproteinase Glu-C. Lane 1 shows the purity of preparations used for protein chemical studies. Gel slices of the 34-kD band were treated according to Cleveland et al. (1977) with endoproteinase Glu-C and again subjected to SDS-PAGE (lane 3 shows a more extensive treatment than lane 2). The resulting two fragments of molecular masses 12 kD and 20 kD as well as the untreated SF-assemblin were blotted onto PVDF and subjected to NH<sub>2</sub>-terminal sequencing. Since intact SF-assemblin and the 12-kD fragment showed the same sequence (PTPSPEARVASRPFLLDSPLPG, see Fig. 2), the 12-kD fragment spans the NH<sub>2</sub>-terminal part of SF-assemblin. Sequencing of the 20-kD fragment for 35 steps revealed two sequences staggered by two residues (SEVKGLQERT . . . and VKGLQERT . . .). The two cleavage sites of Glu-C are glutamic acid 105 and glutamic acid 107, respectively (see text and Fig. 2). The molecular weights of the fragments were calculated from their sequences (Fig. 2). Approximate molecular weights (45, 36, 29, 24, 20, and 14 kD) are indicated on the left.

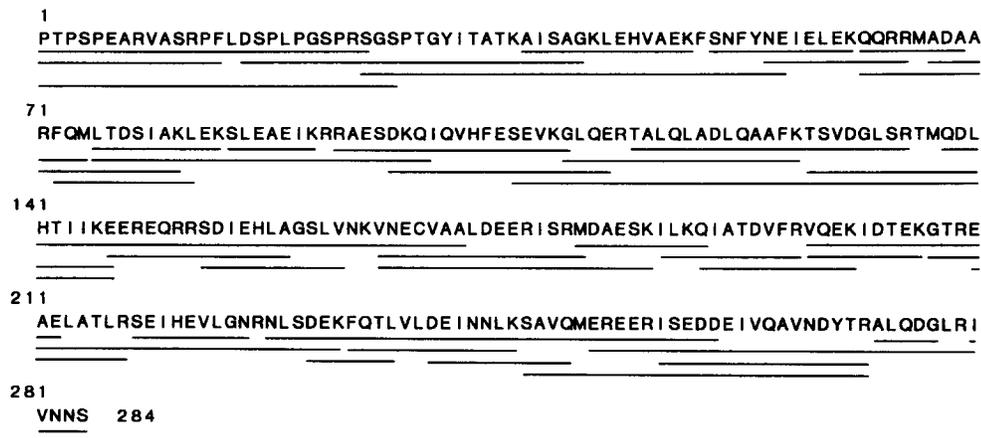
doproteinase Glu-C, two fragments of apparent molecular weights 20,000 and 12,000 were produced (Fig. 1). Both fragments were blotted on PVDF and sequenced. The smaller fragment revealed the NH<sub>2</sub>-terminal sequence of in-

tact SF-assemblin and therefore covers the NH<sub>2</sub>-terminal part of the molecule. The larger fragment was sequenced for 35 residues. It showed two related sequences, which were staggered by two residues. The two cleavage sites of Glu-C, which led to the COOH-terminal fragment of molecular weight 20,000, were later identified as Glu-105 and Glu-107, respectively (Fig. 2). Additional information was obtained from SF-assemblin treated on a blot either with trypsin or endoproteinase Asp-N. Peptides released from the blot were subjected to HPLC and the elution profiles screened for pure peptides by sequencing. The combined results accounted for more than 120 residues and indicated that SF-assemblin could be a new type of  $\alpha$ -helical protein with coiled coil forming ability. Since such proteins are often easy to sequence (see for instance Geisler and Weber, 1982), we obtained a full sequence without resorting to cDNA cloning.

A total of 110  $\mu$ g of SF-assemblin was alkylated under denaturing conditions with 4-vinylpyridine to modify cysteine residues. Two thirds of the recovered protein were subjected to CNBr cleavage. HPLC provided four pure fragments—1 (residues 1–66), 2 (residues 67–74), 4 (residues 138–181) and 6 (residues 252–284), and several mixtures. These arose from overlapping fragments and fragments 3 (residues 75–137) and 5 (residues 182–251). Direct sequencing and the results obtained on secondary products obtained by endoproteinases Asp-N, Lys-C, and Arg-C established the sequences for all CNBr fragments and showed that fragment 6 (residues 252–284) was the only CNBr fragment lacking a COOH-terminal homoserine. Therefore fragment 6 has to span the COOH-terminal region of the polypeptide chain. Due to the presence of larger fragments, which arose by incomplete CNBr cleavage, nearly all information necessary for the linear arrangement of the CNBr fragments was obtained. At this stage the sequence was complete except for a longer direct overlap between fragments 4 and 5. Therefore the last third of the alkylated preparation of SF-assemblin was digested with endoproteinase Lys-C in the presence of 4 M urea. Peptides separated by HPLC were subjected to sequence analysis. The digest with Lys-C provided the missing overlapping fragment and yielded several additional fragments, which confirmed the proposed sequence.

### SF-Assemblin Has a Segmented Coiled Coil

The combined results show that SF-assemblin consists of a single polypeptide of 284 residues (Fig. 2). Its chemical mo-



**Figure 2.** The amino acid sequence of SF-assemblin of the flagellate green alga *Spermatozopsis similis*. The complete sequence was obtained from CNBr fragments and proteolytic fragments as described in Results. The lines underneath the sequence delineate some of the fragments used to establish the sequence. The polypeptide chain covers 284 residues.



tion of the long helical sequences reveals a pronounced coiled coil forming ability. The presumptive coiled coil shows, over most of the rod domain, a 29-residue repeat pattern which is based on four heptads followed by a skip residue (Fig. 4). Five of the first six skip residues are glutamic acid, while one is glutamine.

A segmented coiled coil with a 29-residue repeat pattern has been discussed for  $\beta$ -giardin, a member of the group of giardin proteins which seem associated with the complex microtubular organization of the sucking disc in the protozoan flagellate *Giardia lamblia* (Holberton et al., 1988).  $\beta$ -giardin is indeed the only homologous protein found when the protein data bank is searched with the sequence of SF-assemblin. Fig. 4 shows the sequence alignment of both proteins.  $\beta$ -giardin has a minute head domain of 6 residues, four of which are charged, while SF-assemblin displays a much longer nonhelical domain of 31 residues. Although in the subsequent rod domains the proteins never display more than two identical residues in a row, these domains reveal good homology. Although the level of sequence identity is low (20%), the level of sequence similarity is much higher (43%) and the two sequences are aligned over the entire length of 253 residues without a single gap. In both rods, the third and fourth heptads of each 29 residue repeat show the dominance of hydrophobic residues in the interior *a* and *d* positions, while the first two heptads of the repeat have a much less perfect hydrophobic pattern. Furthermore, in both rod domains there is a preference for the chemical nature of the skip residue. Five of the first six skip residues are glutamic acid, the other is glutamine.

Beyond the sixth repeat, the  $\beta$ -giardin sequence is thought to show a reversal in heptads although it continues with more than two full repeats to the COOH-terminal end (Holberton et al., 1988). We have followed this proposal with the sequence alignment of SF-assemblin (Fig. 4), although alternative ways of heptad arrangements seem to exist in this region. In addition, we note that secondary structure prediction rules raise the possibility of  $\beta$ -sheet at the COOH-terminal 10–20 residues (see above). Nevertheless, because of the more pronounced sequence identity toward the COOH-terminal ends and the length agreement, we assume that both rod domains have a similar folding in the last segment and that the rod domains could continue to the COOH-terminal ends.

Fourier analysis of the rod sequence of SF-assemblin supports the proposed coiled coil structure based on 29 residue repeats formed by four heptads and a skip residue (data not shown; see, however, Fig. 5 as a summary). The Fourier transform showed a strong peak for the hydrophobic residues near 7/2 with the precise peak position located at 29/8. The acidic residues showed the fundamental at 1/29 but the third and fourth orders were stronger indicating that there is probably a repeat near 8–9 residues. Visual examination of the sequence of the 29 residue motifs indicates an approximate threefold repeat and shows a strong clustering of acidic residues near the COOH-terminal end of the motifs. It is present in all complete repeats except for the third one. These periodicities are obvious from the alignment of the consecutive 29-residue repeats given in Fig. 5. Interestingly, a periodic clustering of negative charges at the end of the motif is found also, but less pronounced, in  $\beta$ -giardin (see also Holberton et al., 1988).



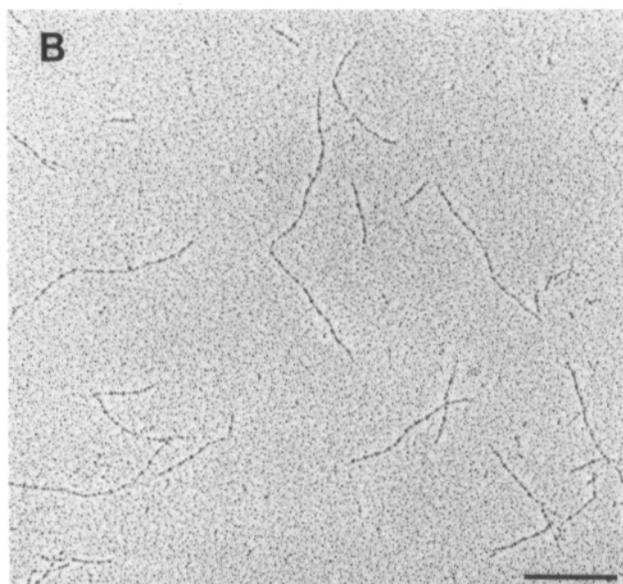
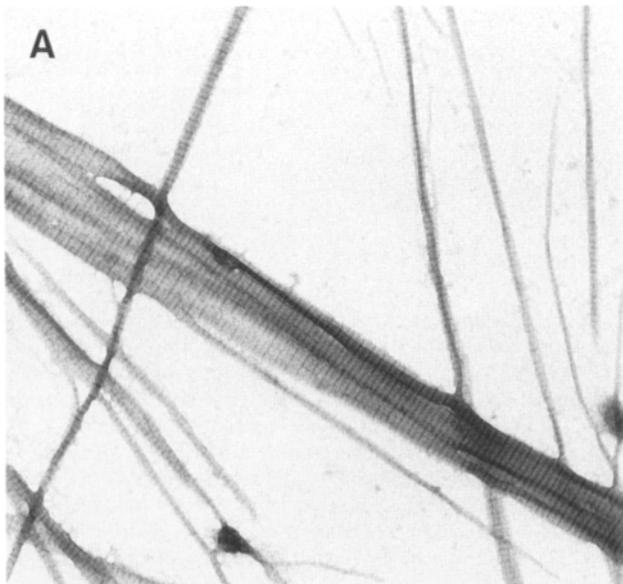
Figure 5. Significant periodicities in the rod domain of SF-assemblin. The sequence of the rod domain (residues 32–284 from Fig. 4) is written as consecutive 29 residue segments. The skip residues in the last column are marked by a star. The number of acidic (*D* and *E*), basic (*K*, *R*, and *H*), and hydrophobic (*A*, *V*, *M*, *I*, *L*, *F*, and *Y*) residues for each column is recorded. Absence of such residues is indicated by dots. The heptad positions in the bottom line refer to the first six segments (see Fig. 4 and text). The sequence repeats every 29 residues and there is a distinct clustering of acidic residues near the COOH terminus of the 29 residue motifs.

### Molecular Characterization of SF-Assemblin

Due to the limited amount of pure SF-assemblin, we have concentrated on two major predictions made by the sequence interpretations. Proteins forming coiled coils renature easily after they have been denatured in 8 M urea (see, for instance, Gessler et al., 1982; Hatzfeld and Weber, 1990; Steinert and Roop, 1988). Therefore, paracrystals were dissolved in 8 M urea and SF-assemblin was purified in this solvent by ion exchange chromatography on Mono Q. Dialysis against reassembly buffer removed the urea and EM documented the extensive formation of paracrystals with the striation pattern of the original fibers (Fig. 6 A). Solubilization of the paracrystals in 50 mM sodium phosphate buffer, pH 7.8, containing 2 M urea resulted in the release of individual protofilaments (Fig. 6 B). These can have a considerable length corresponding to 50 or more 28-nm repeat units. Their average thickness is  $\sim 4$  nm including the metal deposit. They often appear to be regularly beaded with the repeat unit similar or most likely identical to the repeat length of 28 nm found in the crystals (see also Patel et al., 1992). The spherical beads have a diameter which is  $\sim 1.5$ -fold thicker than the diameter of the connecting parts. Similar images are seen when SF-assemblin is dialyzed against 20 mM Tris-HCl, pH 8.8, in the presence of 1 mM EDTA, and 0.1 mM 2-mercaptoethanol.

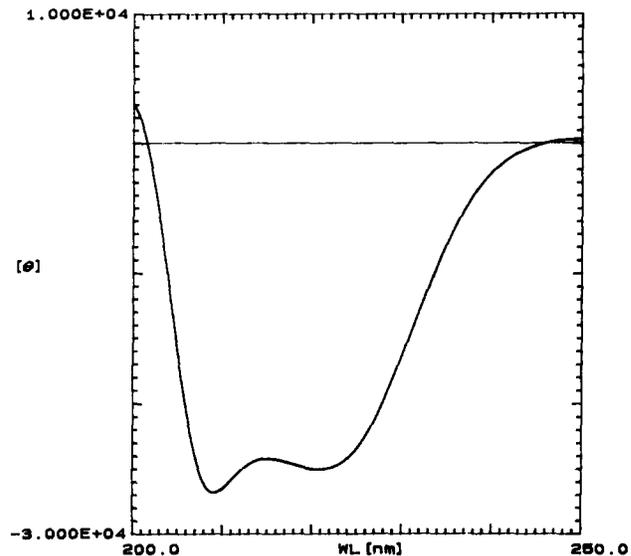
Purification of SF-assemblin by Mono Q chromatography also allowed a spectroscopic analysis. The paracrystals formed in reassembly buffer were harvested and solubilized in 2 M urea. Circular dichroism measurements revealed a high degree of helicity (Fig. 7). Using the approximation method of Greenfield and Fasman (1969) at 208 nm, we calculate an  $\alpha$ -helix content of 76%. Thus at least 216 of the 284 residues of SF-assemblin retain  $\alpha$ -helix conformation in 2 M urea. Therefore the postulated rod domain (residues 32–284; Fig. 4) has an  $\alpha$ -helix content of at least 85%. This value is probably a lower estimate since 2 M urea had to be used to dissolve the paracrystals.

The sole cysteine of SF-assemblin (residue 169; Fig. 2) oc-



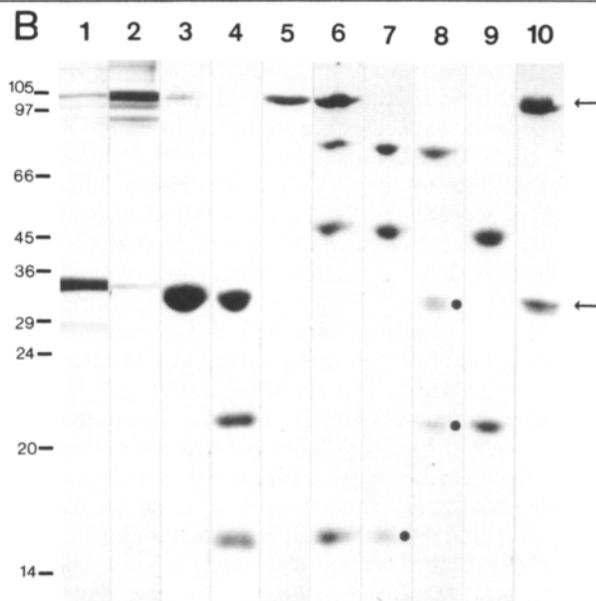
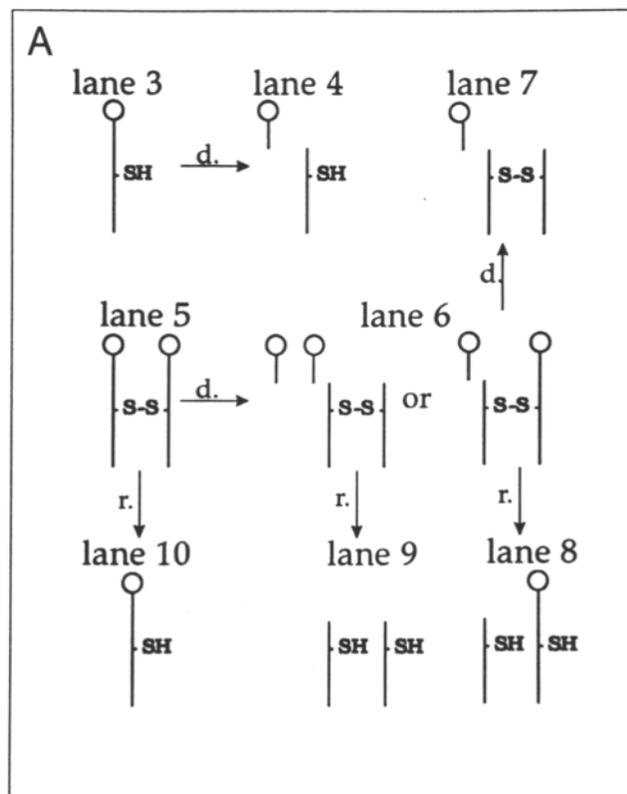
**Figure 6.** Paracrystal formation by renatured SF-assemblin (*A*) and the appearance of protofilaments (*B*). SF-assemblin purified in 8 M urea by anion exchange chromatography was dialyzed against reassembly buffer to remove the urea. An aliquot of the resulting paracrystals was analyzed by negative staining with uranyl acetate (*A*). Paracrystals dialyzed against 2 M urea in 50 mM sodium phosphate buffer, pH 7.8, provide protofilaments which are shown as rotary shadowed structures (*B*). Such filaments often reveal a regular beading with a repeat length similar to that seen in the paracrystals. Both micrographs are at the same magnification. Bar, 0.25  $\mu\text{m}$ .

copies in the proposed  $\alpha$ -helix an interior (*d*) heptad position (Fig. 4). Cysteines present in such positions of the double stranded coiled coils of tropomyosin or recombinantly designed keratins 8 plus 18 yield upon air oxidation a covalent dimer due to the formation of the disulfide between the parallel and in register arranged  $\alpha$ -helices (Lehrer, 1975; Stewart, 1975; Hatzfeld and Weber, 1990). Therefore fibers were solubilized in 2 or 4 M urea. These are conditions which generally do not lead to denaturation of coiled coils, and par-



**Figure 7.** Circular dichroism of SF-assemblin. Measurements were made in 2 M urea buffer (see Materials and Methods). The presence of urea, necessary to solubilize the protein, prevented measurements below 200 nm.  $[\theta]$  is the mean residue ellipticity ( $\text{deg} \times \text{cm}^2 \times \text{decimole}^{-1}$ ). Note the typical  $\alpha$ -helical spectrum.

allel gel filtration experiments showed that SF-assemblin still forms higher oligomers in these solvents (our unpublished results). Subsequent dialysis against the same solvents lacking a reducing agent lead to a 60–80% conversion of SF-assemblin into a component with an apparent molecular weight of 96,000 when analyzed by SDS-PAGE under non-reducing conditions (Fig. 8). That this derivative is indeed a dimer with an aberrant electrophoretic mobility rather than a trimer rests on the presence of a single cysteine residue per polypeptide (Fig. 2) and on the following experiments (Fig. 8). Treatment of the 96,000-molecular wt band with endoproteinase Glu-C yielded in subsequent SDS-PAGE under nonreducing conditions three fragments with apparent molecular weights of 12,000, 45,000, and 75,000. These were identified by  $\text{NH}_2$ -terminal sequencing and SDS-PAGE under reducing conditions. The 12,000-molecular wt component is the monomeric  $\text{NH}_2$ -terminal region of SF-assemblin (see also Fig. 1) and its molecular weight fits the value calculated from the sequence (Fig. 2). The 45,000-molecular wt component is the dimer of the COOH-terminal fragment (positions 106 and 108–284) in line with the location of the single cysteine at position 169. The component of molecular weight 75,000 arises because of incomplete Glu-C digestion. It contains one intact chain of SF-assemblin linked via the disulfide to the COOH-terminal region of a second chain which lacks the  $\text{NH}_2$ -terminal region. Thus the results show that undenatured SF-assemblin readily forms a disulfide-containing dimer, which has an aberrantly low electrophoretic mobility (for similar abnormal behavior of oxidized proteins in SDS-PAGE see also Bretscher and Weber, 1980; Geisler and Weber, 1982; Quinlan and Franke, 1982). Whether this dimer reflects the coiled coil or arises by an inter coiled coil disulfide bond in a higher oligomer remains to be seen.



**Figure 8.** Formation of the disulfide stabilized dimer of SF-assemblin and its characterization. (A) Schematic drawing of SF-assemblin (lane 3), its disulfide stabilized dimer (lane 5) and the fragments resulting by treatment with endoproteinase Glu-C (lanes 4, 6, and 7), or reduction with 2-mercaptoethanol (lanes 8–10). The NH<sub>2</sub>-terminal head domain of SF-assemblin is indicated by a circle, the position of the single cysteine (residue 169) is marked by SH or S-S, indicating the monomer or dimer stage, respectively. Glu-C treatment is marked by d. and reduction of the disulfide is given by r. Lane numbers correspond to the lanes of the SDS-PAGE given in B. (B) Paracrystals of SF-assemblin were dissolved in MT buffer containing 2 M urea and no reducing agent. The protein was allowed to oxidize (see Materials and Methods) before SDS-PAGE with (lane 1) and without (lane 2) reducing agent. The dimer has

## Discussion

The striated microtubule associated fibers (system I fibers) in the basal apparatus of *Spermatozopsis similis*, a flagellate green alga, consist of several layers of 2-nm filaments which arise from a single protein. SF-assemblin originally solubilized by 2 M urea forms paracrystals with the pronounced 28 nm periodicity typical for system I fibers (Lehtreck and Melkonian, 1991; Patel et al., 1992) and the same structures are obtained when SF-assemblin purified in 8 M urea on Mono Q is dialyzed against reassembly buffer (Fig. 6). Because of this self-assembly process, we expected the sequence of SF-assemblin to provide some insights into the structural principles involved in the organization of the striated fibers. The complete amino acid sequence, which was obtained on ~100 μg of protein, documents a single polypeptide chain of 284 residues corresponding to a molecular weight of 32,076 (Fig. 2). Thus the four isoforms observed in two-dimensional gels (Lehtreck and Melkonian, 1991) probably arise by differential and substoichiometric phosphorylation of certain sites.

The sequence of SF-assemblin allows unexpectedly clear structural predictions. It identifies a short nonhelical head domain formed by the NH<sub>2</sub>-terminal 31 residues, which is rich in proline (25%) and hydroxyamino acids (25%). Some of these serines and threonines may serve as targets of *in vivo* phosphorylation. The following rod domain of 253 residues is highly α-helical by secondary structure prediction rules (Fig. 3). Circular dichroism spectroscopy measured in 2 M urea on solubilized protofilaments confirms this prediction (Fig. 7). The pronounced coiled coil forming ability of the rod domain is based on a segmented coiled coil built from 29 residue repeats. The repeat arises from four heptads followed by a skip residue (Fig. 4). Exactly this type of structure has been predicted from cDNA cloning for β-giardin of

an aberrantly low mobility corresponding to a molecular mass of 96 kD. Purified SF-assemblin and the results of a Glu-C treatment are shown under reducing conditions in lanes 3 and 4. Glu-C treatment of the SF-assemblin disulfide (lane 5) yields under nonreducing conditions fragments of 75, 45, and 12 kD in addition to the residual dimer (lane 6). NH<sub>2</sub>-terminal sequencing identified the 12-kD band as the NH<sub>2</sub>-terminal fragment of SF-assemblin also observed when reduced SF-assemblin is treated with Glu-C (compare lane 4). The molecular mass of the 12-kD band fits the value calculated from its sequence. The 45-kD band corresponds by its NH<sub>2</sub>-terminal sequence to the dimer of the COOH-terminal fragment starting at residues 106 and 108, respectively (see also Figs. 1 and 2). Its reduction leads to the expected 20-kD fragment (lane 9). By the same criteria the 75-kD band of lane 6 is the disulfide between an intact polypeptide and a polypeptide lacking the NH<sub>2</sub>-terminal 12 kD region. Redigestion of the band with Glu-C yields 45 and 12-kD bands (lane 7). Reduction of the band provides partial conversion into the intact monomeric SF-assemblin at 34 kD and the monomeric 20-kD fragment (lane 8). Lanes 9 and 10 show the incomplete reduction of the 45-kD band and the SF-assemblin dimer, respectively. Molecular mass markers are indicated on the left. Arrows on the right mark the position of SF-assemblin and its dimer. Weak bands are indicated by a dot at the right. Note that all Glu-C treatments and all reductions are incomplete. The latter point is due to the acidic pH used (see text). The results support the existence of a dimer which has an unusually low electrophoretic mobility.

the protozoan *Giardia lamblia* (Holberton et al., 1988) and  $\beta$ -giardin is the only protein in the data bank obviously related to assemblin. The rod domains of SF-assemblin and  $\beta$ -giardin share 43% sequence similarity but only 20% sequence identity. In spite of a lack of consensus sequences and the low level of sequence identity, we note a strong conservation of sequence principles for both rods. They are aligned without the need to introduce a single gap and have identical lengths (253 residues). They can be described as segmented coiled coils built from 29 consecutive residue repeats, and show a strong preference for the nature of the skip residue. In the first six repeats, the skip residue is five times E and once Q. In the following two repeats the skip residue can also be hydrophobic. Finally, we note that the last repeat, which is incomplete, shows low  $\alpha$ -helical potential. Nevertheless, over this region the rods are well related in sequence, so the same conformation can be assumed for both proteins (Figs. 3 and 4).

A similar structure for proteins forming a microtubule-associated organization of 2-nm filaments in a protozoan and an alga indicates that related molecules can also be expected from higher eukaryotes. In this connection the protozoan molecule still poses some problems in interpretation. *Giardia lamblia* reveals an unexpected complexity of the family of giardins in two-dimensional gels (Peattie et al., 1989) and at least two additional giardins have been described by cDNA sequences. They differ in structural principles from the  $\beta$ -giardin sequence (Peattie et al., 1989; Aggarwal and Nash, 1989). This heterogeneity may be related to the high complexity of the microtubular dominated sucking disc of the protozoan (Holberton and Ward, 1981). Such problems of structural interpretation do not arise for the flagellate green alga. Here the well documented system I fibers of the basal apparatus can be isolated and convincingly reassembled from SF-assemblin (Lechtreck and Melkonian, 1991; Patel et al., 1992; and Fig. 6A), which is now characterized as a single protein species. Since the basal apparatus of the alga is the structural homologue of the centriolar apparatus of mammalian cells (for reviews see Wheatley, 1982; Kalnins, 1992), the architectural knowledge developed on the basal apparatus can most likely be extended to higher eukaryotes. Thus we expect that proteins related to SF-assemblin and  $\beta$ -giardin may be constituents of flagellar root fibers (for example, kinetodesmal fibers of ciliates; Sperling et al., 1991) in other organisms.

Search of the sequence of the SF-assemblin rod domain for internal periodicities clearly shows that the sequence repeats every 29 residues and that there is a very distinct clustering of acidic residues near the COOH terminus of the 29 residue motif (Fig. 5). This finding fits the observation that the microtubule periodicity of 4.2 nm matches 28 residue repeats in coiled coil molecules (McLachlan and Stewart, 1982), and that such a match may be potentially important for the interactions between system I fibers and microtubules. Alternatively, the periodic acidic clusters may also somehow contribute to the self-assembly process of SF-assemblin into the system I fibers.

The molecular properties of SF-assemblin so far established offer also a partial answer about the structure of the striated microtubule associated fibers. A detailed EM and image reconstruction study of system I fibers arrived at an overlap-gap model in which rodlike molecules, 48 nm in length, have a staggered arrangement and identical polarities. The unit particle of 48 nm is thought to be a single dou-

ble stranded coiled coil (Patel et al., 1992). While our results support the coiled coil nature of SF-assemblin and document a 29-residue repeat pattern, they raise a question concerning the molecular identity of the 48-nm particle. In contrast to the assumption of Patel et al. (1992) it cannot reflect solely a single coiled coil. Because of the nonhelical head domain with its eight proline residues, the maximum length of the coiled coil cannot exceed the 253 residues of the entire rod domain and thus a length of  $\sim 37$  nm. A 48-nm particle could arise, however, by additional length contributions from the head domain and the COOH-terminal end of the rod being possibly in a nonhelical conformation (see Fig. 3). Alternatively, a 48-nm particle could reflect the presence of two coiled coils in a staggered arrangement rather than a single coiled coil. This type of structure is found, for instance, as the tetrameric building block of intermediate filaments (Stewart et al., 1989; Potschka et al., 1990). A systematic study is now necessary to define those depolymerization conditions, which provide a homogenous population of SF-assemblin molecules suitable for determining the oligomeric state and the molecular length by hydrodynamic parameters and EM. Such conditions should also allow a decision as to whether the oxidative crosslink in SF-assemblin (Fig. 8) arises from a double stranded coiled coil within register-arranged  $\alpha$ -helices or from a higher oligomeric state.

EM of paracrystals of SF-assemblin solubilized by 2 M urea reveals long thin filaments. They often display a regular beading, which seems very similar to the repeat length observed in the crystals (Fig. 6). If these filaments represent the protofilaments previously identified in the crystals (Patel et al., 1992) their appearance argues against the possibility that consecutive molecules are separated by a gap. Instead it seems conceivable that the dark bands in the crystal represent areas of the molecules which have a different affinity for the stain, as suggested by Patel et al. (1992) as an alternative possibility. The repeat length of 28 nm in the protofilaments and in the crystals is too small to be consistent with a single extended coiled coil ( $\sim 37$  nm) and thus may result from a staggering of neighboring coiled coils as discussed above. In such a case, the protofilaments obtained in 2 M urea could be built from at least two rows of molecules.

The extensive morphological literature on ultrathin filaments present in various cell types and structures sometimes assumes that because of the small diameter, such filaments are related structures. This is, however, often not the case (see also Roberts, 1987). For instance, the ultrathin titin filaments of the myofibril are built by a single gigantic molecule which connects the Z and M lines. Titin arises from 100-residue repeats of the immunoglobulin C2-type domains and the fibronectin type III domains, which are entirely built as  $\beta$  sheet (Labeit et al., 1990; Fürst et al., 1988; Nave et al., 1989). In contrast, the 2-nm filaments of the striated microtubule associated fiber arise from a small molecule with an extended  $\alpha$ -helix able to form a special coiled coil. Thus the thin titin filaments and the 2-nm filaments arise from two entirely distinct molecules which share no sequence relation. Even among coiled coil forming molecules a clear distinction can be made. Fine filaments in close proximity to microtubules have also been described in sea urchin sperm. Such filaments, built from so-called tektins with molecular weights around 50,000, occur in basal bodies and axonemes (Linck and Langevin, 1982). The tektins are  $\alpha$ -helical proteins in circular dichroism analysis and share some epitopes with the proteins of the 10-nm filaments (Chang and

Piperno, 1987; Steffen and Linck, 1989), which are particularly well studied in vertebrates (for review see Steinert and Roop, 1988). However, the primary structure of tektin A1 reveals only a very remote sequence homology with the 10-nm filament proteins (Norrander et al., 1992). While both protein types can form a segmented coiled coil due to the separation of the  $\alpha$ -helices by nonhelical sequences (Geisler and Weber, 1982; Steinert and Roop, 1988; Norrander et al., 1992), the tektin A1 lacks the consensus sequences of the 10-nm filament proteins. Interestingly, tektin A1 shows no obvious homology with  $\beta$ -giardin (Norrander et al., 1992) and SF-assemblin. Direct sequence comparison of intermediate filament proteins and SF-assemblin shows only a marginal homology. This involves a small part of the coil 1a domain of some type II keratins and the phenylalanine/tyrosine rich region around residue 55 of SF-assemblin. In contrast, SF-assemblin and  $\beta$ -giardin show a striking structural relation over the entire rod domains, including a 29-residue repeat pattern based on four heptads followed by a skip residue. These results emphasize that coiled coil forming molecules have a surprising individuality which is readily recognized once extended amino acid sequences are available.

This study was supported in part by grants from the Deutsche Forschungsgemeinschaft (ME 658/9-1).

Received for publication 18 September 1992 and in revised form 3 February 1993.

## References

Aggarwal, A., and T. E. Nash. 1989. Characterization of a 33-kilodalton structural protein of *Giardia lamblia* and localization to the ventral disk. *Infect. Immun.* 57:1305-1310.

Bauw, G., J. Van Damme, M. Puype, J. Vandekerckhove, B. Gesser, G. P. Ratz, J. B. Lauridsen, and J. E. Celis. 1989. Protein-electroblotting and -microsequencing strategies in generating protein bases from two-dimensional gels. *Proc. Natl. Acad. Sci. USA.* 86:7701-7705.

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

Chang, X., and G. Piperno. 1987. Cross-reactivity of antibodies specific for flagellar tektin and intermediate filament subunits. *J. Cell Biol.* 104:1563-1568.

Chou, P. Y., and G. D. Fasman. 1974. Prediction of protein conformation. *Biochemistry.* 13:222-245.

Cleveland, D. W., S. G. Fischer, M. W. Kirschner, and U. K. Laemmli. 1977. Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* 255:1102-1106.

Fürst, D. O., M. Osborn, R. Nave, and K. Weber. 1988. The organization of titin filaments in the half-sarcomere revealed by monoclonal antibodies in immunoelectron microscopy. A map of ten nonrepetitive epitopes starting at the Z line extends close to the M line. *J. Cell Biol.* 106:1563-1572.

Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97-120.

Geisler, N., and K. Weber. 1982. The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* 1:1649-1656.

Geisler, N., E. Kaufmann, and K. Weber. 1982. Protein chemical characterization of three structurally distinct domains along the protofilament unit of desmin 10 nm filaments. *Cell.* 30:277-286.

Greenfield, N., and G. Fasman. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry.* 10:4108-4116.

Hatzfeld, M., and K. Weber. 1990. The coiled coil of *in vitro* assembled keratin filaments is a heterodimer of type I and type II keratins: use of site-specific mutagenesis and recombinant protein expression. *J. Cell Biol.* 110:1199-1210.

Holberton, D. V., and A. P. Ward. 1981. Isolation of the cytoskeleton from *Giardia*. Tubulin and a low-molecular weight protein associated with microtubule structures. *J. Cell Sci.* 47:139-166.

Holberton, D. V., D. A. Baker, and J. Marshall. 1988. Segmented  $\alpha$ -helical coiled coil structure of the protein Giardin from the *Giardia* cytoskeleton. *J. Mol. Biol.* 204:789-795.

Huang, B., A. Mengersen, and V. D. Lee. 1988. Molecular cloning of cDNA for caltractin, a basal body-associated  $\text{Ca}^{2+}$ -binding protein: homology in

its protein sequence with calmodulin and the yeast CDC31 gene product. *J. Cell Biol.* 107:133-140.

Kalnins, V. I. 1992. The Centrosome. Academic Press Inc., New York. 368 pp.

Labeit, S., D. P. Barlow, M. Gantel, T. Gibson, C.-L. Hsieh, U. Francke, K. Leonard, J. Wardale, A. Whiting, and J. Trinick. 1990. A regular pattern of two types of 100-residue motif in the sequence of titin. *Nature (Lond.).* 345:273-276.

Lechtreck, K. F., and M. Melkonian. 1991a. Striated microtubule-associated fibers: identification of assemblin, a novel 34-kD protein that forms paracrystals of 2-nm filaments *in vitro*. *J. Cell Biol.* 115:705-716.

Lechtreck, K. F., and M. Melkonian. 1991b. An update on fibrous flagellar roots in green algae. *Protoplasma.* 164:38-44.

Lehrer, S. S. 1975. Intramolecular crosslinking of tropomyosin via disulfide formation: evidence for chain register. *Proc. Natl. Acad. Sci. USA.* 72:3377-3381.

Linck, R. W., and G. L. Langevin. 1982. Structure and chemical composition of insoluble filamentous components of sperm flagellar microtubules. *J. Cell Biol.* 58:1-22.

McLachlan, A. D., and M. Stewart. 1982. Periodic charge distribution in the intermediate filament proteins desmin and vimentin. *J. Mol. Biol.* 162:693-698.

Melkonian, M. 1980. Ultrastructural aspects of basal body associated fibrous structures in green algae: a critical review. *Biosystems.* 12:85-104.

Melkonian, M., P. L. Beech, C. Katsaros, and D. Schulze. 1992. Centrin-mediated cell motility in algae. *In Algal Cell Motility.* M. Melkonian, editor. Chapman and Hall, New York. 179-221.

Nave, R., D. O. Fürst, and K. Weber. 1989. Visualization of the polarity of isolated titin molecules. A single globular head on a long thin rod as the M band anchoring domain. *J. Cell Biol.* 109:2177-2188.

Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443-453.

Norrander, J. M., L. A. Amos, and R. W. Linck. 1992. Primary structure of tektin A1: comparison with intermediate filament proteins and a model for its association with tubulin. *Proc. Natl. Acad. Sci. USA.* 89:8567-8571.

Patel, H., K.-F. Lechtreck, M. Melkonian, and E. Mandelkow. 1992. Structure of striated microtubule-associated fibers of flagellar roots: comparison of native and reconstituted states. *J. Mol. Biol.* 227:698-710.

Peattie, D. A., R. A. Alonso, A. Hein, and J. P. Caulfield. 1989. Ultrastructural localization of giardins to the edges of disk microribbons of *Giardia lamblia* and the nucleotide and deduced protein sequence of alpha giardin. *J. Cell Biol.* 109:2323-2335.

Potschka, M., R. Nave, K. Weber, and N. Geisler. 1990. The two coiled coils in the isolated rod domain of the intermediate filament protein desmin are staggered: a hydrodynamic analysis of tetramers and dimers. *Eur. J. Biochem.* 190:503-508.

Quinlan, R. A., and W. W. Franke. 1982. Heteropolymer filaments of vimentin and desmin in vascular smooth muscle tissue and cultured baby hamster kidney cells demonstrated by chemical crosslinking. *Proc. Natl. Acad. Sci. USA.* 79:3452-3456.

Roberts, T. M. 1987. Invited review: fine (2-5 nm) filaments: new types of cytoskeletal structures. *Cell Motil.* 8:130-142.

Salisbury, J. L. 1989. Algal centrin: calcium-sensitive contractile organelles. *In Algae As Experimental Systems.* A. W. Coleman, L. J. Goff, and J. R. Stein-Taylor, editors. Alan R. Liss Inc., New York. 19-37.

Salisbury, J. L., A. Baron, B. Surek, and M. Melkonian. 1984. Striated flagellar roots: isolation and partial characterization of a calcium-modulated contractile organelle. *J. Cell Biol.* 99:962-970.

Salisbury, J. L., A. T. Baron, and M. A. Sanders. 1988. The centrin-based cytoskeleton of *Chlamydomonas reinhardtii*: distribution in interphase and mitotic cells. *J. Cell Biol.* 107:635-641.

Sperling, L., G. Keryer, F. Ruiz, and J. Beisson. 1991. Cortical morphogenesis in paramecium: a transcellular wave of protein phosphorylation involved in ciliary rootlet disassembly. *Dev. Biol.* 148:205-218.

Steffen, W., and R. W. Linck. 1989. Relationship between tektins and intermediate filament proteins: an immunological study. *Cell Motil. Cytoskeleton.* 14:359-371.

Steinert, P. M., and D. R. Roop. 1988. The molecular and cellular biology of intermediate filaments. *Annu. Rev. Biochem.* 57:583-625.

Stewart, M. 1975. Tropomyosin: evidence for no stagger between chains. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 53:5-7.

Stewart, M., R. A. Quinlan, and R. D. Moir. 1989. Molecular interactions in paracrystals of a fragment corresponding to the  $\alpha$ -helical coiled coil rod portion of glial fibrillary acidic protein: evidence for an antiparallel packing of molecules and polymorphism related to intermediate filament structure. *J. Cell Biol.* 109:225-234.

Wessel, D., and U. I. Flügge. 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138:141-143.

Wheatley, D. N. 1982. The centriole: a central enigma of cell biology. Elsevier Biomedical Press, North Holland. 350 pp.

Wiche, G., B. Becker, K. Lubert, G. Weitzer, M. J. Castanon, R. Hauptmann, C. Stratowa, and M. Stewart. 1991. Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with a three domain structure based on a central alpha-helical coiled coil. *J. Cell Biol.* 114:83-89.