Journal of Cell Science 106, 909-918 (1993)
Printed in Great Britain © The Company of Biologists Limited 1993

Tektin B1 from ciliary microtubules: primary structure as deduced from the cDNA sequence and comparison with tektin A1

Runtian Chen[1], Catherine A. Perrone[1], Linda A. Amos[2] and Richard W. Linck[1]

[1]University of Minnesota, Department of Cell Biology and Neuroanatomy, 321 Church St, Minneapolis, MN 55455, USA
[2]Laboratory of Molecular Biology, Medical Research Center, Hills Road, Cambridge CB2 2QH, UK

SUMMARY

Tektins are a class of proteins that form filamentous polymers in the walls of ciliary and flagellar microtubules, and they may also be present in centrioles, centrosomes and mitotic spindles. We report here the cloning and sequencing of a cDNA for ciliary tektin B1. Comparison of the predicted amino acid sequence of tektin B1 with the previously published sequence for tektin A1 reveals several features that better define this class of proteins. Like tektin A1, the central region of the tektin B1 polypeptide chain is predicted to form a coiled-coil rod, consisting of four major α-helical regions that are separated by non-helical linkers. Between the central rod domains of tektins A and B there is a 34%/20% amino acid sequence identity/similarity, including equivalent 50-residue segments containing 36 identities, and a high probability of long-range structural homology. The tektin polypeptide chains are divided into two major segments that have significant sequence homology to each other, both within a given tektin chain and between tektins A and B, indicative of gene duplication events. The tektins have a secondary structure and molecular design similar to, but a low primary sequence homology with, intermediate filament proteins. Unlike tektin A1, tektin B1 lacks any part of the C-terminal IFP consensus sequence.

Key words: centrioles, cilia, coiled-coil proteins, flagella, intermediate filaments

INTRODUCTION

Tektins were originally identified as a set of proteins comprising relatively insoluble, 2-3 nm diameter filaments that remain after the extraction of sea urchin sperm flagellar microtubules with chaotropic solvents (Linck and Langevin, 1982). Three tektins have so far been characterized: tektin A (~53 kDa), B (~51 kDa) and C (~47 kDa). Tryptic peptide mapping (Linck and Stephens, 1987) and immunological studies (Linck et al., 1985; Steffen and Linck, 1988) demonstrated that the three tektins were different but related polypeptides and that they were unrelated to tubulin. Initial biochemical and structural data suggested that these novel proteins were similar to intermediate filament protein (IFP) from vertebrates in terms of their relative insolubility, molecular masses, and fibrous, α-helical substructure (see Linck, 1990). Immunological studies further substantiated the similarity between tektins and IFP (Chang and Piperno, 1987; Steffen and Linck, 1989a). It was surprising, therefore, that the first sequence reported for tektin A (Norrander et al., 1992) revealed only a weak homology with IFP.

The organization of tektins in microtubules is not known in detail, but tektins probably interact directly with tubulin. Immunofluorescence microscopy has demonstrated that tektins A, B and C are present in all nine flagellar doublet microtubules and possibly in the central pair microtubules (Steffen and Linck, 1988). Immunoelectron microscopy suggests that tektins form extended polymers in the walls of ciliary and flagellar microtubules (Linck et al., 1985). Sarkosyl detergent extraction of cilia and flagella yields insoluble ribbons of approximately three protofilaments, composed of tubulin and tektins in a close molar ratio, suggesting a direct molecular interaction (Linck, 1976). Finally, at least one form of the insoluble protofilament ribbons is localized to that part of the A-tubule to which the B-tubule and nexin links join (Linck, 1976; Stephens et al., 1989).

Immunological evidence suggests that tektins or tektin-related polypeptides are present in other microtubule systems. Besides staining flagellar axonemes, tektin antibodies also stain basal bodies in sea urchin sperm (Steffen and Linck, 1988). This result is not unexpected, since the doublet tubules assemble from the plus ends of the basal body triplet microtubules (Alberts et al., 1989; Allen and Borisy, 1974), and therefore, some proteins might be common to both doublet and triplet microtubules. Perhaps more surprisingly, tektin antibodies also stain centrioles in a variety of mammalian cells, including HeLa cells (Steffen and Linck, 1988), suggesting that tektins could be evolutionarily conserved components of centrioles. Tektin-like components have also been detected in centrosomes and in spindles and midbodies of dividing cells (Steffen and Linck, 1989b, 1990, 1992, and unpublished observations).

We report here the cloning and sequencing of a cDNA

for tektin B from sea urchin embryos. There were several reasons for this approach. First, sea urchin blastula embryonic cilia had been shown to contain tektins both by immunofluorescence microscopy and by SDS-PAGE immunoblotting, using antibodies raised against sperm flagellar tektins (Amos et al., 1986; Linck et al., 1987). Second, Stephens (1977) had shown that a ciliary microtubule protein of ~55 kDa is synthesized de novo at the onset of ciliogenesis in sea urchin blastulae; this 55 kDa polypeptide has since been identified as tektin A (Stephens, 1989) by crossreaction with a polyclonal antibody against flagellar tektin A (Linck et al., 1987). On the basis of these findings and having had success with cloning tektin A cDNA in this manner, we chose to screen a λZapII cDNA library from *Strongylocentrotus purpuratus* sea urchin blastula with antibodies against flagellar tektin B.

MATERIALS AND METHODS

Screening of cDNA library

A λZapII cDNA library was constructed commercially (Stratagene, Inc., La Jolla, CA) from mRNA of *Strongylocentrotus purpuratus* sea urchin blastula embryos, as prepared by Norrander et al. (1992). The cDNA library was screened with polyclonal antibodies specific to *S. purpuratus* sperm flagellar tektin B (Linck et al., 1987), according to established protocols (Huynh et al., 1985; Young and Davis, 1983). Positive plaques were identified by incubation of filters in a solution of alkaline phosphatase-conjugated goat anti-rabbit IgG (Stratagene, Inc.) for 1 hour, followed by color development with nitro blue tetrazolin (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) provided in the commercial kit.

cDNA cloning and blotting

Procedures for the purification of phage and plasmid DNA and the procedures for RNA blots are based on established techniques with some modification (Sambrook et al., 1989). pBluescript plasmids containing the cDNA inserts were excised from λZapII clones using the helper phage R408. The probe for northern blots was prepared by digesting the purified tek-B1 cDNA clone with *Eco*RI, isolating the tek-B1 insert and then labeling by nick-translation with [^{32}P]dCTP, as described in the manufacturer's protocol (BRL, Bethesda, MD). Hybridizations were carried out at 42°C for 18 hours in a solution containing 50% deionized formamide, 5× SSC (1× =150 mM sodium chloride, 15 mM sodium acetate), 5× Denhardt's (1× =0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 20 mM $NaHPO_4$, pH 6.5, 0.1% SDS, and 50 μg/ml sheared, denatured salmon sperm DNA and 10^6 cpm/ml denatured radiolabeled cDNA probe. Filters were washed four times 10 minutes in 2× SSC, 0.2% SDS at room temperature, followed by two 15-minute washes in 0.1× SSC, 0.1% SDS at 50°C.

cDNA sequencing

The nucleotide sequences of the clones were determined using the dideoxy chain termination method of Sanger et al. (1977). Single-stranded DNA templates were prepared by infecting cells harboring the plasmid pBluescript SK containing cDNA inserts with VCS-M13 filamentous bacteriophage, and then by isolating the resultant progeny phagemids (Sambrook et al., 1989). Partial clones were prepared by constructing overlapping deletion subclones based on controlled digestion of the double stranded DNA with exonuclease III (Putney et al., 1981), as described by Henikoff (1984, 1987) with modification. The United States Biochemical Corporation (Cleveland, OH) Sequenase kit (version 2.0) and the Promega (Madison, WI) *Taq*Track sequencing system were used to prepare the chain termination reactions with [^{35}S]dATP, as described in manufacturer's protocols. We sequenced either single-stranded DNA or double-stranded DNA from cDNA subclones in plasmid pBluescript SK (Stratagene) containing the full-length cDNA (tek-B1) or partial clones (tek-B2, tek-B3).

Fusion protein immunoblot analysis

Purification of inclusion bodies of bacterially expressed protein was performed according to the procedure of Marston (1987; Marston et al., 1984) and Rüther and Müller-Hill (1983) with modification. Tek-B1 cDNA/pBluescript SK$^-$ (Stratagene, Inc.) in LB media with a final concentration of 150 μg/ml ampicillin and 12.5 μg/ml of tetracycline was incubated at 37°C for 4 hours. Expression protein was induced with 1 mM IPTG and incubation overnight at 37°C with shaking. 500 ml of cell culture was centrifuged at 500 ***g*** for 15 minutes at 4°C. Each gram of cells was suspended in 3 ml lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl), 4 μl of 50 mM PMSF, 80 μl lysozyme (10 mg/ml), and 4 mg deoxycholic acid, and sonicated three times 10 seconds. *Escherichia coli* DNA was digested by adding 20 μl DNase I (1 mg/ml) and incubating at 37°C for 30 minutes. Cell lysates containing inclusion bodies were pelleted by centrifugation at 12,000 ***g*** for 15 minutes at 4°C. Inclusion bodies were further treated by resuspension in 15 ml lysis buffer with 0.5% Triton X-100, 10 mM EDTA, pH 8.0, and pelleting as above. The resulting protein was analyzed by SDS-PAGE immunoblotting (Towbin et al., 1979), using affinity-purified, polyclonal anti-(*S.p.*)-tektin B (Linck et al., 1987) and mouse monoclonal (#52-3-3) anti-tektin B (W. Steffen and R. Linck, unpublished observations).

The pBluescript SK phagemid (Stratagene, Inc.) contains 111 nucleotides coding for β-galactosidase (37 amino acid residues, accounting for an increase of 3,863 kDa in the fusion protein), followed by the *Eco*RI site into which our cDNA is cloned. By definition, our screening procedure isolated a clone with the correct reading frame to produce a fusion protein recognized by affinity-purified, polyclonal antibody to *S. purpuratus* tektin B. Thus, any 5′ untranslated region in the cDNA sequence without any stop codons would be readable and would also increase the apparent molecular mass of the fusion protein.

Protein purification and peptide sequencing

Sperm flagellar tektins were purified by *N*-lauryl-sarcosinate/urea extraction and reversed-phase (RP) HPLC, according to Linck and Stephens (1987). The purity of tektin B was assessed by SDS/PAGE. Tektin B was dissolved to ~2.5 mg/ml in 0.4 ml 70% trifluoroacetic acid and treated with cyanogen bromide, at ~30 μmoles of CNBr/mg protein. The mixture was flushed with N_2 and incubated at room temperature for 24 hours in darkness. After termination of the reaction by addition of excess water, the sample was evaporated to dryness. Samples were dissolved in 0.4 ml 6 M guanidine, 25 mM Tris, pH 8.5, 1 mM EDTA, 5 mM DTT, incubated for 30 minutes at room temperature, and subjected to RP-HPLC for peptide separation. Fractions containing pure peptides were collected and sequenced by automated Edman degradation on an Applied Biosystems 470 peptide sequencer, in the Microchemical Facility, University of Minnesota.

Sequence analysis

The sequence was analyzed using the program ANALYSEP (Staden, 1988), based on the studies of Robson and Suzuki (1976), in order to predict stretches of sequence that are more likely to form α-helix than any alternative secondary structure. The poten-

tial for α-helical coiled-coil formation was determined by the method of Lupas et al. (1991), which scores any stretch of α-helix for the presence of hydrophobic residues in the first (*a*) and fourth (*d*) positions of a heptapeptide repeat. Probability of coiled-coil formation was calculated for a 28-residue window.

Different stretches of sequence were compared using the program DIAGON (Staden, 1982) with a proportional matching calculation, first described by McLachlan (1982). The method of scoring depends on a table of amino acid similarity based on the substitution rates in known structures of families of related proteins (Dayhoff, 1972). The program produces a two-dimensional plot in which the *x* axis represents one sequence (tektin X) and the *y* axis represents the same or another sequence (tektin Y). Every point (*x*,*y*) is assigned a score that corresponds to the level of similarity between sequence characters over a chosen SPAN centered on point *x* in sequence X and point *y* in sequence Y, allowing no insertions or deletions. For analysis of tektin sequences, we chose two different span lengths: 11 residues, the default value for short-range sequence comparison, and 49 residues for longer-range comparison. For each point (*x*,*y*), the program calculates the probability of obtaining the same final score by chance with two infinitely long sequences of the same composition as those being compared. We set different minimum values for plotting the probabilities for different spans, in order to show the peaks clearly in each case; matches shown over spans of 11 residues have a probability of <0.0006 of occurring by chance, and matches shown over spans of 49 residues have a probability of <0.00001.

RESULTS

We obtained several positive cDNA isolates by probing a λZapII cDNA library from *Strongylocentrotus purpuratus*. Of these, clone tek-B1 has a size of 2,621 bp, sufficiently large to be a near-full-length cDNA. Two other clones were also isolated, tek-B2 and tek-B3, with sizes of ~1,300 bp and ~950 bp, respectively. Radiolabeled tek-B1, used to probe poly(A)$^+$ RNA isolated from sea urchin blastula, hybridizes with a single mRNA band of ~2,700 bases (Fig. 1). Since the pBluescript plasmid (Stratagene, Inc.) contains the N terminus of the *lac*Z gene, the recombinant β-galactosidase fusion protein can be expressed by IPTG induction of the *lac* promoter upstream from the *lac* gene. The fusion protein from tek-B1 was recognized by both a monoclonal antibody (#52-3-3, W. Steffen and R. Linck, unpublished observations) and a polyclonal antibody (Linck et al., 1987), both specific for sperm flagellar tektin B (Fig. 2). Compared to the apparent mass of flagellar tektin B (~51 kDa) measured by SDS-PAGE, a major immunoreactive fusion polypeptide of ~57 kDa was observed, along with two minor bands, which are presumably degradation products of the fusion protein; degradation of fusion proteins in *E. coli* systems is commonly observed (Sambrook et al., 1989). The increased size of the fusion protein compared to the apparent mass of native tektin B subunit is partially due to the addition of the 36 amino acid N-terminal peptide of β-galactosidase (see further discussion below).

Clones tek-B1, tek-B2 and tek-B3 were sequenced; only that of tek-B1 is presented here (Fig. 3). Compared to the predicted coding region of tek-B1, tek-B2 is identical and tek-B3 differs by only 1 bp (A replaces G at position 960), which does not change the predicted amino acid. tek-B1 is 2,621 nucleotide in length. A stretch of 41 adenine bases defines the 3′ end, and a consensus polyadenylation sequence, AATAAA (Berget, 1984), begins at nucleotide 2,562.

The predicted amino acid sequence of the longest open reading frame beginning with the first methionine in tek-B1 is presented in Fig. 4. Proof that tek-B1 is the cDNA for the ciliary tektin B polypeptide was obtained by comparing the amino acid sequence predicted from the cDNA with actual, partial amino acid sequence of proteolytic fragments of flagellar tektin B1. (Note: the ciliary and flagellar forms of tektin B are immunologically similar, have the same apparent molecular mass, and thus are thought to be homologous; Linck et al., 1987). Flagellar tektin B was purified to homogeneity by RP-HPLC and cleaved by cyanogen bromide. The CNBr-generated fragments were resolved by RP-HPLC, and peak fractions were collected. Three fragments were sequenced by automated Edman degradation (Table 1). There is close correspondence between the sequences predicted from the tek-B1 cDNA and the sequences of these three fragments (totaling 100 residues). The few discrepancies could be due to actual differences between ciliary and flagellar forms of tektin B or to occasional ambiguities in the amino acid sequence analysis. Thus, clone tek-B1 codes for tektin B and not for intermediate filament proteins (IFP, including nuclear lamins)

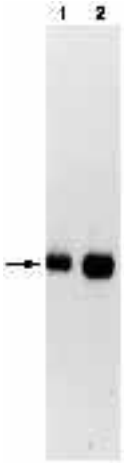


Fig. 1. Analysis of sea urchin embryonic mRNA with tek-B1 cDNA. Poly(A)$^+$ mRNA from blastula stage embryos of *S. purpuratus* was resolved electrophoretically in a 1.5% agarose gel containing formaldehyde, and transferred to nitrocellulose filters; lane 1, 5 μg; lane 2, 15 μg. Filters were then probed with ^{32}P-labeled nick-translated tek-B1 cDNA. The tek-B1 probe hybridizes with an ~2,700 base mRNA (arrow). The pBluescript SK plasmid alone did not hybridize to sea urchin mRNA (not shown). Relative molecular masses were determined by calibration with λ DNA digested with *Eco*RI and *Hin*dIII.

Fig. 2. Immunoblot analysis of the tek-B1 fusion protein. pBluescript SK phagemid with the cloned tek-B1 cDNA was induced for expression of the *lac*Z gene with IPTG. Fusion protein containing tek-B1 was prepared as described in Materials and Methods. The SDS-PAGE blot was probed with monoclonal antibody 52-3-3 against tektin B (lane 2) and polyclonal anti-tektin B antibody (lane 3). Sarkosyl-insoluble protofilament ribbons from sea urchin sperm flagella were probed with polyclonal anti-tektin B as an internal control and marker (lane 1). By SDS-PAGE the fusion protein (B, lanes 2 and 3) has an apparently larger molecular mass of ~57 kDa, compared to ~51 kDa for tektin B isolated from sea urchin sperm flagella.

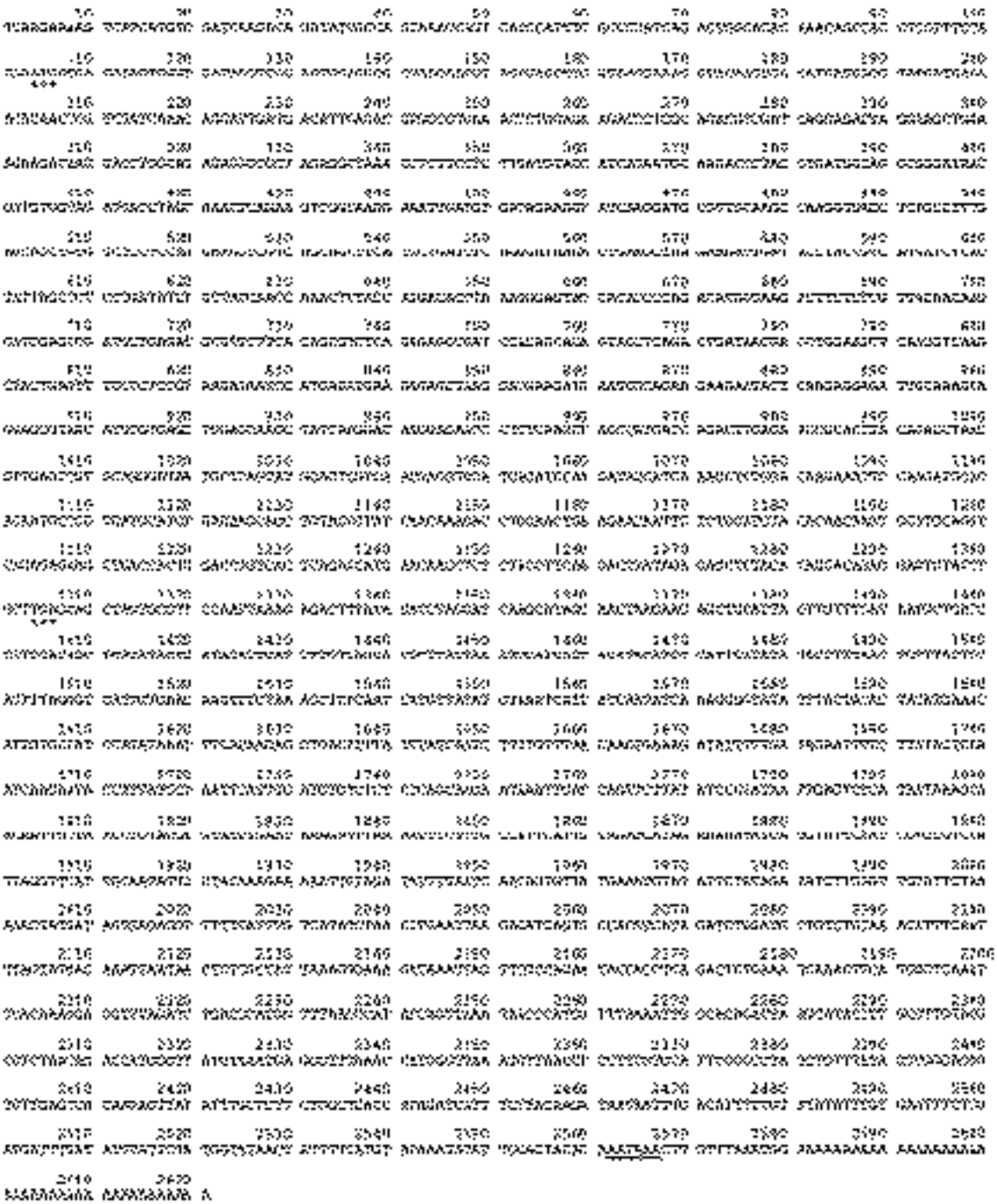


Fig. 3. The cDNA sequence of tek-B1 (encoding tektin B1). The 3′-polyadenylation signal is underlined. Asterisks mark the first AUG (start) codon (position 104-106) and the first in-frame termination codon (position 1306-1308), defining the longest open reading frame. This sequence has been deposited in the GenBank data base (accession no. L21838).

with which anti-tektin antibodies have been shown to cross-react (Chang and Piperno, 1987; Norrander et al., 1992; Steffen and Linck, 1989a).

Translation beginning with the first ATG (AUG) as the initiation site would produce a ciliary tektin B polypeptide with a relative molecular mass of 46,260, which is ~10% smaller than the 51,000 for flagellar tektin B, as estimated from SDS-PAGE (Linck and Langevin, 1982). This discrepancy may be due to tektin B running with an anomalously high relative molecular mass in SDS-PAGE, since the tektin B fusion protein, including the 36 residues of β-galactosidase (~3.9 kDa) and the 34 residues translated before the first Met in Fig. 3, has a predicted relative molecular mass of 53,900 but migrates as a ~57 kDa protein. This situation is similar to tektin A, which behaves as a ~55 kDa protein in SDS-PAGE, though a relative molecu-

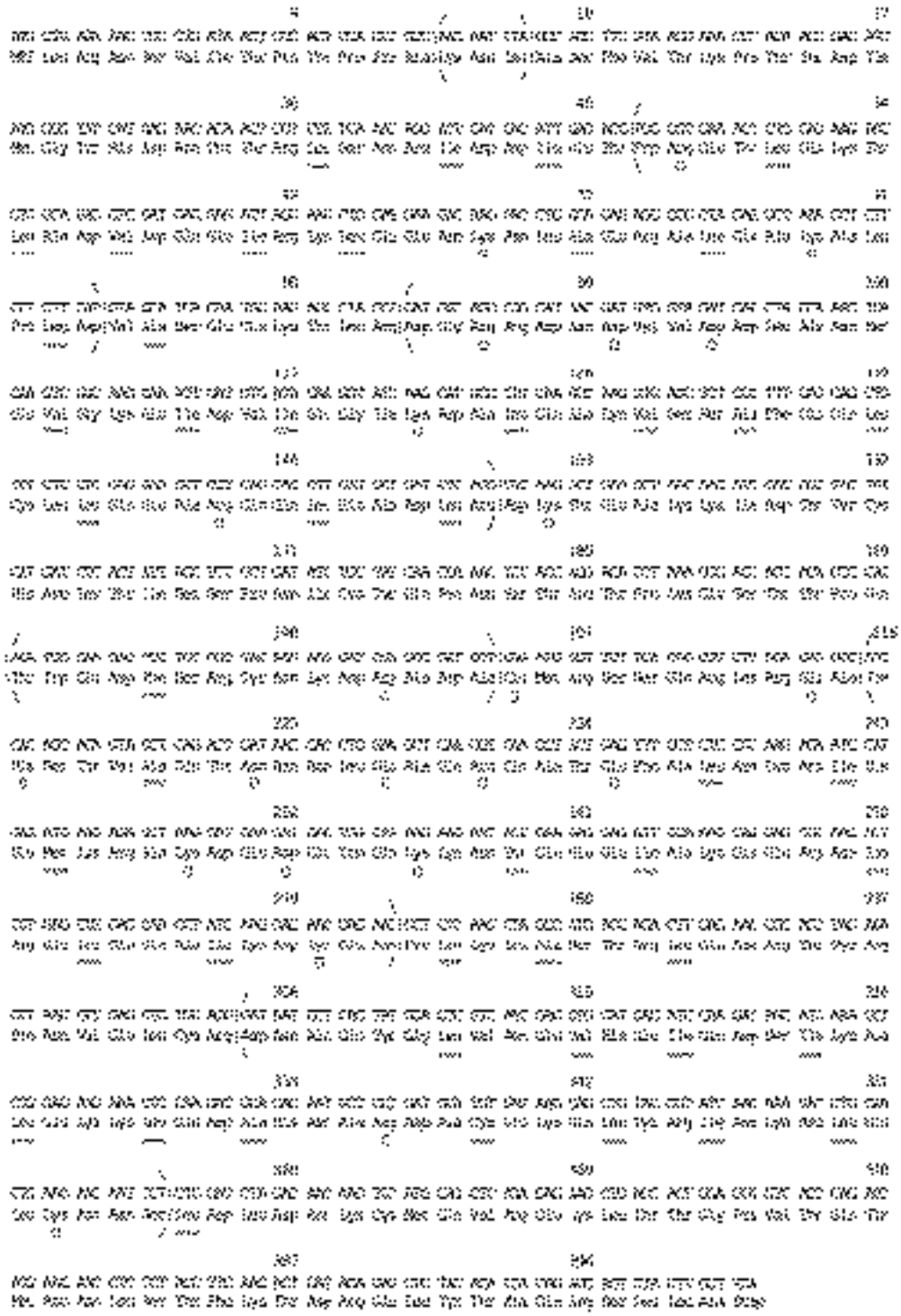


Fig. 4. The predicted amino acid sequence of tektin B1, deduced from the cDNA clone tek-B1. The largest open reading frame beginning with the first Met codes for a protein with a relative molecular mass of 46,260; however, it is possible that the reading frame begins earlier, producing an additional N-terminal sequence (see text). Brackets demarcate regions predicted to be α-helical. Certain of these regions show clear heptapeptide repeats (where the first and fourth residues within a heptad are non-polar), indicative of coiled-coil structure; the first and fourth positions of heptapeptides are marked: hydrophobic and uncharged residues are double-underlined; circles mark negatively and positively charged residues.

lar mass of 52,955 is predicted from its cDNA (Norrander et al., 1992). Alternatively, tek-B1 may be lacking a short N-terminal piece of the tektin B sequence. In this case, the 34 residues possibly encoded by bases 2-103 of Fig. 3 would merely increase the size of the non-helical N-terminal domain. Since there would be no change in the pre-

Table 1. Amino acid sequence of tektin B peptide fragments and of corresponding sequences predicted from the cDNA

Residues 207-245:

```
                                                    ?
Peptide     RSSQRLREAIHSTVAQTDN?LEAQRQATEFALRKRIHET
cDNA        RSSQRLREAIHSTVAQTDNDLEAQRQATEFALRKRIHEM
```

Residues 289-234:

```
                                                L
                                  I   H   F   FW
?AF? ?
Peptide     TRLENRTYRPNVELXRDNAQYGLVNEVXEIQDSIKALEKKLFDFFN-
FRDA?E
cDNA        TRLENRTYRPNVELCRDNAQYGLVNEVHEIQDSIKALEKKLQDAHNAR-
DACE
```

Residues 380-400:

```
Peptide     NNLSTFKTDRELYTAQRSLLA
cDNA        NNLSTFKTDRELYTAQRSLLA
```

Residues numbers are from Fig. 4. Peptide refers to amino acid sequences of CNBr fragments obtained from the sperm flagellar tektin B protein. cDNA refers to the amino acid sequences deduced from matching regions of the tekB1 cDNA for the ciliary form of tektin B. Ambiguities or uncertainties (?) in the peptide sequences are given. The only differences between the peptide and the cDNA-deduced sequences are underlined.

dicted rod domain, our structural analysis of tektin B and comparison with tektin A would not be affected (see Discussion).

DISCUSSION

As summarized in the Introduction, available evidence indicates that tektins A, B and C exist as extended, fibrous polymers associated with a particularly stable segment of the A-microtubule wall, i.e. the Sarkosyl-insoluble ribbon of protofilaments. We are ultimately interested in the function of tektins in microtubules. As a first step, to understand the potential interactions among the tektins and between tektins and tubulin, we have analyzed the primary and secondary structure of the deduced amino acid sequence of tektin B1, and we have compared the structures of tektins A and B.

The predicted secondary structure of tektins

Our analysis of the predicted secondary structures of the two proteins is summarized in Fig. 5, which suggests that the tektin A and B polypeptide chains each consist of several α-helical coiled-coil regions connected by spans of non-helical linkers. Predictions of α-helix (see Materials and Methods) are shown at the top of each panel in Fig. 5(a) and (b). Breaks in the helical structure are expected where each curve dips below the zero line. The potential for α-helical coiled-coil formation was determined by the method of Lupas et al. (1991), as shown in the lower plots in Fig. 5(a) and (b). This method scores each stretch of sequence according to its ability to form an α-helix with non-polar amino acids lined up along a slowly winding helical surface, which would allow a pair of polypeptides to

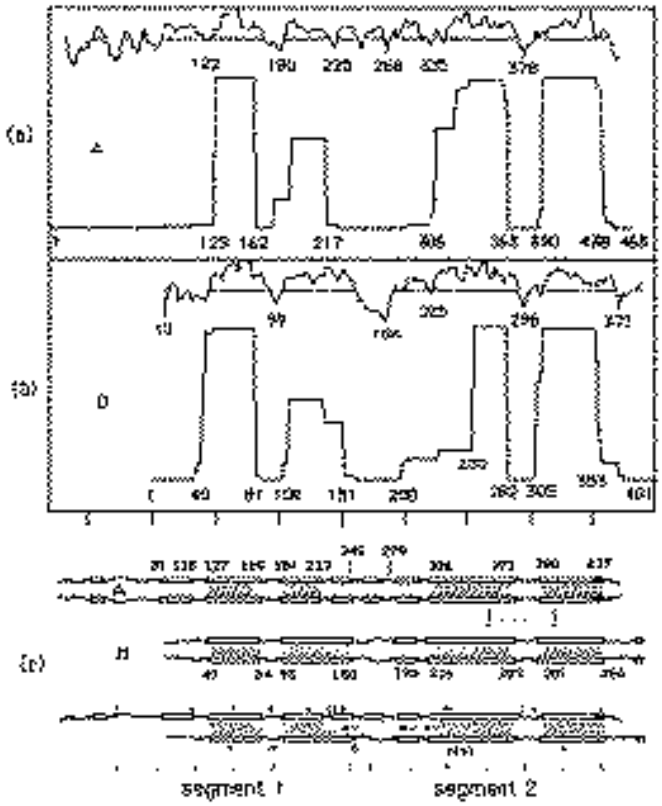
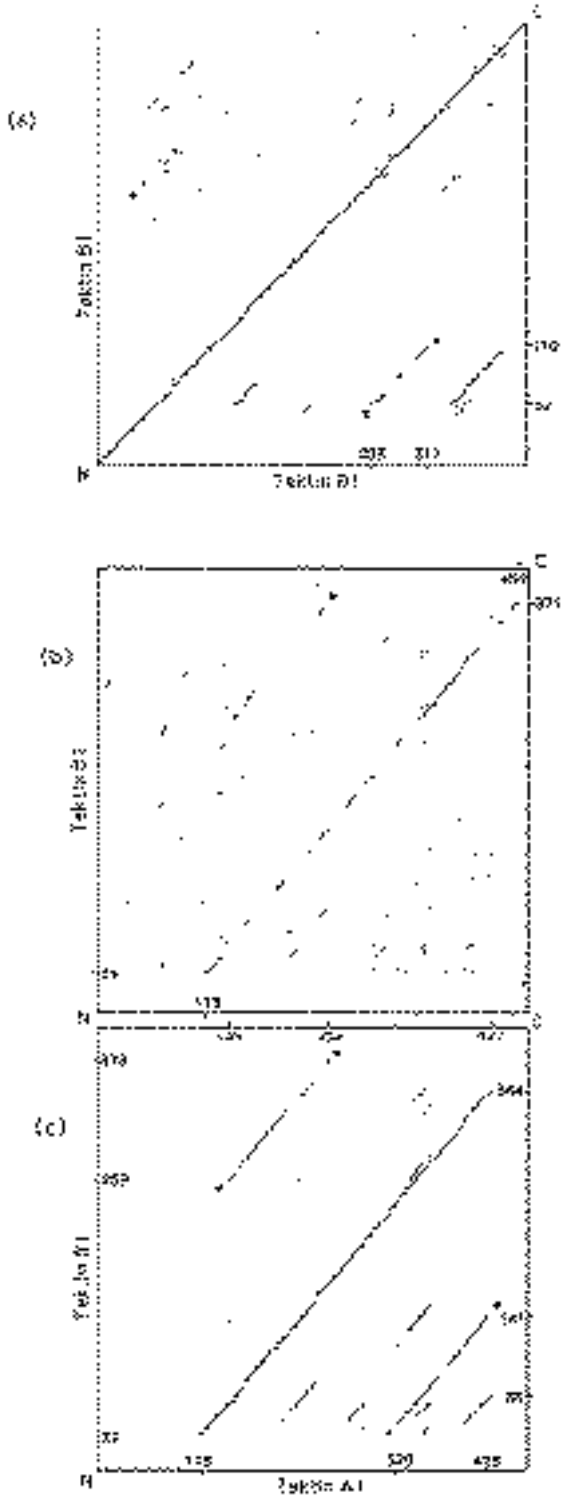


Fig. 5. Predicted secondary structure of tektin A (a) and tektin B (b). At the top of each panel the probability of α-helix is plotted versus position in sequence, as determined by the program ANALYSEP (Staden, 1988), based on the criteria of Robson and Suzuki (1976). Residue numbers are given for the probability minima. At the bottom of each panel the probability of coiled-coil structure is plotted for corresponding regions (for homodimers of tektin A1 or tektin B1), as determined by the program of Lupas et al. (1991), using a 28-residue window. Residue numbers are given for the start and end of every stretch of high probabilities. The alignment between A and B reflects the significant homology between the two sequences (see Figs 6 and 7). Note the correspondence between the prediction of α-helix and coiled-coil in at least four major regions of each tektin polypeptide. (c) Models of tektin A and B homo- and hetero-dimers, based on the data from a and b above, and shown as if they were linear molecules ~50 nm long. Predicted regions of α-helix are shown as open rectangles, non-helical links as irregular lines; brackets [....] demarcate a region of strong consensus between tektins A and B. Predicted regions of hydrophobic bonding (shown by cross-hatching) correspond to the peaks in the Lupas-vanDyke-Stock plots. Homology searches indicate a significant homology between half segments 1 and 2 of each tektin chain (see text and Fig. 6). The shaded region at the C terminus of segment 2 of tektin A matches half the consensus sequence in segment 2B of IFP (Norrander et al., 1992). Residue numbers at the ends of the predicted helical regions are shown alongside the A and B homodimers. The +/− signs between the subunits of the putative heterodimer indicate where hydrogen bonding might supplement the hydrophobic association (compare with Fig. 8). On either side of the subunits, + signs mark sites where two or more positively charged residues occur in tandem. The divisions on the scale at the bottom represent approximate 4 nm intervals.

associate as a coiled-coil. By this method the four major potential α-helical stretches of each tektin are predicted to form a coiled-coil structure (Fig. 5(c)), though there are a few intervening short potential α-helices that may not associate in the same way. In particular, predicted helices at the start of both segment 1 and segment 2 in tektin A (residues 87-115 and 279-298, respectively) and at the start of segment 2 in tektin B (residues 190-205) have too high a ratio of charged to hydrophobic residues for hydrophobic forces to hold identical regions together. (The possible significance of this is discussed below.) The four major regions of coiled-coil predicted for both tektins A and B are in similar relative positions, so already there would appear to be a structural homology between tektins A and B, and indeed they do show significant primary sequence homology (see below). The coiled-coil models for tektins are consistent with the observed ~70% combined α-helical content of tektins A and B (Linck and Langevin, 1982) and the strong α-type X-ray patterns of tektin filaments (Beese, 1984).



Internal sequence homology within tektin B1

The results of a search for internal homology within tektin B1 using the DIAGON program (Staden, 1982; see Materials and Methods) is shown in Fig. 6(a). This analysis, in which a span of 11 residues was used to assess short-range similarities and a span of 49 residues (approximately the size of an individual section of coiled-coil) for longer-range homologies, seems to indicate that the first and second halves (segments 1 and 2) of the polypeptide chain are only weakly homologous, except for two short runs, i.e.

DEEIRKLE vs EEEIAKQE and ERALEAKALPLDVA vs EQAIKDKENPLKLA.

Indeed, only an average of 22% of the residues are identical when long central stretches selected from each half are compared (see Fig. 7). Nevertheless, such levels of homology must be significant when maintained over a long distance (i.e. 27 identities plus 18 highly similar residues in a pair of 124-residue stretches with no inserted gaps). Oddly, the comparison of tektin B1 with tektin A1 (see below) provides stronger evidence for the homology between the two halves of tektin B1.

Sequence homology between tektins A and B

Using the DIAGON program as above, we also searched for sequence homologies between tektins A1 and B1, both for the short-range 11-residue span (Fig. 6(b)) and for the

Fig. 6. DIAGON plots (Staden, 1982) of amino acid homologies for tektin B. (a) Comparison of tektin B1 with itself. Above the diagonal line, short-range homology scores (for different stretches of 11 residues) are plotted. A point here indicates a match that has a probability of only 0.0006 of occurring by chance alone. The level of short-range homology is not high. Below the diagonal line, long-range homology scores (for stretches of 49 residues) are plotted, indicating values that have a probability of 0.00001 of occurring by chance. Even if only a minority of residues match within two long stretches of sequence, a relatively high score occurs here if successive matches are exactly in register. The off-diagonal lines (marked by asterisks) reveal significant homologies between the two halves (segments 1 and 2) of tektin B1. (b) Short-range (11 residue) comparison of tektin B1 with tektin A1. There are several regions of close homology between tektin A1 and B1, particularly in the C-terminal region. (c) Long-range (49-residue) comparison of tektin B1 and tektin A1. The near-continuous diagonal line reveals a strong long-range homology throughout the central rod regions of the two proteins, as further illustrated in Fig. 7. Off-diagonal lines (marked by asterisks) correspond to homologies between different helical segments; in particular, there is significant homology between opposite halves of the A1 and B1 rod domains, e.g. between segment 1 of tektin A and segment 2 of tektin B.

long-range 49-residue span (Fig. 6(c)). By this analysis there is a high probability of short-range and long-range homology between tektin A (residues 113-423) and B (residues 32-344). Furthermore, there is significant homology between segment 1 of tektin A and segment 2 of tektin B, and between segment 1 of tektin B and segment 2 of tektin A. This analysis strongly suggests that sea urchin tektins A and B have a common evolutionary origin; it also suggests that tektins evolved from an early ancestral, half-length gene, which duplicated to produce the full-length sea urchin tektin genes. Since there is more homology between heterologous segments of tektins A and B than between segments within tektin B, it would appear that the gene duplication event preceded the divergence of tektins A and B. Investigations of tektins from evolutionarily simpler organisms should help to confirm this idea.

The sequence homology between tektins A and B is examined at the level of amino acid residues in Fig. 7. Given the homology between tektin segments indicated by the DIAGON plots, Fig. 7 shows segments 1 of tektins A and B aligned over segments 2 of tektins A and B. In this format one can more easily appreciate the amino acid identities (34%) and conservative substitutions (20%) between tektins A and B within a span of 380 residues. The homology between tektins A and B in segment 2 (residues 353-403 of tektin A) is particularly striking, with 36 identities in 51 residues. Within this region the longest run of identities is RPNVELCRD.

Comparison of tektins and intermediate filaments proteins

Except for its homology with tektin A1, tektin B1 has no high degree of homology with other protein sequences available through EMBL and GenBank as of June 1993. At most, tektin B shares ~13% sequence identity with intermediate filament proteins (IFP). We have previously noted a number of biochemical and immunological similarities between tektins and the IFP (see Introduction). At the level of primary sequence, tektin A1 was found to be distinctly different from the IFP superfamily (Norrander et al., 1992). First, tektin A1 lacks the IFP consensus sequence LNDRL(or F)AXYI at the start of the IFP rod domain 1A. Tektin B1 also lacks this sequence. Tektin A1 has only part of the IFP consensus sequence LD(or E)XEIAXYRKL-LEGEEXR(or K) at the C terminus, lacking the LLEGE-EXR motif. Within this region tektin B1 has no significant homology with either IFP or tektin A. Yet, despite these major differences, tektins A and B appear to have an overall structural pattern rather similar to IFP (see Fig. 5c); i.e. both tektins and IFP are predicted to have a central rod composed of helical segments joined by non-helical linkers, and the length and positions of these domains are approximately equal (compare Norrander et al., 1992, with Conway and Parry, 1988, and Steinert and Roop, 1988; Stewart, 1990, 1993). Thus, tektins are a class of proteins distinct from IFP, but the possibility remains that tektins and IFP are related by a common distant ancestor. Verifi-

Fig. 7. Detailed comparison of amino acid sequences of tektins A1 and B1. Both halves of each rod domain are co-aligned. The regions most strongly predicted to form coiled-coil are marked by double dashes (= = =); see Fig. 5. Within these main rod segments, the number of inserted blanks required to obtain good alignment is minimal, as expected for the type of long-range homology revealed by Fig. 6. * indicates an exact match; | is drawn between residues similarly charged or with equivalent hydrophobic properties. Many other pairs of residues contribute to the homology scores of Fig. 6, which are based on the frequency with which given amino acids are observed to replace one another in known protein families (Dayhoff, 1982).

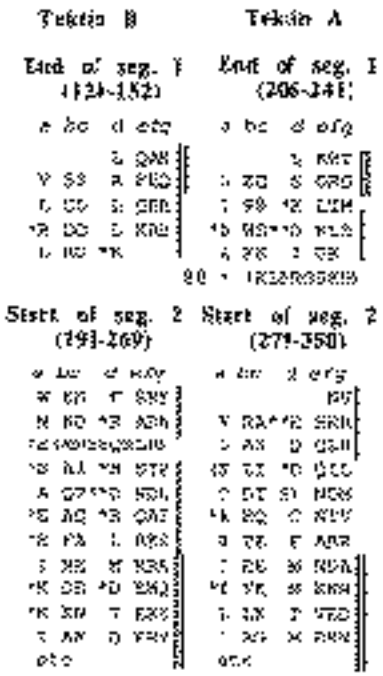


Fig. 8. Sections of the tektin A and B sequences predicted to form α-helix (marked by vertical lines) but where the heptapeptide hydrophobic repeat (indicated by double vertical lines) is heavily interrupted by polar residues in positions *a* or *d* (preceded by asterisks); i.e. the single vertical lines indicate helix without hydrophobic residues, and breaks in lines indicate that helix is not strongly predicted. Residues in parentheses may possibly form non-helical linkers; in particular, this model assumes that the non-helical sequence in tektin A1 (residues 218-225, marked @@), with no counterpart in tektin B, loops out from the coiled-coil. Most charges in *a* and *d* would be cancelled by neighboring charged residues in a homodimer, while those marked with double asterisks would not.

cation of such an evolutionary relationship might come from sequence studies of tektins and nuclear lamins in evolutionarily more primitive organisms.

Structural interactions of tektins

The mechanisms by which tektins associate to form extended polymers and interact with tubulin are not known. Filament preparations can be obtained that are composed of equimolar amounts of tektins A and B only (unpublished observations), with the possibility of these filaments being either homopolymers or heteropolymers. The only (circumstantial) evidence available is consistent with a heteropolymer model. First, it has so far not been possible to separate tektins A and B without solubilizing both subunits. Second, tektins A and B have significantly different pI values, i.e. ~6.9 and ~6.2, respectively. In these respects, tektins would appear more similar to keratin types I and II, which are obligate heteropolymers, than to IF protein type III (e.g. desmin, GFAP and vimentin), which form homopolymers (Albers and Fuchs, 1992; Coulombe and Fuchs, 1990, Hatzfeld and Weber, 1990; Quinlan et al., 1986; Steinert, 1990). If and how tektin C is associated with the core filaments of tektins A and B are not understood at this time.

The basic units of tektin polymers are assumed to be polypeptide dimers, which could be homo- or heterodimers, given the close structural similarity predicted for the two polypeptides (Fig. 5c). In either case, we assume that two polypeptides assemble in parallel to form a polar dimer, as is the case for all known coiled-coil structural proteins (Lupas et al., 1991). Tektins differ from IF proteins in the center of their rod domains by having minor helices, which apparently lack sufficient hydrophobic residues to bond them. However, it is possible that if certain negatively charged residues in positions *a* or *d* of a heptapeptide on one strand were associated with positively charged residues in positions *a* or *d* on the other strand, then these sections of rod might be held together by ionic bonds.

This possibility is explored in Fig. 8. In tektin B, the arrangement of charged residues in heptad positions *a* and *d* at the beginning of segment 2 would bring most of them close to an opposite charge on the neighboring strand in a *homodimer*. However, Asp224 (marked by a double asterisk), which would lie next to Glu228 on an α-helix, would not be counter-balanced in a homodimer. In a *heterodimer*, Asp224 of tektin B might bond to Arg309 of tektin A. Similarly, Glu284 of tektin A might be compensated for by Arg201 of tektin B, and Asp232 of tektin A by Arg142 of tektin B. Thus, it is possible that tektin AB heterodimers might be more stable than either type of homodimer. Crosslinking studies of tektin oligomers, which are underway, should answer this question.

While tektins probably interact with tubulin (see Introduction), direct evidence is lacking. By way of comparison with the interactions of tropomyosin and actin, a 14-fold periodicity due to the arrangement of negatively charged and non-polar amino acids is present in tropomyosin, which interacts with ~7 actin subunits along the actin filament (McLachlan and Stewart, 1976). However, neither tektin A (Norrander et al., 1992) nor tektin B has any obvious periodicities, apart from their 1 nm non-polar heptapeptide repeats, that would indicate potential interactions with tubulin monomers or dimers, with axial spacings of 4 and 8 nm, respectively. If they are linear molecules, tektin dimers, whether homo- or heterodimer, would measure nearly 50 nm long, a distance that could span six tubulin dimers or 48 nm along a protofilament. The strong homology between segments 1 and 2 further creates the possibility of a ~24-nm spacing in tektin filaments. It may worth noting that the approximate 24 and 48 nm spacings potentially associated with tektins are multiples of the tubulin dimer repeat and correspond to spacings of dynein arms and subspacings of radial spokes (Amos and Klug, 1974; Darnell et al., 1986). Current studies are designed to elucidate the mechanism of tektin polymerization and to investigate whether tektins interact with tubulin and possibly with other axonemal components.

This work was supported by US Public Health Service grant GM35648 and Minnesota Medical Foundation grant SMF-527-86 to R.W.L. We thank Dr Eric Eccleston and Mr Bruce McKenzie in the Microchemical Facility at the University of Minnesota for the amino acid sequencing, and Mr Dave Lee for facilitating our work within the Department of Cell Biology and Neuroanatomy. The sequence reported in this paper has been deposited in the GenBank data base (accession no. L21838).

REFERENCES

Albers, K. and Fuchs, E. (1992). The molecular biology of intermediate filaments. *Int. Rev. Cytol.* **134**, 243-279.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1989). *Molecular Biology of the Cell*. Garland Publishing, New York.

Allen, C. and Borisy, G. (1974). Structural polarity and directional growth of microtubules of *Chlamydomonas* flagella. *J. Mol. Biol.* **90**, 381-402.

Amos, W. B., Amos, L. A. and Linck, R. W. (1986). Studies of tektin filaments from flagellar microtubules by immunoelectron microscopy. *J. Cell Sci. Suppl. 5*, 55-68.

Amos, L. A. and Klug, A. (1974). Arrangement of subunits in flagellar microtubules. *J. Cell Sci.* **14**, 523-549.

Beese, L. (1984). Microtubule structure: an analysis by X-ray diffraction. Ph.D. thesis, Brandeis University, Waltham, MA.

Berget, S. M. (1984). Are U4 small nuclear ribonuclear proteins involved in polyadenylation? *Nature* **309**, 179-182.

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

Conway, J. F. and Parry, D. A. D. (1988). Intermediate filament structure: 3. Analysis of sequence homologies. *Int. J. Biol. Macromol.* **10**, 79-98.

Coulombe, P. A. and Fuchs, E. (1990). Elucidating the early stages of keratin filament assembly. *J. Cell Biol.* **111**, 153-169.

Darnell, J., Lodish, H. and Baltimore, D. (1986). *Molecular Cell Biology*, pp. 786-787. Scientific American Books, Inc., NY.

Dayhoff, M. O. (1972). *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Spring, MD.

Hatzfeld, M. and Weber, K. (1990). The coiled coil of in vitro assembled keratin filaments is a heterodimer of type I and II keratins: Use of site-specific mutagenesis and recombinant protein expression. *J. Cell Biol.* **110**, 1199-2010.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 1-359.

Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Meth. Enzymol.* **155**, 156-165.

Huynh, T. V., Young, R. A. and Davis, R. W. (1985). Constructing and screening cDNA libraries in λgt10 and λgt11. In *DNA Cloning: A Practical Approach,Volume 1* (ed. D. M. Glover), pp. 49-78. IRL Press, Oxford.

Linck, R. W. (1976). Flagellar doublet microtubules: fractionation of minor components and α-tubulin from specific regions of the A-tubule. *J. Cell Sci.* **20**, 405-439.

Linck, R. W. (1990). Tektins and microtubules. *Advan. Cell Biol.* **3**, 35-63.

Linck, R. W., Amos, L. A. and Amos, W. B. (1985). Localization of tektin filaments in microtubules of sea urchin sperm flagella by immunoelectron microscopy. *J. Cell Biol.* **100**, 126-135.

Linck, R. W., Goggin, M. J., Norrander, J. M. and Steffen, W. (1987). Characterization of antibodies as probes for structural and biochemical studies of tektins from ciliary and flagellar microtubules. *J. Cell Sci.* **88**, 453-466.

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

Linck, R. W. and Stephens, R. E. (1987). Biochemical characterization of tektins from sperm flagellar doublet microtubules. *J. Cell Biol.* **104**, 1069-1075.

Lupas, A. Van Dyke, M. and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

Marston, F. A. O. (1987). The purification of eukaryotic polypeptides expressed in *Escherichia coli*. In *DNA Cloning: A Practical Approach, Volume 3* (ed. D. M. Glover), p. 59. IRP Press, Oxford.

Marston, F. A. O., Lowe, P. A., Doel, M. T., Schoemaker, J. M., White, S. and Angal, S. (1984). Purification of calf prochymyosin (prorennin) synthesized in Escherichia coli. Bio/Technology **2**, 800-804.

McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallog.* **38**, 871-873.

McLachlan, A. D. and Stewart, M. (1976). The 14-fold periodicity in α-tropomyosin and the interaction with actin. *J. Mol. Biol.* **103**, 271-298.

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

Putney, S. D., Benkovic, S. J. and Schimmel, P. R. (1981). A DNA fragment with an α-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo. Proc. Nat. Acad. Sci. USA* **78**, 7350-7354.

Quinlan, R. A., Hatzfeld, M., Franke, W. W., Lustig, A., Schulthess, T. and Engle, J. (1986). Characterization of dimer subunits of intermediate filament proteins. *J. Mol. Biol.* **192**, 337-349.

Robson, B. and E. Suzuki. (1976). Conformational properties of amino acid residues in globular proteins. *J. Mol. Biol.* **107**, 327-356.

Rüther, U. and Müller-Hill, B. (1983). Easy identification of cDNA clones. *EMBO J.* **2**, 1791-1794.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci. USA* **74**, 5563-5567.

Staden, R. (1982). An interactive graphics program for comparing and aligning nucleic-acid and amino-acid sequences. *Nucl. Acids Res.* **10**, 2951-2961.

Staden, R. (1988). Methods to define and locate patterns of motifs in sequences. *Comput. Appl. Biosci.* **4**, 53-60.

Steffen, W. and Linck, R. W. (1988). Evidence for tektins in centrioles and axonemal microtubules. *Proc. Nat. Acad. Sci. USA* **85**, 2643-2647.

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

Steffen, W. and Linck, R. W. (1989b). Tektins in ciliary and flagellar microtubules and their association with other cytoskeletal systems. In *Cell Movement*, vol. 2 (ed. J. R. McIntosh and F. D. Warner), pp. 67-81. Alan R. Liss, Inc., New York.

Steffen, W. and Linck, R. W. (1990). The presence of tektin-related components in centrosomes of *Spisula solidissima*. *J. Cell Biol.* **111**, 181a.

Steffen, W. and Linck, R. W. (1992). Evidence for a non-tubulin spindle matrix and for spindle components immunologically related to tektin filaments. *J. Cell Sci.* **101**, 809-822.

Steinert, P. M. (1990). The two-chain coiled-coil molecule of native epidermal keratin intermediate filaments is a type I-type II heterodimer. *J. Biol. Chem.* **265**, 8766-8774.

Steinert, P. M. and Roop, D. R. (1988). Molecular and cellular biology of intermediate filaments. *Annu. Rev. Biochem.* **57**, 593-625.

Stephens, R. E. (1977). Differential protein synthesis and utilization during cilia formation in sea urchin embryos. *Dev. Biol.* **61**, 311-329.

Stephens, R. E. (1989). Quantal tektin synthesis and ciliary length in sea-urchin embryos. *J. Cell Biol.* **92**, 403-413.

Stephens, R. W., Oleszko-Szuts, S. and Linck, R. W. (1989). Retention of ciliary ninefold structure after removal of microtubules. *J. Cell Sci.* **92**, 391-402.

Stewart, M. (1990). Intermediate filaments: Structure, assembly and molecular interactions. *Curr. Opin. Cell Biol.* **2**, 91-100.

Stewart, M. (1993). Intermediate filament structure and assembly. *Curr. Opin. Cell Biol.* **5**, 3-11.

Towbin, H., Staehlin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Nat. Acad. Sci. USA* **76**, 4350-4354.

Young, R. A. and Davis, R. W. (1983). Yeast RNA polymerase II genes: isolation with antibody probes. *Science* **222**, 778-782.

(Received 15 June 1993 - Accepted 7 July 1993)