Journal of Cell Science 111, 2967-2975 (1998)
Printed in Great Britain © The Company of Biologists Limited 1998
JCS4612

Common and variant properties of intermediate filament proteins from lower chordates and vertebrates; two proteins from the tunicate *Styela* and the identification of a type III homologue

Dieter Riemer and Klaus Weber*

Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, Am Fassberg 11, D-37077 Goettingen, Germany
*Author for correspondence

Accepted 31 July; published on WWW 9 September 1998

SUMMARY

The chordates combine the vertebrates and the invertebrate phyla of the cephalo- and urochordates (tunicates). Two cytoplasmic intermediate filament (IF) proteins of the urochordate *Styela plicata* are characterized by cDNA cloning, gene organization, tissue specific expression patterns in the adult animal and the self assembly properties of the recombinant proteins. In line with metazoan phylogeny St-A and St-B have the short length version of the coil 1b domain found in all vertebrate and cephalochordate IF proteins while protostomic IF proteins have the longer length version with an extra 42 residues. St-A is the first IF protein from a lower chordate which can be unambiguously related to a particular vertebrate IF subfamily. St-A shares 46% sequence identity with desmin, displays the N-terminal motif necessary for filament assembly of type III proteins and forms normal homopolymeric 10 nm filaments in vitro. St-A but not St-B is present in smooth muscle cells of the body wall musculature. St-A and St-B are found as separate networks in some interior epithelia. St-B shares 30 to 35% identity with keratin 8, St-A and desmin and does not form IF under in vitro assembly conditions. Its relation to a particular vertebrate IF type or to the eight currently known IF proteins from the cephalochordate *Branchiostoma* remains unresolved. The striking relation between St-A and desmin predicts that the common progenitor of the urochordate (tunicate) and the cephalochordate/vertebrate lineages already possessed a type III homologue. Unlike in vertebrates intron patterns cannot be used to classify the tunicate IF genes. Although St-A is a type III homologue its gene shows an intron position which in vertebrates is restricted to keratin type II genes.

Key words: Cephalochordate, Chordate, Desmin, Deuterostome, Intermediate filament, Tunicate, Urochordate, Vertebrate

INTRODUCTION

Intermediate filaments (IF) form a flexible cytoskeleton which resists mechanical stress. Mutations that weaken this framework can lead to cell rupture and cause a variety of hereditary human disorders including the epidermal keratin diseases (McLean and Lane, 1995; Fuchs and Cleveland, 1998). The common structural principle of all IF proteins is a central alpha helical rod domain of coiled coil forming ability which is flanked by hypervariable head- and tail domains. The rod is divided into subdomains (coils 1a, 1b, 2a and 2b) which are connected by short non helical linkers. True consensus sequences occur at the ends of the rod domain and are involved in filament assembly and stability. Biochemical properties, protein sequences, cell and tissue specific expression patterns and the organization of the corresponding genes divide the 50 members of the vertebrate IF protein family into 5 subfamilies. Types I and II cover the two distinct keratin types necessary to form the obligatory heteropolymeric keratin filaments of the epithelia and the epidermal appendages. Desmin, vimentin, glial fibrillary acidic protein and peripherin form the type III, which can form homopolymeric IF. Type IV describes the various neurofilament proteins. Types I to IV are cytoplasmic IF proteins while type V covers nuclear proteins, the lamins. They differ from the vertebrate cytoplasmic IF proteins by an extra 42 residues (6 heptads) in their coil 1b subdomain and by unique tail domains which harbor a nuclear localization signal and in most cases a terminal CaaX box (Fuchs and Weber, 1994; Parry and Steinert, 1995). The two constituent proteins of the mammalian eye lens filaments (Merdes et al., 1993) seem to lie outside types I to IV. The four IF subfamilies extend from the mammals to fish (Druger et al., 1994; Glasgow et al., 1994).

In contrast to vertebrates, protostomic phyla display cytoplasmic IF proteins which have an extra 42 residues in their coil 1b domain and in most cases harbor a lamin homology segment of 120 residues in their tail domain. This was originally documented for molluscs, nematodes and annelids (Weber et al., 1988, 1989; Dodemont et al., 1990, 1994; Szaro et al., 1991; Way et al., 1992; Tomarev et al., 1993; Adjaye et

al., 1995; Bovenschulte et al., 1995; Johansen and Johansen, 1995) and more recently the long coil 1b version has also been found in eight further protostomic phyla (D.R., K.W., A. Erber and M. Bovenschulte, unpublished results). Parallel work showed that the short coil 1b version present in all vertebrate IF proteins also holds for the eight cytoplasmic IF proteins established for the cephalochordate *Branchiostoma* (Riemer et al., 1992, 1998). Since no molecular information is available for cytoplasmic IF proteins from echinoderms the short coil 1b version is either a property shared by all deuterostomia or a property restricted to the chordates. Thus given the relatively recent origin of the short coil 1b version a key question is whether the early chordates have clearly identifiable homologues of the type I to IV IF subfamilies established for vertebrates. While these may be difficult to detect in the current collection of cephalochordate IF proteins (Riemer et al., 1998) we show here that one of the two urochordate IF proteins cloned in this study is a true homologue of type III IF proteins.

MATERIALS AND METHODS

Animals

Styela plicata was collected around Los Angeles (CA, USA), frozen in liquid nitrogen and stored at −80°C.

Preparation of RNA, genomic DNA and isolation of cDNA clones

Total cellular RNA was prepared from 1 g of frozen animals and poly(A)$^+$ RNA was isolated (Riemer et al., 1998). Preparation of genomic DNA was as described (Riemer et al., 1992). A plasmid cDNA library representing total poly(A)$^+$ RNA was constructed using the Zap Express cDNA kit (Stratagene, Heidelberg, Germany) and *Eco*RI-*Xho*I digested pBluescriptII vector (Stratagene) with *E. coli* XLIBlue as bacterial host. An initial screening at reduced stringency (Dodemont et al., 1994) of approximately 100,000 primary transformants with a nick translated cDNA probe comprising the entire rod domain of *Branchiostoma* IF-1 (Riemer et al., 1992) yielded a full length clone of 2,138 bp coding for St-A protein. In a second screen using the St-A rod domain as cDNA probe the St-B cDNA clone (2,282 bp) was identified. In northern blotting experiments glyoxylated poly(A)$^+$RNA was probed with the full length cDNA inserts from St-A and St-B under conditions of high stringency (Dodemont et al., 1990).

Isolation and characterization of St-A and St-B genes

Restriction fragments comprising *St-A* and *St-B* genes were detected by southern blot hybridization of *Bgl*II- and *Sal*I digested genomic DNA at conditions of high stringency (Dodemont et al., 1990). Entire *St-A* and *St-B* cDNA inserts were used as probes. Genes were isolated from subgenomic libraries established with gel purified 4.3 kb and 7.5 kb *Pvu*II (*St-A*) and 9.3 kb *Sal*I (*St-B*) fragments. Library construction and screening with appropriately cut and dephosphorylated pUC18 vector DNA were as described (Riemer et al., 1992). Extensive restriction enzyme maps were made for the cloned genomic DNA fragments, relevant fragments were isolated and subjected to sequence analysis.

Expression and purification of St-A and St-B oligo histidine tagged polypeptides; production of antibodies

For production of a St-A fusion polypeptide comprising amino acid residues E100 to P265 St-A cDNA was amplified by PCR using primers 5′-CTGAATTCCAGGGACAGCGAATTAATG-3′ (sense, nt 371 to 397) and 5′-CGAAGAGCTTCTTCAAGCTTTGGTGC-3′ (antisense, nt 896 to 871). A 507 bp restriction fragment was isolated upon digestion with *Eco*RI and *Hin*dIII. To produce a polypeptide covering coil 1b and part of coil 2 of the St-B protein a 383 bp *Xho*I-*Pvu*II cDNA fragment (amino acid residues L84 to L214) was isolated. Both fragments were cloned in frame into appropriately digested pRSET vectors (Invitrogen, Leek, The Netherlands), expressed in *E. coli* JM109 and affinity purified on a Ni^{2+}-NTA column (Qiagen, Hilden, Germany) for antibody production (Riemer et al., 1995). Rabbits were immunized according to standard methods. To remove cross reacting activities St-B antiserum was preabsorbed on recombinant St-A protein (see below) bound to cyanogenbromide activated Sepharose beads (Pharmacia, Uppsala, Sweden). St-A antiserum was used directly at high dilutions. Specificity of antisera was tested by immunoblots.

Protein gel electrophoresis and western blotting

Total protein extracts from whole animals were prepared with Trizol reagent (Gibco-BRL, Eggenstein, Germany). Precipitated proteins were boiled in sample buffer. Polypeptides separated by SDS-PAGE in 10% gels were transferred to nitrocellulose membranes and incubated with the following antibodies for 1 hour at room temperature. Monoclonal antibody IFA (Pruss et al., 1981) as a 1:30 dilution of hybridoma supernatant; St-A rabbit antiserum diluted 1:6,000; preabsorbed St-B antiserum (dilution 1:500). Immunoreactive polypeptides were visualized with biotinylated secondary antibodies and horseradish peroxidase-conjugated to Streptavidin (Amersham, Braunschweig, Germany). IFA reacts with many but not all IF proteins (Bartnik and Weber, 1989; Riemer et al., 1991; Johansen and Johansen, 1995).

Immunofluorescence microscopy

Sections from frozen animals were prepared for immunofluorescence microscopy as described (Riemer et al., 1998). Sections were incubated with St-A rabbit antiserum (1:450 dilution) or double labeled with monoclonal IFA antibody used as undiluted hybridoma supernatant and preabsorbed St-B rabbit antiserum (1:15 dilution) followed by incubation with the appropriate secondary antibodies. DNA was stained with Hoechst 33258 dye (Hoechst, Frankfurt, Germany).

Expression of St-A and St-B proteins in *E. coli*

For expression of the full length St-A polypeptide a 1,404 bp *Eco*RI-*Xho*I fragment from the St-A cDNA was subcloned in pKK388-1 vector (Clontech, Heidelberg, Germany) creating an expression construct covering amino acid residues E100 to S490 of the St-A protein. PCR amplification of St-A cDNA with primers 5′-CTACACAACCATGGCTCAACAACAAAGCAGC-3′ (sense, nt 66 to 93) and 5′-CGCTGTCCCTGGAATTCAGTTCC-3′ (antisense, nt 389 to 367) provided the entire head domain and introduced a novel *Nco*I restriction site at the putative start codon. After digestion with *Nco*I and *Eco*RI a 299 bp restriction fragment was isolated and ligated into the appropriately digested pKK388/E100-S490 expression construct to generate a plasmid that drives expression of full length St-A polypeptide in *E. coli* JM109. To express the full length St-B polypeptide the coding sequence of St-B cDNA was amplified by PCR with primers 5′-CAGTTACCAAGCCATGGTTGGTCGTG-3′ (sense, nt 198 to 223) and 5′-GACTATAGTCTAGAGGCTTGACCG-3′ (antisense, nt 1525 to 1502) creating an *Nco*I site at the 5′ end and an *Xba*I site flanking the 3′ end of the coding sequence. The 1,314 bp *Nco*I-*Xba*I restriction fragment was ligated into pKK388-1 vector DNA and expressed in *E. coli* JM109.

Purification of recombinant proteins and filament formation

Recombinant proteins, highly enriched in the inclusion body preparation, were solubilized in 8 M urea containing 10 mM Na_2HPO_4, 1 mM 2-mercaptoethaonl, pH 6.6. Proteins were purified using Mono S cation exchange chromatography in 8 M urea buffer

A

```
   1                                                                                   M  S  Q  Q  S  S  K  R  F  L  E  E  R  T  S
   1 TCTAGAGGATCCCCAGACAAGCTTCAAAAAAGTTCATTTATTTCAAAGAATTGTCTGAAAGATAACTACACAACGATGTCTCAACAAAGCAGCAAAAGATTTCTTGAAGAAAGAACCTCA

  16 S  Y  R  R  N  F  G  K  N  R  S  L  F  D  Y  G  T  P  R  T  G  G  S  Y  Q  S  Y  S  R  Y  Q  T  S  G  G  A  G  G  A  V
 121 AGTTACAGGAGAAATTTCGGAAAAAACCGATCTCTTTTTGACTATGGAACACCACGTACTGGTGGCTCATACCAAAGCTACAGCAGATACCAAACTTCAGGGGGTGCCGGTGGAGCAGTT

  56 G  G  A  V  G  G  A  T  G  G  H  R  I  V  Q  E  R  V  S  V  L  R  G  A  P  V  H  S  G  I  S  M  K  D  I  A  S  M  E  E
 241 GGAGGTGCAGTGGGAGGGGCAACTGGTGGACACAGAATTGTACAGGAACGTGTGTCTGTTTTGAGAGGAGCACCCGTACATAGCGGGATTAGTATGAAAGACATCGCATCTATGGAAGAA

  96 S  M  G  T  E  F  Q  G  Q  R  I  N  E  K  A  E  L  Q  Q  L  N  S  R  F  A  E  Y  I  T  K  V  R  N  L  E  Q  A  N  R  V
 361 AGTATGGGAACTGAATTCCAGGGACAGCGAATTAATGAAAAAGCTGAACTTCAACAGCTTAACAGCAGATTTGCTGAATACATCACAAAGGTGAGAAATCTTGAGCAAGCAAACAGAGTT

 136 L  E  E  Q  V  A  H  L  S  S  T  K  P  S  R  L  A  E  T  Y  E  E  E  L  N  R  L  R  R  E  I  E  K  L  T  N  E  R  S  S
 481 CTTGAAGAACAAGTTGCTCACTTGTCATCAACAAAACCATCAAGATTGGCAGAAACATACGAAGAAGAACTCAACAGACTTCGTCGTGAAATAGAAAAACTTACTAACGAGCGATCCTCT

 176 L  A  L  Q  L  E  N  A  F  I  D  V  A  K  W  K  D  M  Y  E  E  E  Q  V  T  R  K  E  V  E  D  D  L  A  G  M  R  K  D  C
 601 CTTGCTCTGCAATTGGAGAATGCATTTATTGACGTTGCTAAGTGGAAAGATATGTATGAAGAGGAACAAGTCACCAGGAAGGAAGTTGAAGATGATCTTGCAGGAATGCGTAAAGATTGT

 216 D  D  A  T  L  V  R  L  D  L  E  R  R  L  E  T  L  Q  E  E  I  E  F  L  K  K  A  H  Q  Q  E  V  E  E  L  E  D  R  I  R
 721 GATGATGCAACTCTTGTTCGACTTGATCTTGAGAGAAGACTTGAAACCTTGCAGGAGGAAATTGAGTTCTTGAAGAAAGCTCATCAACAAGAAGTTGAGGAATTGGAGGACAGAATCAGA

 256 S  T  E  I  K  I  E  T  T  P  A  P  D  L  E  E  A  L  R  D  V  R  A  Q  Y  E  N  I  A  R  K  N  R  E  D  A  E  K  W  Y
 841 TCAACCGAAATCAAGATCGAGACAACACCCGCACCAGATCTTGAAGAAGCTCTTCGTGATGTCCGTGCTCAATACGAAAATATTGCTCGTAAGAACAGAGAAGATGCTGAGAAATGGTAT

 296 E  D  K  V  V  N  L  K  S  Q  A  L  H  N  E  E  A  M  R  A  V  K  N  E  M  S  E  Y  R  K  N  V  Q  T  L  T  L  E  I  D
 961 GAAGACAAGGTTGTAAACCTGAAGTCTCAAGCACTTCACAACGAAGAGGCAATGAGGGCTGTCAAGAATGAAATGAGCGAATACAGAAAGAACGTACAGACTCTCACCCTCGAGATTGAC

 336 S  M  R  G  S  N  E  A  L  Q  R  N  L  G  D  L  E  D  R  Y  N  R  D  M  D  A  S  Q  D  T  I  N  G  L  Q  A  E  C  D  D
1081 TCAATGAGAGGATCGAATGAAGCGCTTCAAAGAAATCTTGGTGATTTGGAAGACAGATACAACAGAGATATGGATGCAAGTCAAGATACCATCAATGGTTTGCAGGCTGAGTGTGACGAC

 376 L  K  T  Q  M  A  Q  H  L  R  Q  Y  Q  E  L  M  D  V  K  M  A  L  D  L  E  I  A  T  Y  R  K  L  L  E  G  E  E  V  R  I
1201 TTGAAAACTCAAATGGCTCAACATCTTCGTCAATACCAAGAACTTATGGATGTCAAGATGGCGCTTGACCTTGAAATTGCAACTTACAGAAAACTTTTGGAAGGAGAAGAAGTTAGAATC

 416 S  D  K  V  Q  T  M  R  T  S  V  P  V  Y  S  S  I  Q  R  S  Y  G  G  R  Y  S  D  V  Q  Q  E  T  A  D  S  R  H  T  D  D
1321 AGCGACAAAGTCCAAACCATGAGAACTTCAGTACCAGTTTACTCATCAATTCAGAGATCATATGGAGGAAGATATTCAGATGTTCAACAAGAAACCGCTGATTCCCGACACACCGATGAC

 456 V  E  T  V  T  T  K  K  V  V  V  K  T  I  E  T  K  D  G  K  V  V  S  Q  T  E  D  V  R  E  Q  S  T  D  S  ***
1441 GTTGAGACGGTCACGACCAAGAAAGTAGTTGTCAAAACAATCGAGACGAAAGACGGCAAGGTTGTCAGTCAGACTGAAGATGTTCGTGAACAATCAACCGACTCTTAAACTACTTGCGTT

1561 AATTGGAAATCGCTTTTCTACTAGCCGGAATTTATTTAGATAATACTGTAAAATTTCTTATTGTAAACTTTTAAAATAGAAATAAAATTGAGGTTTTTTACAACCGTGATTGATTTGTAA
1681 ATGAATAAAAACTAAATGAACGGATTTTACTTTTCCGAATTGTTCGCTTCTAAATGGTTTGCGGAGGAATGTTATTTTTAGGTCGTTTTGTTTACATCTAGATGAAGTATTTTTTGATTG
1801 TTTTGGTTTGTAATTGACCTTTTGTTATTTCTTTTGAAACATATTGTTGACAATGTATAGTTTTATAATTTAATTTGTTACCAATATACTCCCGATAAGAATATCTATCTTGTCAAGTTT
1921 GAACCACAGATAAGCGGTCTGATAGACGCGTTTTATCTGTGGATCCTACTTGAAAATAATTATGAGCTGCTTAAATAATTGTTTAAAAACTAGTATTTCTTGTCTATCACATTTTGAAAA
2041 TGAAATTCTGTTTATATAATGTCTTGATTATGCTTTGTTGTAGAAAATATTAACAAATAAAAAATTGATCCTCAAAAAAAAAAAAAAAAAA
```

B

```
   1 CTTAATAGGTTTGTAGTTTAATGTTCGATAGTCCGTTTGGGAAGGCAGAGTACGTAACATCAACACCGAAATCTTATTTGCTGAGCCAAGAGAAGACGAAGAATATAGGAACAAAAGGAA

   1                                                                                                 M  F  G  R  G  H  K  G  G  H
 121 GCAAGTTACAACAGTTATGATGGAGTCAGCGGTGGAAAGGTCACGTACCAGTCGACGTCGGTTTCGTATGGGCCATCCAGTTACCAAGCTATGTTTGGTCGTGGACACAAGGGCGGGCAT

  11 V  A  Y  G  R  L  S  R  Y  G  Y  G  I  E  G  P  I  Q  S  V  S  V  S  E  S  F  A  Q  L  P  G  I  L  S  T  R  S  E  E  K
 241 GTTGCGTATGGAAGGTTGAGCAGATATGGATACGGAATTGAAGGACCAATTCAGTCGGTGTCGGTGAGTGAGAGTTTTGCTCAACTACCTGGGATTTTGAGCACACGAAGTGAAGAGAAG

  51 Q  E  L  G  V  L  N  N  R  F  A  T  Y  I  D  K  V  R  R  L  E  S  Q  N  K  A  L  A  T  R  V  T  E  L  E  S  T  K  S  V
 361 CAAGAATTAGGGGTCCTGAACAATCGGTTTGCCACCTACATTGATAAAGTCAGAAGATTGGAATCACAAAACAAAGCTCTGGCGACTAGAGTCACCGAGCTCGAGTCTACCAAGAGCGTC

  91 V  S  R  S  G  D  I  Y  D  D  E  L  A  R  L  R  R  E  I  E  K  L  T  H  D  K  A  D  V  E  I  Q  L  H  N  C  Y  Q  E  L
 481 GTAAGTAGATCTGGTGATATTTACGACGATGAACTTGCAAGGCTCAGACGTGAGATTGAGAAGTTAACGCATGATAAAGCAGATGTTGAGATTCAACTTCATAATTGCTACCAAGAATTG

 131 K  D  Y  E  K  K  L  K  A  E  S  D  A  R  R  D  A  E  K  L  V  K  S  L  R  K  D  V  D  D  A  T  L  A  R  I  D  L  E  R
 601 AAGGACTACGAGAAAAAGCTAAAAGCAGAGAGTGATGCAAGACGTGACGCGGAAAAATTGGTGAAATCTCTTCGAAAAGACGTTGATGATGCAACTCTAGCAAGGATTGATTTGGAAAGG

 171 K  L  E  T  L  Q  E  E  L  E  L  L  K  A  T  T  S  E  D  I  E  V  L  K  S  Q  V  T  V  K  H  E  E  V  F  D  A  P  A  P
 721 AAGTTGGAAACCCTCCAAGAAGAGCTTGAACTTCTCAAGGCTACAACTTCTGAGGACATCGAAGTGTTGAAATCTCAAGTAACAGTGAAACATGAAGAAGTATTCGATGCTCCAGCTCCG

 211 S  A  D  L  T  E  S  L  R  D  I  R  V  A  Y  E  Q  L  S  K  S  N  A  Y  D  V  E  K  V  Y  K  N  T  I  A  D  L  Q  Q  Q
 841 TCAGCTGATTTGACCGAGTCTTTGAGAGATATTCGTGTTGCTTACGAACAACTTTCAAAGAGCAATGCTTATGACGTCGAAAAAGTTTACAAAAATACGATTGCCGATCTTCAACAACAG

 251 I  R  N  S  N  Q  A  L  G  D  A  K  S  A  L  M  D  T  R  R  Q  L  Q  T  V  T  V  E  I  E  G  L  R  S  T  N  S  S  L  E
 961 ATCAGAAACAGCAACCAGGCACTTGGCGATGCCAAGTCTGCTCTCATGGATACTCGACGTCAATTGCAAACCGTGACTGTTGAAATTGAAGGATTGAGAAGCACTAACTCATCGTTGGAA

 291 G  Q  I  A  E  L  Q  D  R  Q  D  K  E  S  E  Q  Y  Q  A  R  I  T  E  L  E  S  E  L  Q  K  C  R  D  N  M  A  R  H  L  A
1081 GGACAGATTGCTGAATTACAAGATCGTCAAGATAAAGAAAGCGAGCAATATCAAGCAAGAATTACTGAATTGGAAAGCGAATTACAGAAATGCAGAGATAATATGGCACGTCATCTTGCT

 331 D  Y  N  K  L  M  N  I  K  L  S  L  D  L  E  I  S  T  Y  R  K  L  L  E  G  E  E  G  R  I  S  R  N  T  S  M  E  V  S  G
1201 GATTACAACAAGTTGATGAACATCAAATTATCACTCGATCTTGAGATTTCAACTTACAGAAAATTATTGGAAGGAGAAGAGGGAAGGATAAGTCGAAACACAAGCATGGAAGTAAGTGGA

 371 Y  D  T  I  E  K  S  V  T  V  Q  S  V  S  V  S  S  T  K  S  V  E  S  H  P  K  I  L  V  R  E  E  I  T  T  R  S  G  S  S
1321 TACGACACAATTGAGAAGAGTGTGACAGTGCAATCTGTGTCGGTCTCATCAACTAAATCAGTTGAAAGTCATCCAAAGATTCTGGTGAGGGAAGAAATAACTACCAGATCAGGATCAAGT

 411 S  S  S  S  S  S  G  S  D  S  D  ***
1441 TCGTCCTCATCTTCTTCTGGATCAGATTCTGACTAAGAAGTATATACTTATCTCCGATCCACGGTCAAGCCTTCACACTATAGTCTATTAGCATGTTCAAAATAGTTAATTTTGTCTGAA

1561 ATTGCCACGCACGAGCTACATATAGATGCAAAGATGCAAATTCGAATCTCGACAAAGGCAATAAAGATAATATTGTGAAATCTTTGCTGGACAGAAGTCAGTGGTGAATTATGTTTGATA
1681 GAGATCAAGTTTCATTGATTTCAAAGTTATTAAACAAATTTTGGTTTAATCGTGAATTTACGTACCGATATGTTAATCTGTAGGGATAACTTTTCTCAAGCCTCAGTTCGCCAATAAGCA
1801 CCGCCAAAACATTGTGGTGAATGAATAAAAAAATGTGTTATTTAGTTGATCCATGTGTTATTTTCCGACTTTACAACGTTGCATTGTCACTTTGGCTGGTAGAAAGTTCAGAGACTCGTT
1921 AAATTAAATAATTTCGATTACTTATTGTTATGCATTCGGAAAATGTCATACATAACTTCGAAATGTGTGGGTGATAGGTGGTTTATGTTTTGACTTTATAGAAATATGAAAGAAATCATC
2041 ATTAAAAAAAATTAACCGCAATTAACCGCATAGCATATAAATTGAAATTCGCCGTTTCTTAGGTCTATTTTTGTACTTCCATGATGTGGTTTAAATTCGTCAAATATTAATAACAATATA
2161 CTAATATAAATGCTTCTTTGTGCTTGTGCTTTTGCTCGTCATAATATTCAAATCACAATTTTGTAAATTGTATCACTAATCAAACACACACTAGTAAAAACTGTAAAAAAAAAAAAAAA
```

Fig. 1. Nucleotide and predicted amino acid sequences of the cDNAs for two tunicate IF proteins. (A and B) St-A and St-B from *Styela plicata*. The start codons are underlined and the stop codons are marked by asterisks. Arrowheads mark the position of introns in the corresponding genes (see also Fig. 4). Sequences of the exon / intron boundaries follow consensus sequences. The lengths of the 11 introns in St-A are 0.33, 0.28, 0.69, 0.64, 0.53, 0.86, 1.23, 0.75, 0.53, 0.72 and 0.66 kb, respectively. Corresponding values for the 9 introns in St-B are 1.45, 3.4, 0.39, 0.14, 0.39, 0.33, 0.24, 0.38 and 1.8 kb, respectively. The sequences are available from EMBL/GenBank under accession numbers AJ005020 (St-A) and AJ005021 (St-B).

and a gradient formed between 0 and 0.4 M NaCl. Fractions were monitored by SDS-PAGE. Aliquots of the pure proteins (40 µl; 0.2 mg/ml protein) were dialyzed at room temperature for three or more hours against filament buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM 2-mercaptoethanol, pH 7) using dialysis filters (Millipore, Eschborn, Germany). Structures formed were analyzed by electron microscopy after negative staining with 2% uranylacetate.

RESULTS

Characterization of cDNAs encoding cytoplasmic IF proteins

A plasmid cDNA library from total adult *Styela plicata* was screened at reduced stringency with a cDNA probe from the rod domain of *Branchiostoma lanceolatum* IF-1 (Riemer et al., 1992). Eight weakly positive signals were obtained upon colony hybridization of 100,000 primary transformants. Southern analysis of the inserts recovered from purified plasmid DNAs was used to select the longest cDNA of 2,138 bp for sequence analysis. The cDNA, designated St-A, contains a single reading frame of 1,470 bp flanked by 5′- and 3′ untranslated sequences of 75 and 593 bp, respectively. A canonical polyadenylation signal is found 18 bp upstream of a poly(A) remnant (Fig. 1). The open reading frame translates into a protein of 490 amino acid residues. The calculated molecular mass and isolelectric point of St-A are 56,232 and 4.93, respectively.

In a second screen a probe from the rod domain of St-A was used at reduced stringency. Of 60 positive signals 50 persisted after washing to high stringency as putative St-A clones. Among the 10 positive colonies with a lower affinity for the St-A probe a cDNA of 2,282 bp was identified as St-B by sequence analysis. The full length cDNA is flanked by 210 and 808 bp of untranslated sequences. The 3′-untranslated region shows a poly (A) tract located 27 bp downstream of an AATCAA polyadenylation signal. The predicted protein has a length of 421 amino acid residues (Fig. 1). The calculated molecular mass and isoelectric point of St-B are 47,466 and 5.18, respectively.

In northern analysis of total mRNA, St-A and St-B cDNA inserts detected single mRNAs of 2.5 and 2.4 kb, respectively (data not shown). The differences in signal intensities indicate that the St-A transcript is nearly five times more abundant than the St-B transcript.

St-A protein is a tunicate homologue of vertebrate type III proteins

The sequences of St-A and St-B have the characteristic

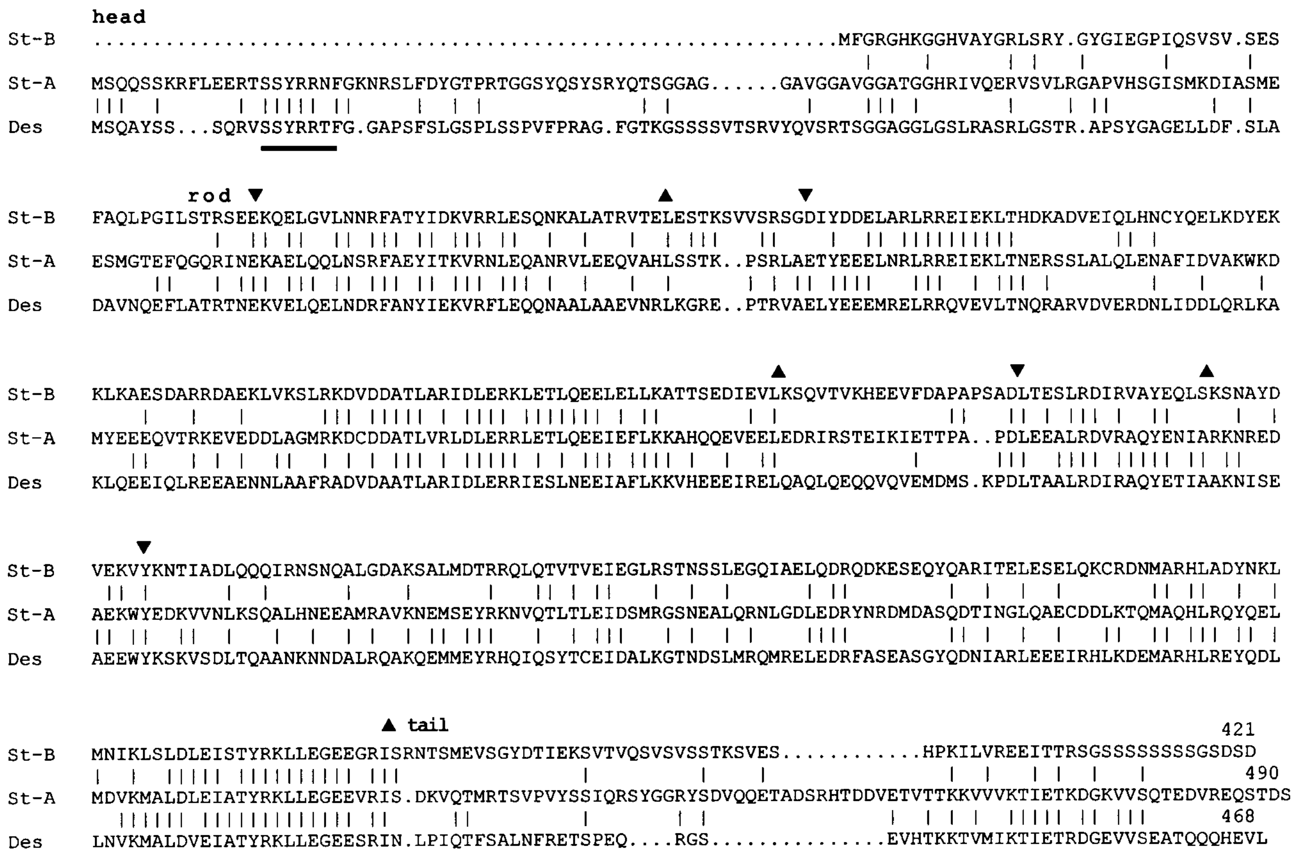

Fig. 2. Sequence alignment of the tunicate IF proteins St-A and St-B and hamster desmin. The three domain organization (head, rod and tail domain) is marked. Arrowheads pointing down or up delineate the starts and ends of the α helical segments with coiled coil forming ability (coils 1a, 1b, 2a and 2b) which are connected by short linkers. Identical residues in St-B (upper line) and St-A (middle line) or St-A and hamster desmin (lower line) are marked by vertical lines. The desmin sequence is from Quax et al. (1985). The two tunicate sequences are from Fig. 1. Dots indicate gaps introduced to optimize the alignments. The N-terminal motif sequence necessary for IF formation by type III proteins is underlined. It is present in desmin and St-A but not in St-B.

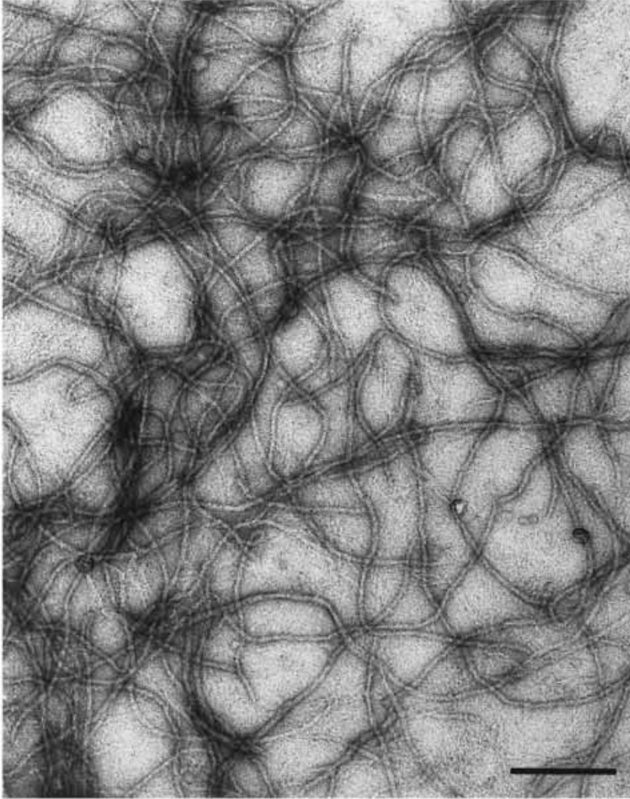

Fig. 3. Electron micrograph of IF assembled from recombinant St-A in vitro. St-A was purified in the presence of 8 M urea (see Materials and Methods) and dialyzed to standard filament buffer (10 mM Tris-HCl, 50 mM NaCl, 1 mM 2-mercaptoethanol) without urea. Aliquots of samples were negatively stained with 2% uranylacetate. The preparation shows typical and very long IF. Bar, 0.2 µm.

tripartate structural organization (head-rod-and tail domain) of IF proteins (Fig. 2). St-A and St-B show the short length version of the coil 1b subdomain of the rod as do the vertebrate type I to IV and the *Branchiostoma* IF proteins while all nuclear lamins and the protostomic cytoplasmic IF proteins have an extra 42 residues in coil 1b (for references see Introduction).

Over its entire sequence St-A is clearly related with vertebrate type III proteins, particularly with desmin and vimentin. St-A and hamster desmin share 46% identity (Fig. 2) while the various type III proteins (desmin, vimentin, GFAP and peripherin) share 55 to 65% identity. The sequence homology extends through the rod domain. In addition the carboxyterminal tail domain shows a stretch of 22 residues of high homology and the head domain displays the motif SSYRRXF found in similar location in all type III proteins (Fig. 2). This motif is involved in the 10 nm filament formation by type III proteins (Hatzfeld et al., 1992; Hermann et al., 1992).

The relation of the second tunicate protein to other IF proteins is much less clear. St-B and St-A share 32% identity (Fig. 2) indicative of different protein types. St-B and hamster desmin share 35% identity but in keratin II molecules this value is reduced to 30% (human keratin 8) while keratin I molecules (human keratin 18) show only 26% identity. St-B lacks the aminoterminal motif of type III proteins and St-A necessary for filament formation. Interestingly there is also no clear relation between St-B and any of the four IF prototypes of the cephalochordate *Branchiostoma* (Riemer et al., 1998).

St-A forms homopolymeric IF in vitro

The sequence predictions on St-A and St-B were tested in functional terms with recombinant proteins. Full length proteins were expressed in *E. coli*. The proteins present in inclusion bodies were solubilized with 8 M urea and purified to homogeneity by chromatography in 8 M urea. Removal of the urea by dialysis resulted in the formation of long and smooth IF in the case of St-A (Fig. 3). In contrast St-B did not form filaments under a variety of conditions. Only irregular aggregates were observed.

Isolation and organization of the St-A and St-B genes

Since vertebrate type I to IV IF genes show distinct and type specific intron patterns (Fuchs and Weber, 1994) we explored whether the tunicate protein St-B can be related to a specific vertebrate IF type by the organization of its gene. Genomic DNA from *Styela plicata* was digested with several restriction enzymes and hybridized to full length St-A and St-B cDNAs under conditions of high stringency. Two *Pvu*II fragments of 4.3 and 7.5 kb spanning the entire *St-A* gene were detected while the *St-B* gene was isolated in a single 9.3 kb *Sal*I fragment (data not shown). Hybridization patterns were consistent with the presence of single genes for both *St-A* and *St-B*. Genomic fragments were isolated from size fractionated genomic DNA libraries. All exons were fully sequenced and introns were identified by aligning genomic and cDNA sequences. Intron sizes in the *St-A* gene range from 276 bp to 1.23 kb compared with 144 bp to 3.4 kb in the *St-B* gene (Fig. 1). All exon/intron boundaries confer with the consensus 5′- and 3′-splice site sequence. Intron positions are summarized in Fig. 4.

The organization of *St-A* and *St-B* genes is clearly unrelated to vertebrate type IV neurofilament genes and shows a relatively good relation to vertebrate type I to III genes although a direct connection to a single of these three types cannot be made (Fig. 4). Both genes show 5 introns at homologous positions with vertebrate type I to III genes and 6 (*St-A*) or 4 (*St-B*) additional introns at unique positions which have no counterparts in currently known metazoan IF genes. With 11 (*St-A*) and 9 (*St-B*) introns the *Styela* genes are clearly richer in introns than type I to III genes with their usual 7 to 8 intron patterns. The tunicate genes are the only chordate IF genes with an intron in the head domain. Curiously the gene for *St-A*, which by protein sequence and homopolymeric IF formation is a type III homologue (see above), shows an intron in coil 1a which in vertebrate genes is absent from type III genes and only found in keratin type II genes (Fig. 4).

Intron positions in the tail domain of vertebrate type I to III genes are not precisely conserved in position (Fig. 4), sometimes not even in a particular gene type. However, the first intron is always situated in phase 1 (types I to III) while the second intron, when present, uses phase 3 (type III). In this context *St-A* relates to type III genes and *St-B* to types I plus II.

Tissue specific expression of St-A and St-B

For studies using frozen sections of *Styela* in

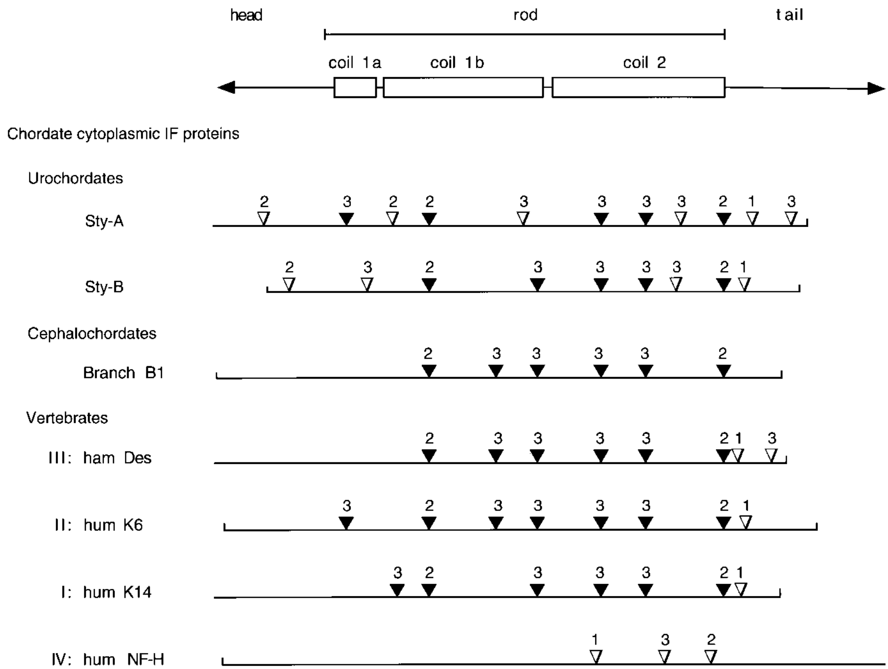

Fig. 4. Summary of intron positions for chordate IF genes. The tripartate structural organization of the proteins is indicated at the top. Solid triangles specify homologous intron positions; non-related intron positions are marked by open triangles. Numbers above the triangles give the phase of the introns. 1, 2 and 3 indicate that the intron is inserted after the first, second or third nucleotide of the corresponding codon. The urochordate genes for St-A and St-B proteins are from Fig. 1. The cephalochordate gene for B1 (Riemer et al., 1992) represents also the *Branchiostoma* genes for A1, A2, B2 and D1 although in these 4 genes the tail encoding region was not characterized. A minor variation on the intron pattern is found in *Branchiostoma* C2 where again the tail encoding region was not analyzed (Riemer et al., 1998). Vertebrate type I to IV genes are illustrated by the genes for human keratin 14 (hum K14), human keratin 6 (hum K6), hamster desmin (ham Des) and human neurofilament protein NF-H (hum NF-H). For references to these and additional vertebrate IF genes see Dodemont et al. (1990).

immunofluorescence microscopy rabbit antibodies raised against oligo histidine tagged proteins containing partial rod domains of St-A or St-B proteins were used. Monospecificity of both antisera was demonstrated on immunoblots with total protein extracts from *Styela*. St-A antiserum detected a single band of ~60 kDa, preabsorbed St-B antiserum reacted with a polypeptide of ~50 kDa (Fig. 5). Monospecificity of both antisera was also documented in immunoblotting experiments with recombinant full length St-A and St-B proteins (not shown). The monoclonal antibody IFA detected only the 60 kDa St-A protein and did not recognize St-B (Fig. 5). The murine monoclonal antibody IFA was therefore taken together with rabbit St-B antibodies for double labelling experiments in immunofluorescence. The strong reaction of St-A with the monoclonal antibody IFA is in agreement with cDNA cloning experiments which lead to a partial clone of St-A (M. E. White and W. R. Jeffery, personal communication).

Strong St-A reactivity was seen in body wall musculature (Fig. 6) which consists of smooth muscle tissue (Burighel and Cloney, 1997) while a lower level of reactivity was observed in atrial and pharyngeal epithelia. St-B reactivity measured by a specific antibody occured exclusively in epithelial cells of the atrium and pharynx and not in muscle. Although St-A and St-B decoration was observed in internal epithelia their exact fibrous staining patterns did not coincide indicating the presence of two independent filament systems (Fig. 6H,I). Neither IFA or St-A and St-B antibodies decorated the epidermis (Fig. 6). Thus epidermal IF seem to be built from IF molecules differing from St-A and St-B.

DISCUSSION

We have cloned two cytoplasmic IF proteins (St-A, St-B) from the tunicate *Styela plicata*, determined the organization of their genes, monitored their cell and tissue specific expression patterns by immunofluorescence microscopy and studied the self assembly properties of the purified proteins.

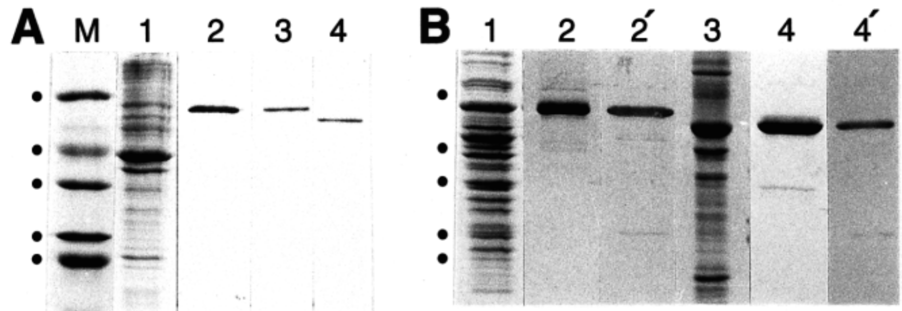

Fig. 5. SDS-PAGE and immunoblot analysis. (A) Lane 1, total protein extract from *Styela* stained with Coomassie Brilliant Blue. Lanes 2, 3 and 4 show the corresponding immunoblot reactions after incubation with murine monoclonal IFA (2), St-A rabbit antiserum (3) and preabsorbed St-B antiserum (4). Note that IFA reacts only with St-A but not with St-B. Lane M, marker proteins with molecular mass of 66,000 (BSA), 45,000 (ovalbumin), 36,000 (glyceraldehyde phosphate dehydrogenase), 29,000 (bovine carbanhydrase), and 24,000 (bovine trypsinogen). (B) Lanes 1 and 2, Coomassie stained protein extract from bacteria expressing St-A (1) and purified St-A protein (2). Lane 2′ shows the corresponding immunoblot reaction with St-A antiserum. Lanes 3 and 4, Coomassie stained protein extract from bacteria expressing St-B and purified St-B (4). Lane 4′ shows the corresponding imunoblot reaction with purified St-B antiserum.

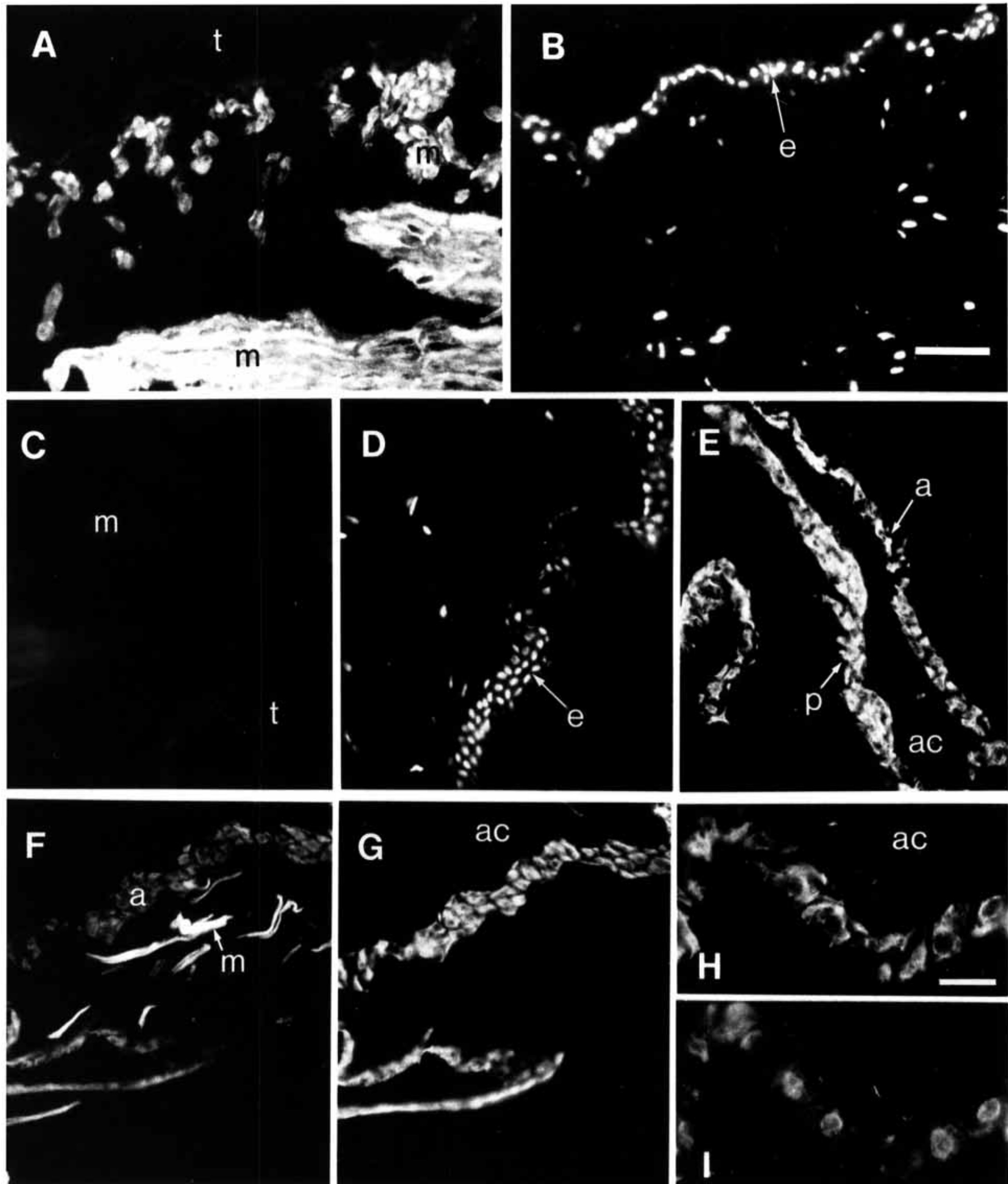

Fig. 6. Immunofluorescence microscopy on frozen sections of *Styela plicata.* Sections were fixed in acetone and incubated with St-A rabbit antiserum (A), purified St-B antiserum (C,E,G,H) and monoclonal murine IFA antibody (F,I). Staining of DNA was with Hoechst 33258 (B,D). Body wall musculature (m) reacts strongly with St-A antiserum (A) and is negative with St-B antiserum (C). St-B antiserum recognizes St-B protein in atrial (a) and pharyngeal (p) epithelia (E). Monoclonal antibody IFA, which reacts only with St-A and not with St-B (Fig. 5 and Results), gave the same staining as St-A antiserum. IFA antibody decorates body wall musculature and shows a significantly weaker reaction in atrial epithelium (F). St-B antiserum reveals coexpression of St-A and St-B protein on the same section (G). High power view of atrial epithelium double labelled with St-B antiserum (H) and IFA (I). Note that St-B and IFA reactivities do not colocalize in epithelial cells. The single layered epidermis (e) of *Styela* does not react with antisera to St-A (A) or St-B (C) and is also negative with IFA antibody. ac, atrial chamber, t, tunic. Magnifications: (A-G) ×250; (H,I) ×625. Bars: 50 μm (B); 15 μm (H).

St-A is the first IF protein from a lower chordate which can be unambiguously related to a particular vertebrate IF type. It shares 45 to 46% sequence identity with the type III proteins desmin and vimentin while the various vertebrate type III proteins show 55 to 65% identity. In contrast type I (keratin 18) and type II (keratin 8) proteins share only 26 and 32% identity with St-A. The sequence homology between desmin and St-A is not restricted to the rod domain but covers also the carboxyterminal part of the tail domain. Even the head domain, which is often the most variable region of IF proteins, reveals a hallmark of vertebrate type III proteins (Fig. 2). The short motif SSYRRXF is not only common to all type III proteins but is also a necessary requirement for in vitro and in vivo filament assembly (Hatzfeld et al., 1992; Hermann et al., 1992; Ralton et al., 1994). In perfect agreement with these sequence relations the purified recombinant protein forms 10 nm filaments in vitro (Fig. 3). This selfassembly process shows that St-A like vertebrate type III proteins can form homopolymeric IF. Additionally we note that St-A is richly expressed in the smooth muscle cells of the body musculature (Fig. 6).

The division of vertebrate cytoplasmic IF proteins into the subfamilies I to IV is based on sequence characteristics, intron patterns of the corresponding genes and their expression patterns. While St-A is clearly the urochordate homologue of vertebrate type III proteins by sequence criteria, homopolymeric IF formation and the expression in smooth muscle tissue, this relation is obscured in the gene organization. With 11 introns the *St-A* gene has the highest number of introns of all currently known IF genes (see Fig. 4). In addition the *St-A* and *St-B* genes are the only chordate IF genes with an intron in the head domain encoding region. The *St-A* gene has an intron position in coil 1a, which in vertebrates, is characteristic of type II keratin genes. The *St-A* gene lacks, like type I keratin genes, an intron position shared by type II and III genes (second intron in type III genes). Thus the gene organization cannot be used to relate a tunicate IF protein to a particular vertebrate IF subfamily. To some extent this problem has also emerged with the eight currently known *Branchiostoma* IF proteins. Although their divergent sequences allow for four subfamilies the corresponding genes follow essentially the vertebrate type III intron pattern (Riemer et al., 1998; see also Fig. 4). Thus intron patterns can only be used in vertebrates to define a particular IF subfamily. The very close relation in organization observed for all genes of a particular vertebrate IF type most likely reflects multiple gene duplication events typical for the onset of the vertebrate lineage as discussed for *Hox* genes (Holland and Garcia-Fernandez, 1996; Holland, 1996).

St-B, the second tunicate IF protein, is difficult to relate to any particular chordate IF-type. Vertebrate IF subfamilies I to III usually share only 30% sequence identity, and thus the 32% identity between St-A and St-B probably indicates two distinct types. Sequence identity values for St-B and vertebrate type III (desmin), II (keratin 8) and I proteins (keratin 18) are 35, 32 and 26%, respectively. The lack of IF forming ability of recombinant St-B and its exclusive presence in some interior epithelia (Fig. 6) raises the possibility that St-B could be related to vertebrate keratins, and given the sequence identity values a possible relation to keratin type II molecules can be raised. Future experiments must concentrate on cytoskeletal preparations of readily dissected tissues such as the epidermis and the epithelia containing St-B. One- and two-dimensional gels, followed by microsequencing of candidate proteins could identify additional IF proteins, which can be approached by PCR cloning methods. Once such a collection of full length IF proteins is available in recombinant form in vitro assembly studies can decide whether some of them participate only in obligatory heteropolymer filament assembly like the vertebrate type I and II keratins. Until this is done it remains unclear whether lower chordates have homologues of vertebrate keratins.

The chordate group of the deuterostomia is defined by embryological and anatomical features but a monophyletic nature of the chordates is difficult to deduce from 18S ribosomal DNA sequences because of the high rate of evolution in the urochordate lineage (Turbeville et al., 1994). However certain sequence features of the muscle actins of vertebrates, cephalochordates and urochordates separate the chordates from all other metazoa including the echinoderms and hemichordates (Vandekerckhove and Weber, 1984; Kovilur et al., 1993; Bovenschulte and Weber, 1997; Kusakabe et al., 1997). The finding that the short coil 1b length holds for all cytoplasmic IF proteins of vertebrates, cephalochordates and urochordates conforms with chordate phylogeny but provides no independent proof since we do not yet know whether the short coil 1b is restricted to chordates or reflects a feature typical of all deuterostomia (see Introduction). Phylogenetic studies also suggest that cephalochordates are the sister group of the vertebrates while the urochordates (tunicates) are the probable sister group of the clade containing vertebrates and cephalochordates (Turbeville et al., 1994; Holland 1996). The identification of St-A as a tunicate homologue of vertebrate type III IF proteins suggests that at least the type III predates the divergence of the chordate phyla. Whether this also holds for type I and II proteins can only be decided once *Styela* St-B and the currently established group of eight *Branchiostoma* IF proteins (Riemer et al., 1998) are characterized in homo- and heteropolymeric filament assembly studies. However, at least some of these IF proteins could also resemble new IF types which did not evolve in the vertebrate lineage or were lost on this branch. Thus the C1 and C2 proteins of the cephalochordate *Branchiostoma* seem unique in the entire collection of metazoan IF proteins. Their long tail domains harbor extended alpha helical regions with coiled coil forming ability (Riemer et al., 1998).

We thank T. Eisbein and J. Schünemann for excellent technical assistance and Dr N. Johnsson for a supply of *Styela plicata.* The help of Dr U. Welsch in *Styela* histology is greatly appreciated.

REFERENCES

Adjaye, J., Plessmann, U., Weber, K. and Dodemont, H. (1995). Characterisation of neurofilament protein NF70 mRNA from the gastropod *Helix aspersa* reveals that neuronal and non-neuronal intermediate filament proteins of cerebral ganglia arise from separate lamin-related genes. *J. Cell Sci.* **108**, 3581-3590.

Bartnik, E. and Weber, K. (1989). Widespread occurrence of intermediate filaments in invertebrates; common principles and aspects of divergence. *Eur. J. Cell Biol.* **50**, 17-33.

Bovenschulte, M., Riemer, D. and Weber, K. (1995). The sequence of a cytoplasmic intermediate filament (IF) protein from the annelid *Lumbricus terrestris* emphasizes a distinctive feature of protostomic IF proteins. *FEBS Lett.* **360**, 223-226.

Bovenschulte, M. and Weber, K. (1997). Deuterostomic actin genes and the definition of the chordates: cDNA cloning and gene organization for cephalochordates and hemichordates. *J. Mol. Evol.* **45**, 653-660.

Burighel, P. and Cloney, R. A. (1997). Urochordata: Ascidiacea. In *Microscopic Anatomy of Invertebrates,* vol. 15 (ed. F. W. Harrison and E. E. Ruppert), pp. 221-347. Wiley-Liss Inc. New York.

Dodemont, H., Riemer, D. and Weber, K. (1990). Structure of an invertebrate gene encoding cytoplasmic intermediate filament proteins: implications for the origin and the diversification of IF proteins. *EMBO J.* **9**, 4083-4094.

Dodemont, H., Riemer, D., Ledger, N. and Weber, K. (1994). Eight genes and alternative RNA processing pathways generate an unexpectedly large diversity of cytoplasmic intermediate filament proteins in the nematode *Caenorhabditis elegans*. *EMBO J.* **13**, 2625-2638.

Druger, R. K., Glasgow, E., Fuchs, C., Levine, E. M., Matthews, J. P., Park, C-H. and Schechter, N. (1994). Complex expression of keratins in goldfish optic nerve. *J. Comp. Neurol.* **340**, 269-280.

Fuchs, E. and Weber, K. (1994). Intermediate filaments: Structure, dynamics, function and disease. *Annu. Rev. Biochem.* **63**, 345-382.

Fuchs, E. and Cleveland, D. W. (1998). A structural scaffolding of intermediate filaments in health and disease. *Science* **279**, 514-519.

Glasgow, E., Druger, R. K., Fuchs, C., Levine, E. M., Giordano, S. and Schechter, N. (1994). Cloning of multiple forms of goldfish vimentin: differential expression in CNS. *J. Neurochem*. **63**, 470-481.

Hatzfeld, M., Dodemont, H., Plessmann, U. and Weber, K. (1992). Truncation of recombinant vimentin by ompT. *FEBS Lett.* **302**, 239-242.

Hermann, H., Hofmann, I. and Franke, W. W. (1992). Identification of a nonapeptide motif in the vimentin head domain involved in intermediate filament assembly. *J. Mol. Biol.* **223**, 637-650.

Holland, P. W. H. (1996). Molecular biology of lancelets: insights into development and evolution. *Israel J. Zool.* **42**, 247-272.

Holland, P. W. H. and Garcia-Fernandez, J. (1996). Hox genes and chordate evolution. *Dev. Biol.* **173**, 382-395.

Johansen, K. M. and Johansen, J. (1995). Filarin, a novel invertebrate intermediate filament protein present in axons and perikarya of developing and mature leech neurons. *J. Neurobiol.* **27**, 227-239.

Kovilur, S., Jacobsen, J. W., Beach, R. L., Jeffrey, W. R. and Tomlinson, C. R. (1993). Evolution of the chordate muscle actin gene. *J. Mol. Evol.* **36**, 361-368.

Kusakabe, T., Araki, I., Satoh, N. and Jeffery W. R. (1997). Evolution of chordate actin genes: evidence from genomic organisation and amino acid sequences. *J. Mol. Evol.* **44**, 289-298.

McLean, W. H. I. and Lane, E. B. (1995). Intermediate filaments in disease. *Curr. Opin. Cell Biol.* **7**, 118-125.

Merdes, A., Gounari, F. and Georgatos, S. D. (1993). The 47-kD lens-specific protein phakinin is a tailless intermediate filament protein and an assembly partner of filensin. *J. Cell Biol.* **123**, 1507-1516.

Parry, D. A. D. and Steinert, P. M. (1995). Intermediate filament structure. Springer Verlag, New York.

Pruss, R. M., Mirsky, R., Raff, M. C., Thorpe, R., Dowding, A. J. and Anderton, B. H. (1981). All classes of intermediate filaments share a common antigenic determinant defined by a monoclonal antibody. *Cell* **27**, 419-428.

Quax, W., van den Broek, L., Vre Egberts, W., Ramaekers, F. and Bloemendal, H. (1985). Characterization of the hamster desmin gene: expression and formation of desmin filaments in nonmuscle cells after gene transfer. *Cell* **43**, 327-338.

Ralton, J. E., Lu, X., Hutcheson, A. M. and Quinlan, R. A. (1994). Identification of two N-terminal non-alpha-helical domain motifs important in the assembly of glial fibrillary acidic protein. *J. Cell Sci.* **107**, 1935-1948.

Riemer, D., Dodemont, H. and Weber, K. (1991). Cloning of the non-neuronal intermediate filament protein of the gastropod *Aplysia californica*; identification of an amino acid residue essential for the IFA epitope. *Eur. J. Cell Biol.* **56**, 351-357.

Riemer, D., Dodemont, H. and Weber, K. (1992). Analysis of the cDNA and gene encoding a cytoplasmic intermediate filament (IF) protein from the cephalochordate *Branchiostoma lanceolatum*; implications for the evolution of the IF protein family. *Eur. J. Cell Biol.* **58**, 128-135.

Riemer, D., Stuurman, N., Berrios, M., Hunter, C., Fisher, P. A. and Weber, K. (1995). Expression of *Drosophila lamin* C is developmentally regulated: analogies with vertebrate A-type lamins. *J. Cell Sci.* **108**, 3189-3198.

Riemer, D., Karabinos, A. and Weber, K. (1998). Analysis of eight cDNAs and six genes for intermediate filament proteins in the cephalochordate *Branchiostoma* reveals differences in the multigene families of lower chordates and the vertebrates. *Gene* **211**, 361-373.

Szaro, B. G., Pant, H. C., Way, J. and Battey, J. (1991). Squid low molecular weight neurofilament proteins are a novel class of neurofilament protein. *J. Biol. Chem*. **266**, 15035-15041.

Tomarev, S. I., Zinovievam R. D. and Piatigorsky, J. (1993). Primary structure and lens-specific expression of genes for an intermediate filament protein and β-tubulin in cephalopods. *Biochim. Biophys. Acta* **1216**, 245-254.

Turbeville, J. M., Schulz, J. R. and Raff, R. A. (1994). Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Mol. Biol. Evol.* **11**, 648-655.

Vandekerckhove, J. and Weber, K. (1984). Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of different vertebrate muscle actins. *J. Mol. Biol.* **179**, 391-413.

Way, J., Hellmich, M. R., Jaffe, H., Szaro, B., Pant, H. C., Gainer, H. and Battey, J. (1992). A high-molecular-weight squid neurofilament protein contains a lamin-like rod domain and a tail domain with Lys-Ser-Pro repeats. *Proc. Nat. Acad. Sci. USA* **89**, 6963-6967.

Weber, K., Plessmann, U., Dodemont, H. and Kossmagk-Stephan, K. (1988). Amino acid sequences and homopolymer-forming ability of the intermediate filament proteins from an invertebrate epithelium. *EMBO J.* **7**, 2995-3001.

Weber, K., Plessmann, U. and Ulrich, W. (1989). Cytoplasmic filament proteins of invertebrates are closer to nuclear lamins than are vertebrate intermediate filament proteins; sequence characterization of two muscle proteins of a nematode. *EMBO J.* **11**, 3221-3327.