

The epidermal intermediate filament proteins of tunicates are distant keratins; a polymerisation-competent hetero coiled coil of the *Styela* D protein and *Xenopus* keratin 8

Jian Wang, Anton Karabinos, Jürgen Schünemann, Dieter Riemer, Klaus Weber¹⁾

Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, Goettingen, Germany

Received February 26, 2000

Received in revised version March 29, 2000

Accepted March 29, 2000

Coiled coil – intermediate filament protein – keratin – metazoa – urochordate

Two novel cytoplasmic intermediate filament (IF) proteins (C and D) from the tunicate (urochordate) *Styela* are characterised as putative keratin orthologs. The coexpression of C and D in all epidermal cells and the obligatory heteropolymeric IF assembly of the recombinant proteins argue for keratin orthologs, but the sequences do not directly reveal which protein behaves as a keratin I or II ortholog. This problem is solved by the finding that keratin 8, a type II keratin from man or *Xenopus*, forms chimeric IF when mixed with *Styela* D. Mutant proteins of *Styela* D and keratin 8 with a single cysteine in equivalent positions show that these chimeric IF are, like vertebrate keratin filaments, based on the hetero coiled coil. We propose that *Styela* D retains, in spite of its strong sequence drift, important molecular features of type I keratins. By inference *Styela* C reflects a type II ortholog. We discuss that type I to III IF proteins are expressed along the chordate branch of metazoa.

Introduction

Cytoplasmic intermediate filaments (IF) give rise to a flexible cytoskeleton which resists mechanical stress. Thus mutations, which weaken this framework, can lead to cell rupture as seen in a variety of hereditary human disorders including the epidermal keratin diseases (Fuchs and Cleveland, 1998). The common structural principle of all IF proteins is a central alpha helical rod domain, able to form double stranded coiled coils,

which is flanked by hypervariable head- and tail domains. Biochemical properties, protein sequences, the organisation of the genes and their cell- and tissue-specific expression patterns allow most of the 50 members of the human IF multigene family to be placed 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 and their relatives form the type III, which is able to form homopolymeric IF. Type IV covers five neuronal IF proteins while type V refers to the nuclear lamins (Fuchs and Weber, 1994; Parry and Steinert, 1995).

Metazoan cytoplasmic IF proteins come in two general prototypes which are phylogenetically defined. The cytoplasmic IF proteins from vertebrates (Fuchs and Weber, 1994; Parry and Steinert, 1995), cephalochordates (Riemer et al., 1992, 1998; Karabinos et al., 2000) and urochordates (Riemer and Weber, 1998) have the short version of the coil 1b subdomain. In contrast, the cytoplasmic IF proteins of 11 protostomic phyla show, as do all nuclear lamins (Erber et al., 1999), the long version of coil 1b due to 6 additional heptads (Weber et al., 1989; Erber et al., 1998). This difference must be of clonal origin given the whole spectrum of vertebrate IF proteins and the full complement of IF proteins in the nematode *C. elegans*. Not only the 8 *C. elegans* IF proteins earlier reported (Dodemont et al., 1994), but also the additional 3 proteins, available through the completed genome project, have the long coil 1b version. Since no sequences are available for the echinoderms, the precise origin of the short coil 1b domain remains unknown. It could be a marker of the chordate branch of metazoa or shared by all deuterostomic phyla (Erber et al., 1998). These results suggest that the precursors of the four subfamilies I to IV defined in the vertebrates (see above) arose by gene duplications and sequence drift after a common precursor gene had introduced the short coil 1b subdomain.

¹⁾ Prof. Dr. Klaus Weber, Max Planck Institute for Biophysical Chemistry, Department of Biochemistry, Am Fassberg 11, D-37077 Goettingen, Germany, e-mail: rlongo@gwdg.de, Fax: +49 551 201 1578.

Recent reports show that at least some of the four subfamilies I to IV of cytoplasmic IF proteins, originally defined for vertebrates, can be recognised in the early chordates. Molecular phylogenetic calculations of the cytoplasmic IF protein sequences from the cephalochordate *Branchiostoma* (Karabinos et al., 1998; Karabinos et al., 2000; Luke and Holland, 1999) were necessary to indicate five keratin orthologs – three of type I and two of type II. In each case this assignment was verified by demonstrating the obligatory heteropolymeric IF formation in vitro. In addition, a possible type III ortholog with a strongly drifted sequence is recognised by its homopolymeric IF-forming ability (Karabinos et al., 1998). In the urochordate *Styela* the muscle IF protein A is well related to vertebrate desmins and consistent with its designation as a type III ortholog forms homopolymeric IF in vitro (Riemer and Weber, 1998).

Here we have tried to extend the number of cytoplasmic IF proteins for urochordates. *Styela* C and D are novel IF proteins, but from their sequences alone one cannot decide to which of the type I to IV subfamilies of vertebrate IF proteins they are related. Both their coexpression in the epidermis and their obligatory heteropolymeric IF formation indicate a keratin relation. This is confirmed by the demonstration that *Styela* D and the type II keratin 8 of man and *Xenopus* readily form IF. Mutant proteins containing a single cysteine in equivalent positions show that such chimeric filaments like mammalian keratin filaments are based on the hetero coiled coil. Thus, in spite of its sequence drift, *Styela* D is a keratin I ortholog and by inference *Styela* C seems to reflect a keratin II ortholog. Taken together with the finding that *Styela* A and B form homopolymeric IF, the results show that orthologs of type I to III IF proteins are expressed along the chordate branch of metazoa.

Material and methods

Animals

Adult *Styela clava* were collected around Los Angeles (CA, USA). Animals were frozen in liquid nitrogen, shipped on dry ice and stored at -80°C .

Isolation of *Styela clava* IF cDNA clones

IF clones were isolated from a *Styela clava* tailbud stage λ ZapII cDNA library kindly provided by Dr. B. J. Swalla, University of Washington, Seattle, USA. Approximately 300 000 plaques grown on *E. coli* XL1Blue were hybridised at conditions of reduced stringency (Riemer et al., 1998) to DIG-labeled DNA fragments comprising the entire rod domain of the *Styela plicata* IF proteins A and B (Riemer and Weber, 1998). Labeling was performed by PCR amplification using the PCR DIG-Labeling Kit (Boehringer, Mannheim, Germany) and primers 5'-CTGAATTCAGGGACAGCGAATTAATG-3' (A143, sense, nt 371–397), 5'-GTTCTCATGGTTTGACTTTGTGCTG-3' (A144, antisense, nt 1346–1320) for *S. plicata* A and 5'-GAGTGA-GAGTTTGTCTCAACTACCTG-3' (B141, sense, nt 305–330), 5'-CTTCCATGCTTGTGTTTCGACTTATC-3' (B142, antisense, nt 1311–1286) for *S. plicata* B. Nucleotides (nt) refer to positions in the respective cDNAs. Following hybridisation, equilibration of filters in maleic acid buffer, blocking, incubation with anti-DIG Fab fragments coupled to alkaline phosphatase and visualisation of immunoreactive plaques were according to the manufacturer's instructions (Boehringer). Filters were exposed overnight at room temperature to Kodak XAR-5 films. Inserts from purified plaques were amplified by PCR, cloned into the pCR2.1 vector (Invitrogen, Leek, The Netherlands) and completely sequenced on both strands.

Expression and purification of recombinant proteins

For expression of full-length polypeptides the entire coding regions of the *S. clava* cDNAs B, C and D were amplified by PCR with primers "SB5" (sense, nt 72–107) 5'-CAAGTCGAATCAAAAGCTAGC-GAACAGAAAGAAGTC-3' and "SB3" (antisense, nt 1504–1463) 5'-GCTATAAATCCTATTTGTGGATCCAGACTGGAGAAAAA-TATC-3' for B, "SC5" (sense, nt 8–43) 5'-GAGCAAATAATTTTC-CATATGTATTCTTCTCCAATC-3' and "SC3" (antisense, nt 1409–1373) 5'-GTGACGTGTGCAGTGAATTCCTTGATCTAGTCT-GAATC-3' for C and "SD5" (sense, nt 144–178) 5'-CTAAAA-CAAATTAACCATATGAATACCAG-AGGACG-3' and "SD3" (antisense, nt 1549–1515) 5'-CCAAGTGGACGTTTAGAATTC-CTTTCTTTCTACTC-3' for D. Primers were designed to create unique restriction sites flanking the 5'- and 3'-ends of the coding sequences. PCR products were digested with the appropriate enzymes, gel purified and ligated in frame into pET23a vector DNA (Novagen, Madison, USA) and expressed in *E. coli* BL21(DE3)pLysS (Novagen).

The recombinant *S. clava* proteins were highly enriched in inclusion body preparations. Inclusion bodies were solubilised in 10 mM Na_2HPO_4 , pH 6.6, containing 8 M urea and 1 mM 2-mercaptoethanol and subjected to cation exchange chromatography on Mono S 5/5 using a gradient formed by the solubilisation buffer and the same buffer containing 0.4 M NaCl. Elution profiles were monitored by SDS-PAGE. Fractions containing pure protein D were pooled, dialysed against 8 M urea in 20 mM Tris-HCl, pH 8.0, 1 mM 2-mercaptoethanol and stored at -20°C until use. Proteins B and C, which were present in the flow through fractions, were dialysed against the Tris/urea buffer and subjected to anion exchange chromatography on Mono Q 5/5 using a gradient formed by this buffer and the buffer containing 0.3 M NaCl. Fractions highly enriched in B or C were pooled, diluted with equilibration buffer and again subjected to anion exchange chromatography to remove some minor contaminants. Purified proteins B and C were stored as above. Recombinant protein A from *S. plicata* was obtained as described (Riemer and Weber, 1998) and recombinant human keratin 8 was a kind gift from Dr. H. Herrmann, DKFZ, Heidelberg, Germany. All proteins were monitored by automated Edman degradation to verify their N-terminal sequences.

In vitro filament assembly

Aliquots of the pure proteins were dialysed at room temperature for three or more hours against assembly buffers using dialysis filters (Millipore, Eschborn, Germany). Structures formed were analysed by electron microscopy after negative staining with 2% uranylacetate. *S. clava* B was assembled at 0.2 mg/ml in 5 mM Tris-HCl, pH 7.1, 1 mM DTT. Equal molar mixtures of C and D were used at a total protein concentration of 0.3 mg/ml in 10 mM Tris-HCl, pH 6.8, 1 mM DTT. Equal molar mixtures of human keratin 8 and *Styela* D protein (total protein 0.15 mg/ml) were dialysed for 2 hours against 2 mM Tris-HCl, pH 9.0, 1 mM DTT and then for 20 hours against 10 mM Tris-HCl, pH 7.6, 1 mM DTT. Assembly conditions for A from *S. plicata* were 10 mM Tris-HCl, 50 mM NaCl, 1 mM 2-mercaptoethanol, pH 7.0 (Riemer and Weber, 1998).

Production and purification of antibodies

Full-length polypeptides B, C and D were further purified by preparative gel electrophoresis. Rabbits were immunised according to standard protocols. To remove cross-reacting activities, antisera were preabsorbed on a mixture of other recombinant proteins. These were C plus D for the B antiserum, B plus D for the C antiserum and B plus C for the D antiserum. Preabsorbed antisera were antigen affinity purified using the corresponding polypeptides. The *S. plicata* A antiserum raised against the A protein from *S. plicata* (Riemer and Weber, 1998) was preabsorbed on a mixture of *S. clava* proteins B, C and D. Columns used for the purification of the antisera were prepared with recombinant proteins bound to cyanogen bromide-activated Sepharose beads (Pharmacia, Uppsala, Sweden). The specificity of the polyclonal antibodies was tested by immunoblotting.

Protein gel electrophoresis and immunoblotting

Small pieces of frozen animals were dissected under a stereo microscope and dissolved directly in boiling sample buffer. Polypeptides were separated on 10% SDS-polyacrylamide gels and transferred electrophoretically to nitrocellulose membranes. Blots were incubated with the rabbit antibodies as described (Riemer and Weber, 1998). Preabsorbed and affinity-purified B, C and D antibodies were used as 1:50, 1:10 and 1:20 dilutions. Preabsorbed antibodies to *S. plicata* A protein were used at 1:100 dilution. Immunoreactive polypeptides were visualised with horseradish peroxidase-conjugated secondary antibodies using the ECL chemiluminescence kit (Amersham, Braunschweig, Germany).

Immunofluorescence microscopy

Frozen sections were cut from adult *S. clava*, fixed in cold acetone and incubated with antibodies as described (Bartnik and Weber, 1989). Preabsorbed and affinity-purified antibodies against *S. plicata* protein A and *S. clava* proteins B, C and D were used. Bound antibodies were detected with rhodamine-conjugated secondary antibodies (Dako, Glostrup, Denmark; dilution 1:40). DNA was stained with Hoechst 33258 (Hoechst, Frankfurt, Germany).

Isolation of the polymerisation-competent heterodimer of keratin 8 and *Styela* D with a single disulfide bridge

Site-directed mutagenesis of the *S. clava* cDNA for protein D, cloned into the pET23a expression vector (see above) was performed using the QuickChange mutagenesis kit (Stratagene, Heidelberg, Germany). The position of the cysteine mutation was confirmed by DNA sequencing. The cDNA for the corresponding cysteine mutant of *Xenopus* keratin 8 was described (Hatzfeld and Weber, 1990). Mutant proteins were expressed in *E. coli* and purified as described (Hatzfeld and Weber, 1990).

Equimolar amounts of both recombinant proteins (total protein concentration 0.6 mg/ml) were dialysed for 3 h at room temperature against 2 mM Tris-HCl, pH 9, 1 mM 2-mercaptoethanol and then for 18 h against 10 mM Tris-HCl, pH 7.5, 1 mM in reducing agent. An aliquot was removed to monitor filament formation by electron microscopy. The following steps involved buffers without reducing agent and were performed at room temperature. Filaments were dialysed for 20 h against 5 l 10 mM Tris-HCl, pH 7.5, with one change and harvested by centrifugation. The filament pellet was washed with 10 mM Tris-HCl, pH 8, 5 mM EDTA and again centrifuged. The final pellet was resuspended in the same buffer and left for 20 h. Filaments were dissolved in 10 mM Tris-HCl, pH 8, containing 8.5 M urea, followed by Mono Q 5/5 chromatography using the same buffer and a gradient formed with 0.4 M NaCl in the same buffer. The elution profile showed a single major peak which was analysed by SDS-PAGE under non-reducing as well as reducing conditions. Fractions containing the pure heterodimer with a single disulfide bond were pooled. An aliquot was dialysed for 4 h against 10 and 20 mM Tris-HCl, pH 7.5, to monitor filament assembly of the crosslinked heterodimer by electron microscopy.

Results

Cloning of four tunicate IF proteins from a tailbud stage library

Screening of the *S. clava* library with cDNA probes corresponding to the rod domains of the IF proteins A and B from adult *S. plicata* (Riemer and Weber, 1998) provided four distinct IF cDNA types A, B, C and D. The cDNA sequence of A from *S. clava* (accession number AJ271143) shows that the A proteins from the two *Styela* species share 89% sequence identity.

The cDNA sequence of *S. clava* B protein (Fig. 1; accession number AJ271144) predicts a protein of molecular weight 51 923 with an isoelectric point of 5.02, which extends at the N-terminal side the sequence previously given for the *S. plicata* protein (Riemer and Weber, 1998). Reinvestigating the cDNA sequence extended the open reading frame upstream by 97 nucleotides (Fig. 1). The sequence preceding the previously assumed ATG start codon (nucleotides 210–212) shows two putative start codons in frame, of which only the start codon at positions 105–107 fits into the Kozak sequence context required for translational initiation. This assumption was confirmed by expression in *E. coli* using the entire open reading frame between nucleotides 21 and 1389. Microsequencing of the recombinant protein (see below) provided the N-terminal sequence MEQK (Fig. 1), which strongly indicates that the second ATG codon is the presumptive site of translation. The extension of the *S. plicata* B cDNA sequence is available from the EMBL/GenBank (accession number AJ249990). The recombinant *S. clava* B protein starts also with the sequence MEQKE (Fig. 1). The B proteins from the two *Styela* species share 86% sequence identity.

The two additional *S. clava* cDNA clones, C and D (Fig. 1; accession numbers AJ271145 and AJ271146), define novel IF proteins (Fig. 2) with calculated molecular weights and isoelectric points of 51 643 and 4.77 for C and 51 165 and 5.28 for D. Three peptide sequences covering a total of 46 residues from a polypeptide of 50 kDa, present in a cytoskeletal preparation of epidermis dissected from *S. plicata*, are readily aligned with the *S. clava* D protein (Fig. 1).

Fig. 2A gives a sequence alignment of the four cytoplasmic IF proteins of *S. clava*. Their rod domains share 36 to 42% sequence identity (Fig. 2B). Judging by the relation between the type I to IV vertebrate IF subfamilies, such diversity would indicate distinct IF types. So far only the A protein is understood in its relation to the type I to IV families. In line with its sequence relation with vertebrate type III proteins (51% identity for the rod domain of *S. plicata* A and hamster desmin) and the presence of homology segments in head and tail domains, the recombinant A protein forms homopolymeric IF in vitro. In addition, like vertebrate desmin, the *Styela* A protein is expressed in muscle (Riemer and Weber, 1998). Curiously, the other 3 proteins seem, at least by sequence identity values (Fig. 2B), also more closely related to type III (up to 41 to 43%) than to type I (29 to 36%) or type II proteins (29 to 37%). In addition, A and D share a N-terminal motif observed in type III proteins, and also a C-terminal motif, which occurs in all type III proteins and also in the type II keratin 8 and in a shorter version in various type I proteins (underlined motifs in Fig. 2A). On the other hand, the presence of D in epidermis indicated by microsequencing results of three peptides (see above and Fig. 1) raised the question of a keratin relation, since the epidermis of vertebrates and cephalochordates is rich in keratins (Karabinos et al., 2000).

Styela C and D are epidermal IF proteins

To characterise the IF proteins in more detail *S. clava* proteins B, C and D were expressed in *E. coli*. Inclusion bodies, highly enriched in recombinant proteins, were dissolved in 8 M urea and the proteins were purified by ion exchange chromatography. The purity of the proteins was monitored by SDS-PAGE (Fig. 3) and automated protein sequence analysis. Aliquots of proteins B, C and D, which were additionally subjected to

S. plicata B

GTTTAATGTTTCGATAGTCGGTTTGGGAAGGCAGAGTACGTTAACTCAACACCCGAAATCTTATTGCTGCAGCCAAAGAGAAGCAAGAAATATGGAACAAAAGGAAGCAAGTTACAAACAGTTATGATGGAGTACCGCGTGGGAGGATTCACGAT
 (M) F D S P F G K A E Y V T S T P K S Y L L S Q E K T K N (M) E Q K E A S Y N S Y D G V S G G K A V T

S. clava B

[illegible]

S. claya C

GGCAGCAGAGCAAACTAATTTCGCCATGTAATTTCTTCTCAATCAAGTCAGGCGTAACGCGACAGCGTGTGCTGCACACAGTGTACTCATCGTCTGTGGAAAGTCTCTGCAGGAGGATATTCTTCATCAACTCGTGTTCCTGTGTTACGGAACT	150
M Y S P I K S G V T Q Q R V T T Q S Y A S S S G S P A G G Y S S T T R V S G Y G T	
TAACCCGGGATTACATCTTCTCGTATTCTACAGCTTTGGATTTCACCCAGCATGTGTCATCTCTACTATTTCGTCTTATGGAGGATATGGTGAGGAGGAAGTAATTCATCAGTGAAGAAAGATTCTCAAGAATCTGGAGGAGAGGAGG	300
N P G F T T S S R Y T G A F A P T M S S S S Y S Y A Y D G G Y G G G C G G T Y I S E K D L Q E L G G E G G	
AGCAAAATTCGAAGCATGAGTACAGGAGAAAGAAAGAACTCAAGTCTTCGAAACACGGTTTCTGCTCTCAATGAGCGATGAGATATTCTGGACGCAAAAACAACTCTGGATGCTCAACTTCGACAAATGCTCTGTGAAGTATG	450
A E F R S M R S G E K K E L K S L N N R F A S Y I E R V R Y L E Q Q K L L D A Q L R Q M S T C T Y D	142
TTCAACAATGATGATTGTCAATGCTGAAGTGAAGGAAATGAAGGAGCTTTGGATGCTTTGAAACGCGTATGTCGAAATGATTGAAGCCGAAGTGAACAACAATCAATCATGATGTCGGTGAATGAACAGAGAAATCGAAATCCGAAT	600
S Q L D D L L Y N T A E V R R L K G L L D A L N G D T R K M I E A E V E H M Q T D V G D L K R E Y E N A I	192
TCAGATATGTCGAAGAATGGAAGCGTGAATGGCAAAATTCAGAAATATAGTTGATGAGTGCACAGTCCCGGATCGATCTCGAAAGAAACATGTTTCTACATCCGAGGAAGATCGACTTCGAGAATCTTGCTTCACTCAGGATTAATCAA	750
Q D R E L E R E A L L A N L R N N V D E C T V T R V D L E R K L V L R E E L L R E E L A H I E I I N	242
CGAACTCAAACTCAACTTGTCCACCACCATCTCGAGTCGAAATGGACACCCATGTCCTGATCTAAGCGATCTTCTCCGAGACATCCGAGCTCAATACGATCTTCGACCGCAAGAAAACAAGGAAGAAGCCGAGAACTGTACAGCAG	900
E L K S Q L V T D H V R Y Q M D T H S F D L S D L L R D I R A Q Y Q D L A A K K N K E A E N T W S Y S	292
CAAGCTCAATGATCTCAACCAACAAGTATCCGGCAGATGCCACAGACTCAGAAGATCGCAATCTGAACTCAGCAAGATCAGAGCAGCGGCATACGCGCAAACTGCAATCCCTCCGATCTAACAAAGACTATCTTGAAGCGCCA	1050
K L N D L N T Q T V S G D A N R L K E S Q S E L S E Y R N R V S G H T A Q I E S L R N S K D I A E R L Q	1200
ACTGCGGATGGAGGATCGGTTCAACCGTGATATCGAGGATATCAGATCAAACTCAGCATCTCAAGCCCAACTCAGCAAGATTAAGTCAGAGATCTCGCAGCATTTGAGGAGATACCAAGAACTCTTTCGGTCAAACTTGGTCT	1350
L A D V E D R F N R D I G G Y Q D I T D L Q A Q L L D K I K S E M S D R L R E Y Q E L L S V K L G L L	392
CGATTTTGAGATCAGCAGATACCAAGAACTCTGGGAAGGAAGAATGGAGGCTTCAAGCTGTACCGTCCGCTCTCGCAGTGGAGAACAAATGTCACTACGAAGGCGTGTGAGGAATATGATCGGCGTACAAAAGCGCCTCATTTGTA	1450
D F E I S T Y Q E L L E G E W R L Q A V P S R L L A T G E Q M S L R R V V E E Y S A G K S A H S D	1320
TTCTTCGCCACGGAAATGTTCAGATTACAGATGATCAACAACTGACTGCACACGTCACATCATAAAAAGGAAATAATATTGTATTGTTCAGCTTCATTGTATATTTTAAGGCTACATCTTGCACTAATGTATTCTTAATTTTAAAAAT	1500
S R Q R K M S D S D *	1502
TCCTTTTATATGCTTAGTTTGGGACAAATGCATATCTCTGCACATTATAGCAAGTTTGTCCAACTCTGGTCGATTGCTCTGTGACAAAAGGGTTTGGTGATAAATGTTCTTCTGCTTGAATTTTATGTACGTAAATTTGTCGGGT	1650
GGCTATAAA	1659

S. claya D

TGGACGAGTATCGCAGTAACCTAAACAAATTAACAAATTAAGTACACGAGGACGAAACTCAAGCTACAGCGCTACTTTCCGAGGACAAACCGCATATTTCAACAGCTGGATCAAAAGGATTGGTTCCTCGTCTTATCCATGAAACTCCGA	150
M N T R G R N S S Y R F T F G G Q T A Y S T A G T S G K F G S R L I H E T T	37
TTGTTTCGAAAAAGTTTCATACCTTATTTCTTCAAGTGGGGTGAGTGGGAGCGCAGGACGCTTAATGGGAATATCGCTTTCGTCATCTCCGACGGATGGCATTAAGGTGGTTTTCAGCTGAAATTTGAATGAGAGGAACCTT	300
I V S K K T V S Y S Y S S S G G V G G S A G A S M G I S A S G M S A G G I E G G F S S L K L L N E K E L	87
TGAATGGACTCAATTCCTCGGTTGCATCATACCTCGAAAAAGTTCGTTATTTGGAGGACAAACCAATGCTTGAAGGAAGATCCGTGAAGCAAGTGTCAAGAAAGATGTGATCTGGGTGAAGATAAAGCTGAAATTCAGAAAAAC	450
L N G L N S R F A S Y V E K V R Y L E E Q N Q M L E K K I R E A S A K K D V D L G E D K R E N L R K	137
TCGCGTCCAAATAGACAGTCCGACGGTTCGCAAAATAAAGCTGAAATGGCAGAGGATAACTTGAGAGAGAGAACCTCAGAATTTGAAATGAAACTTGACTTCGAGGTCAACAACTCAAGAATTCGAGCATGAATTCGACAGATTGC	600
L R A Q I I D D A T V A K I K A E M A R D N L R G E A S E L K W K L D F E G Q R Q E L D D E L A R	187
G	
GTAAAGACGTTGATGACCAACAATGTGCGCGTGATCTTGAGCGAAAGATGAAACTTCGCGGAGAACTTGAATCAACCGAAAGATCATGACGAGAAAGATTAAGATCTTTAAGAAACAAATTCCTGCAACCAAGGATTTAAACATTG	750
R K D V D D C G A T T M V R V D L R K E I T E L R E E L A Y N R K I E D L K K A E A T T A E G Q V N I	237
Y L X K	
AAGTTGAGGAATCGCTCGCTAGCTGGGAAACTTTTTCGACAGATTTAGAGCTCAATACGAAGATATTTGATCTCAGAACAGAGATGAAGCTGAAGCTTGOTATAAGAGAAGTTCGAAGATCTTGAAGATAAAGCAAACTCAAAACGAAG	900
E V D G I A I P D V G E L L R Q I R A Q Y E A D I V I Q N R D E A E A W Y K K K F F E D L E D K A K S N Q	287
CGCAGCTTGAAGAGTCAAGGAGACATTTGGAGAAATACAGAAAACAAAGTTCTCACTCAACTCGEAATGGAGCTTGAATCTCTTCGAGGAACGAATGACTACTTTGGAGCGCAATCTGGCTGATGAGAGAAACGTTTGAATCAGAAAGTTACAA	1050
G D L E K V Q G D I G E Y R K Q V T Q L E M E L E S L R G T N D Y L E R N L A D V E K R Y E S E V T	137
GGTATCAGATGACTTGGTCGCATACGACCGACCTTGGACATGCTCTGGTGAATGAAGCAAGCTCTTCGCGAATATAAGCGTTAATGAGCGCTCAACATCTCTTGAAGAGAAATCAGAAACTACCGGGCGCTACTTGAAGGAG	1200
R Y Q D R L A R I G T D L E H A T G E M K R H L A E Y K R L M S V K L S L E K I E R T Y R R L L E G	387
AAGGAGACGGAAGCTTTCAAGCAGCTCGAGTTCGGAAGTGATGACGAGGTTTCTACAAAAACAAAACCTTCAAGAAAGTTTATATCAAGACAATAGAAACTAGGAGTGGCAAAATTTGATCTCTTAAGACATCTGAAAAAAGAGTCG	1350
E G D G K L S S S S S S D E V S T K T K T S K K V I I K T I E T R D G K I V S S K T S E K R V	437
ATAGATCAGCGAATTTCTATCTTCTATCTCTGACGGAAGAGAGTAGAAGAAAGGAATCTTAAAGCTCCACTTGGAACCTTTTGGCTTGAACAAATCTCACAAAAAAGAAATTTAAACATATTGTAGGAAACGATCAATATTGGAATATA	1500
D R S S S S S S S S S S D E E E *	452
AGTTTCCAAAACTCCTCGACCTTCTCTCATTAGGCCAAGCTCTCATAAAGCCAGGCGCTCATGAACACTCTCTACTTTCATCTAATTTGTTGTGTAGATCTCTGTAATTTCACTAATTTTGGCAAAATTTTCTGAAAAATGGCAAGAAA	1650
AGTAAAGTCTCCCTCAACCATTAATGTGCTTTTAAACTATTGAGCATACGAAGATTAAAGCTTAGTATTCTATATAAGACATATTGCAACATTTTATCAACTCATATTGCTTCAAGCTTCCACATTTTACATATAATGGAGAAA	1800
CTGCGCTTATGATCATATCTCAAAAAAATATTATTATTAGCTCTTATACCACTCAACATGGCGTTTCGAGAGCAGTATATATATATCTTTTAAAGTAACTCTTTTCAAGCGCAAAATGGTTTGTACAGATATTGTTTAAATTTTGGCAAAACCAAGCATCG	2100
TTTTCGAGATTATAAA	1950

Fig. 1. Nucleotide and predicted amino acid sequence of the cDNAs for the three IF proteins B, C and D from the tunicate *S. clava*. The stop codons are marked by an *asterisk*. The top shows the extended 5' end of the *S. plicata* B protein previously thought to start with the sequence MFGR (Riemer and Weber, 1998). Recombinant B proteins of *S. clava*

and *S. plicata* start with the sequence MEQ (arrowhead) at the second circled M. Horizontal lines give the sequences of three peptides obtained from a polypeptide (molecular mass 50 kDa) present in a cytoskeletal preparation of epidermis dissected from *S. plicata*. Amino acid exchanges are marked.

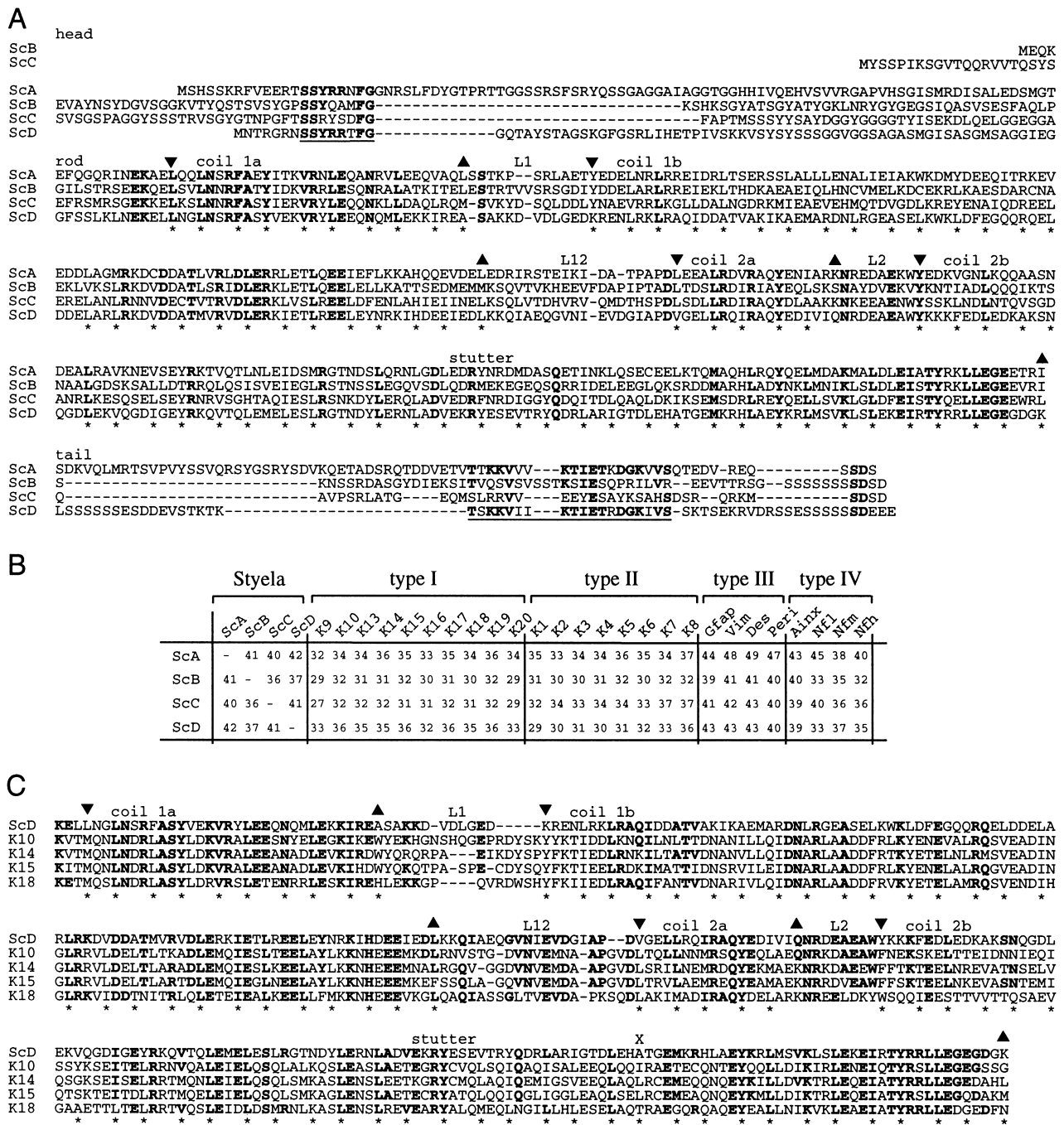


Fig. 2. Sequence properties of the four *S. clava* IF proteins. (A) Alignment of the four sequences. The tripartite domain organisation and the subdomains of the central rod domain are indicated. Identical residues are given in **bold** type. The horizontal lines mark conserved motifs in the head and tail domains of A and D. (B) Sequence identity values (%) for the rod domains in the four *Styela* proteins and human

type I to IV proteins. (C) Alignment of the rod domains of *Styela D* and four human type I keratins. Residues identical in D and one or more type I keratins are shown in **bold** type. The X above the top line marks the position where the alanine codon GCT was changed to the cysteine codon TGT in the mutant D protein.

preparative SDS-PAGE, were used to raise antibodies in rabbits. Individual antisera were crossabsorbed on the other recombinant proteins and then subjected to antigen affinity purification. Fig. 3 documents the specificity of the purified antibodies by immunoblotting experiments.

Immunofluorescence microscopy on frozen sections of adult *S. clava* showed that C and D are exclusively expressed in the epidermis (Fig. 4). The parallel DNA stain indicates that the

epidermis is a simple epithelium (Bartnik and Weber, 1989; Burighel and Cloney, 1997). The immunofluorescence images clearly showed that C and D are coexpressed in all epidermal cells of the sections. Earlier electron micrographs of tunicate epidermis indicated the presence of IF meshworks (Bartnik and Weber, 1989). As earlier reported for *S. plicata*, protein B is present in some interior epithelia, while muscle tissue is strongly decorated by antibodies to A (Riemer and Weber,

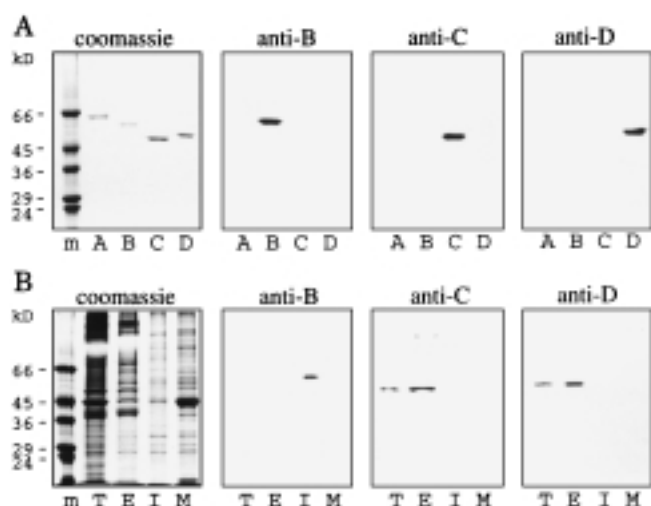


Fig. 3. Immunoblotting results on recombinant proteins (**A**) and tissue extracts (**B**). Recombinant proteins (A from *S. plicata*; B, C and D from *S. clava*) and protein extracts of *S. clava* (T total animal extract; E epidermis; I internal epithelia; M body wall musculature) were separated by SDS-PAGE and either stained with Coomassie brilliant blue or transferred to nitrocellulose membranes. (**A**) shows the specificity of the purified rabbit antibodies for proteins B, C and D respectively. (**B**) shows that protein B was detected only by B antibodies in internal epithelia. C and D are found in the epidermis and not in muscle or internal epithelia. Marker polypeptides are given at the left side.

1998). This strong reaction for A is in line with earlier reported electronmicrographs which document a wealth of IF in tunicate muscle tissue (Terakado, 1988).

IF assembly: C and D as well as human keratin 8 and D form obligatory heteropolymers

Mammalian keratin filaments are obligatory heteropolymeric filaments based on equal numbers of keratin I and II polypeptides arranged as double-stranded hetero coiled coils (Coulombe and Fuchs, 1990; Hatzfeld and Weber, 1990; Steinert, 1990). In contrast, type III proteins form homopolymeric filaments based on homo coiled coils. We made use of these molecular features to decide whether the coexpression of C and D in the tunicate epidermis could be related to the presence of keratin-like filaments.

Purified recombinant proteins were dialysed either alone or as an equimolar mixture to remove the urea and the self-assembly products were analysed by electron microscopy after negative staining. While C or D alone resulted only in aggregated material, the stoichiometric mixture of the two epidermal proteins provided the IF in Fig. 5A. Since the two proteins behaved as keratin orthologs, we explored the possibility of obtaining chimeric IF based on a human and a tunicate IF protein. Fig. 5B shows that human keratin 8, a type II protein, assembles with D into normal IF. This IF formation also occurred when *Xenopus* keratin 8 and *Styela* D were used (data not shown). Thus, D behaves like an ortholog of type I keratins. The reverse experiment using human keratin 18, a type I protein, and the tunicate C protein did not lead to IF (see Discussion).

We previously found with *S. plicata* that the muscle protein A formed excellent homopolymeric IF, while protein B did not form filaments (Riemer and Weber, 1998). We now observed

that protein B of *S. clava* (N-terminal sequence MEQK), which shows an extension of 35 residues versus the previous *S. plicata* protein (N-terminal sequence MFGR), can form homopolymeric IF (Fig. 5C). The same result was obtained when the corresponding longer recombinant *S. plicata* B protein was used (N-terminal sequence MEQK). Thus, the N-terminal 35 residues make, as in other IF proteins, a major contribution to filament formation (Fuchs and Weber, 1994; Parry and Steinert, 1995). The filaments formed by B do not seem to reach the length observed in assembly studies of A (compare Figs. 5C and D). In addition, the presence of some spheroid particles indicates that the assembly of B is less optimal than that of A.

Isolation of the polymerisation-competent heterodimer of keratin 8 and *Styela* D containing a cysteine bond

To decide whether the chimeric filaments formed by keratin 8 and *Styela* D are built from hetero coiled coils, we explored one of the approaches which was previously used to document the keratin I/II hetero dimer (Hatzfeld and Weber, 1990). It is based on the observation that in parallel and unstaggered α helices of a double-stranded coiled coil a cysteine in an *a* or *d* position of the heptads can form a disulfide crosslink (Lehrer, 1975; McLachlan and Stewart, 1975). A *Xenopus* keratin 8 mutant containing a single cysteine located 28 residues before the conserved tyrosine in the sequence YRKLL was available (Hatzfeld and Weber, 1990). Using site-specific mutagenesis, the same *d* position in *Styela* D, which lacks cysteines, was changed to cysteine. Heterodimers with a single disulfide bridge were observed (Fig. 6). The oxidised filaments were dissolved in 8 M urea without reducing agent and the heterodimer was purified by ion exchange chromatography in the same solvent (Fig. 6). Subsequent dialysis against assembly buffer without reducing agent provided the filaments shown in Fig. 7. Thus, the chimeric filaments arise like normal keratin filaments from a hetero coiled coil and *Styela* protein D can substitute for a keratin I molecule (for sequence comparison of the rod domains see Fig. 2C).

Discussion

Most vertebrate cytoplasmic IF proteins are readily recognised as a member of one of the four subfamilies I to IV. Usually different members of the same subfamily share 65 to 70% sequence identity over the rod domains, while different subfamilies share only 30 to 35% identity (Fuchs and Weber, 1994; Parry and Steinert, 1995). An additional important diagnostic tool is the fact that the keratin filaments of epithelia are obligatory heteropolymers containing equal numbers of type I and II polypeptides. The molecular basis for this type of heteropolymer is the first assembly product which is identified as the heterodimeric coiled coil, built from one type I and one type II polypeptide (Coulombe and Fuchs, 1990; Hatzfeld and Weber, 1990; Steinert, 1990). We have exploited this heterodimeric feature of keratin filaments in an attempt to understand the two new IF proteins C and D cloned from a tailbud stage library of the tunicate (urochordate) *Styela*.

Immunofluorescence microscopy with antibodies raised against the recombinant proteins showed that C and D are coexpressed in the epidermis (Fig. 4), which is a simple

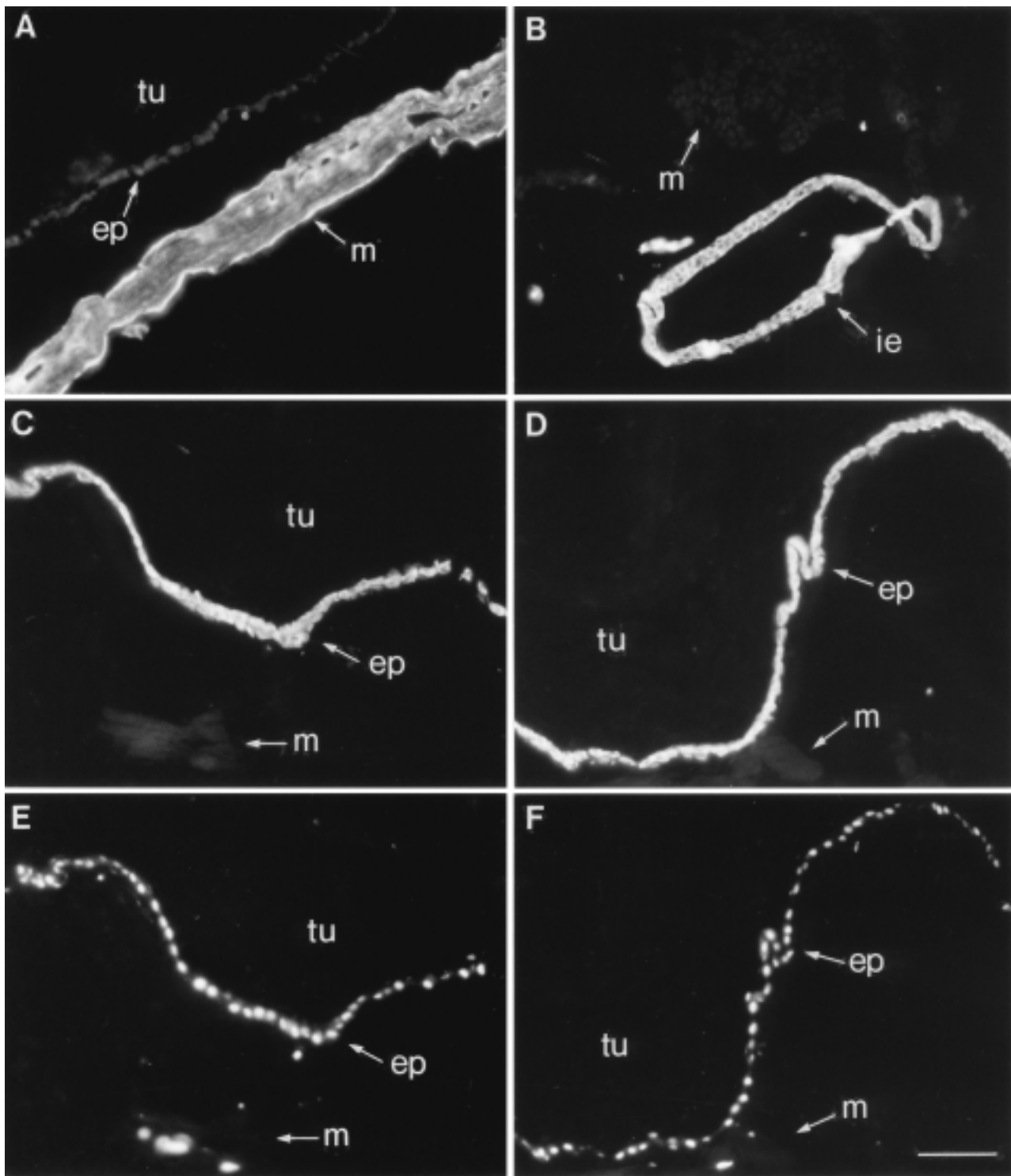


Fig. 4. Immunofluorescence microscopy on cryosections of *S. clava*. Monospecific rabbit antibodies against A detect the IF protein A only in smooth muscle (A). Monospecific antibodies to B detect the protein only in interior epithelia (B). The epidermis is only decorated by monospecific rabbit antibodies to C (C) and D (D). Abbreviations are

ep epidermis, ie internal epithelia, m muscle and tu tunic. Parallel sections to C and D stained for DNA are shown in (E and F). Note the co-expression of C and D in all cells of the epidermis, which is a simple epithelium. Bar indicates 50 μ m.

epithelium. Although the epidermal nature of C and D raised the possibility of a keratin relation, the two sequences were difficult to analyse (Fig. 2). Sequence identity values for the rod domains seemed even compatible with type III orthologs. In this situation the self-assembly properties of recombinant C

and D were very informative. Since C and D alone formed only aggregated material, while their equimolar mixture provided normal IF, a type I/II keratin relation was directly indicated. However, the sequences did not directly indicate which of the two *Styela* proteins behaves as a type I and which as a type II

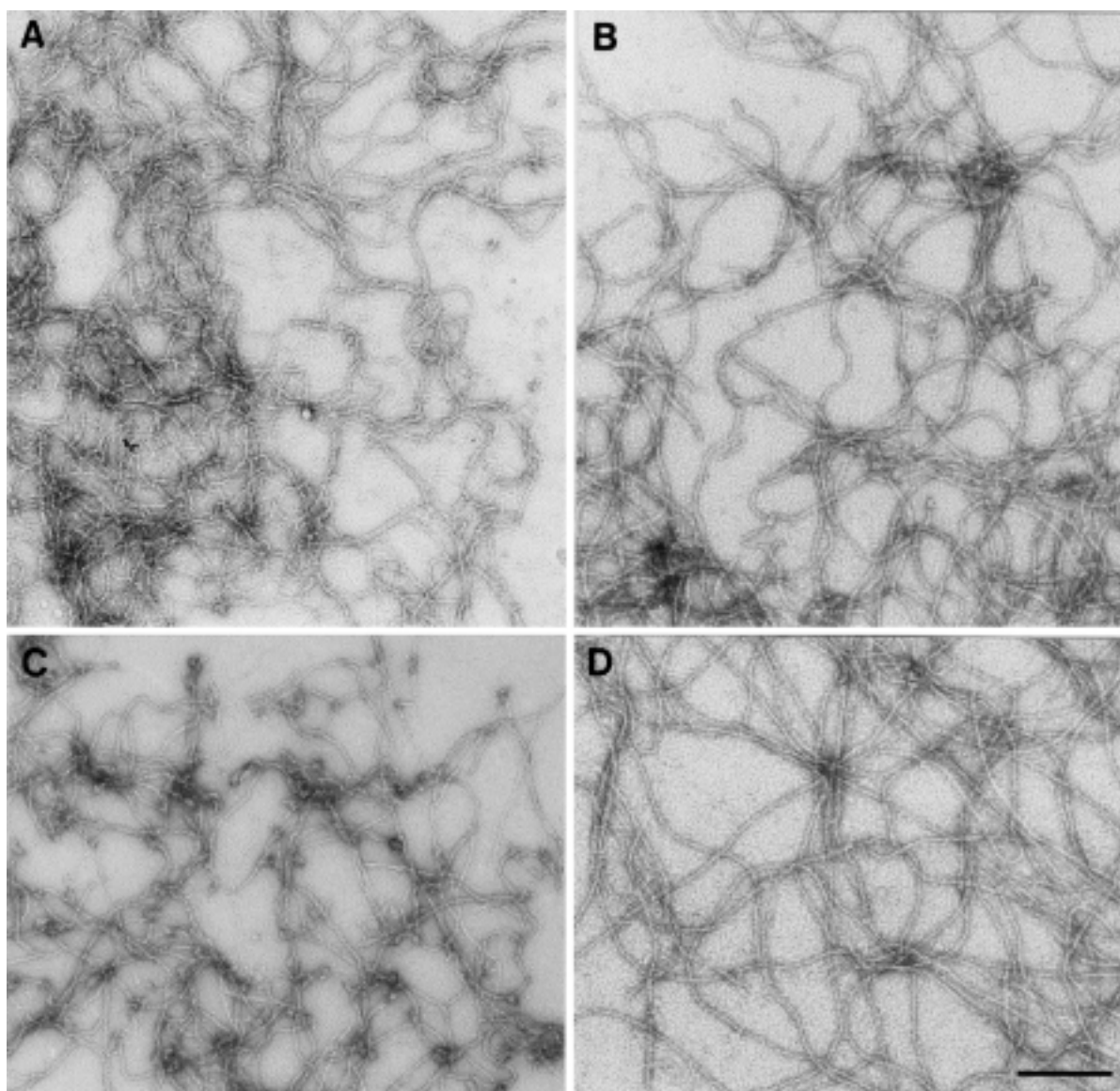


Fig. 5. Electron micrographs of IF formed by recombinant proteins. Purified recombinant proteins were dialysed alone or in equal molar mixtures against filament buffers and negatively stained with 2%

uranyl acetate. (A) stoichiometric mixture of *Styela* proteins C and D. (B) Stoichiometric mixture of human keratin 8 and *Styela* D protein. (C) *Styela* protein B. (D) *Styela* protein A. Bar 200 nm.

keratin ortholog. This difficulty was also not overcome by phylogenetic tree calculations (data not shown). However, the problem was solved by the finding that keratin 8 from *Xenopus* or man, a type II keratin, forms chimeric IF when mixed with equimolar amounts of *Styela* D (Fig. 5). Mutant proteins of *Xenopus* keratin 8 and *Styela* D with a single cysteine, which was placed into the same heptad *d* position in helix 2b, confirmed that this assembly process directly mimics normal keratin filament formation. When the chimeric filaments were allowed to oxidise, the hetero coiled coils formed the single disulfide bond in very high yield. The purified heterodimer directly formed the IF shown in Fig. 7. Thus, in the type I/type II keratin heterodimer keratin 18, a natural partner of keratin 8, can be replaced by *Styela* D, although the two proteins share only 35% sequence identity over the rod domains (Figs. 2B and C). The combined results document a keratin type I ortholog character for *Styela* D. They also imply that *Styela* C,

which forms obligatory heteropolymeric IF with *Styela* D, can be considered to reflect a keratin type II ortholog.

The reverse experiment to obtain chimeric filaments from *Styela* C and human keratin 18, a type I keratin, was so far not successful. Possible reasons may be the limited set of assembly conditions tried, a too strong divergence of the rod sequence to allow the formation of stable heterodimers, or incompatible contributions by the head domains, which are also required for IF formation.

The difficulties encountered in establishing the character of keratin orthologs with strongly drifted sequences document the need for in vitro assembly studies with the recombinant proteins. It is interesting to note that in the case of the cephalochordate *Branchiostoma*, which is considered to be the closest relative of the vertebrates (Turbeville et al., 1994), identification of keratin orthologs was not possible by sequence identity values alone (Riemer et al., 1998), and

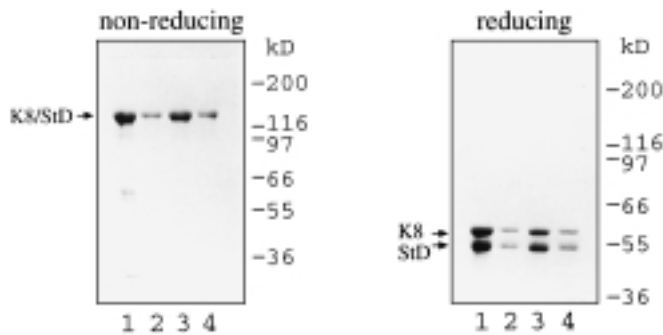


Fig. 6. Filaments assembled from the cysteine mutants of *Xenopus* keratin 8 and *Styela* D protein form a cystine bond in the heterodimer. Filaments assembled in the presence of 2-mercaptoethanol were freed of reducing agent and allowed to oxidise over a time period of more than 40 h. After solubilisation in 8 M urea the protein was subjected to Mono Q chromatography and analysed by SDS-PAGE under non-reducing (left panel) and reducing conditions (right panel). Lanes 1 show the material before Mono Q chromatography. Lanes 2 to 4 give three fractions from the major protein peak. A molecular mass standard in kDa is shown at the right of each panel. Note the heterodimer of K8 and StD containing the cystine bond (left panel) and the dissociation into monomers after reduction. The fractions from the Mono Q column contain pure heterodimer with a cystine bond. Note the nearly complete oxidation of the filaments.

phylogenetic tree calculations were necessary to identify three type I keratin orthologs and two type II keratin orthologs. In each case this assignment was verified by the demonstration of obligatory heteropolymeric IF assembly in vitro (Karabinos et al., 1998; Luke and Holland, 1999; Karabinos et al., 2000). Urochordates (tunicates) are generally considered to be the evolutionary oldest phylum of the chordates. Currently, we do not know if the difficulty in identifying keratin orthologs in tunicates is related to their evolutionary distance. Although *Styela* C and D proteins seemed to be keratin orthologs by their heteropolymeric IF formation, the key experiment to identify *Styela* D as a type I keratin ortholog was the observation of chimeric filaments formed by the *Styela* protein and keratin 8, a type II keratin from *Xenopus* or man.

Of the four urochordate IF proteins currently established, A and B follow the type III subfamily by their homopolymeric filament-forming ability, while C and D reflect keratin II and keratin I orthologs. Since type I to III orthologs are already known in the cephalochordate *Branchiostoma* (Karabinos, 1998; Luke and Holland, 1999; Karabinos et al., 2000), they seem a feature common to all three phyla of the chordate branch of metazoa. All chordate cytoplasmic IF proteins display the short coil 1b version, while the cytoplasmic IF proteins from 11 protostomic phyla show, like the nuclear lamins, the long coil 1b version with six extra heptads and have in general also a lamin homology segment in their tail domains (Weber et al., 1989; Erber et al., 1998). Thus, the common precursor of chordate type I to III proteins, which lost the six extra heptads, arose either with the origin of the chordates or the origin of the deuterostomes. A decision between these two models may be possible once sequences for cytoplasmic IF proteins of echinoderms and hemichordates become available.

A major remaining problem of the IF multigene family of tunicates concerns the presence of possible orthologs of neuronal type IV proteins. Although they are well documented in fish (Markl and Schechter, 1998), we have so far not detected such proteins either in the cephalochordates (Karabinos et al., 2000) or the tunicates (this study).

Acknowledgements. We thank Dr. Billie J. Swalla, University of Washington, Seattle, USA, for the tailbud stage cDNA library of *Styela clava*. Dr. H. Herrmann kindly provided recombinant human keratins K8 and K18.

References

- Bartnik, E., Weber, K. (1989): Widespread occurrence of intermediate filaments in invertebrates; common principles and aspects of diversification. *Eur. J. Cell Biol.* **50**, 17–33.
- Burighel, P., Cloney, R. A. (1997): Urochordata: Ascidiacea. In: F. W. Harrison, E. E. Ruppert (eds.): *Microscopic Anatomy of Invertebrates*. Wiley-Liss Inc., New York, Vol. 15, pp. 221–347.
- Coulombe, P. A., Fuchs, E. (1990): Elucidating the early stages of keratin filament assembly. *J. Cell Biol.* **111**, 153–169.

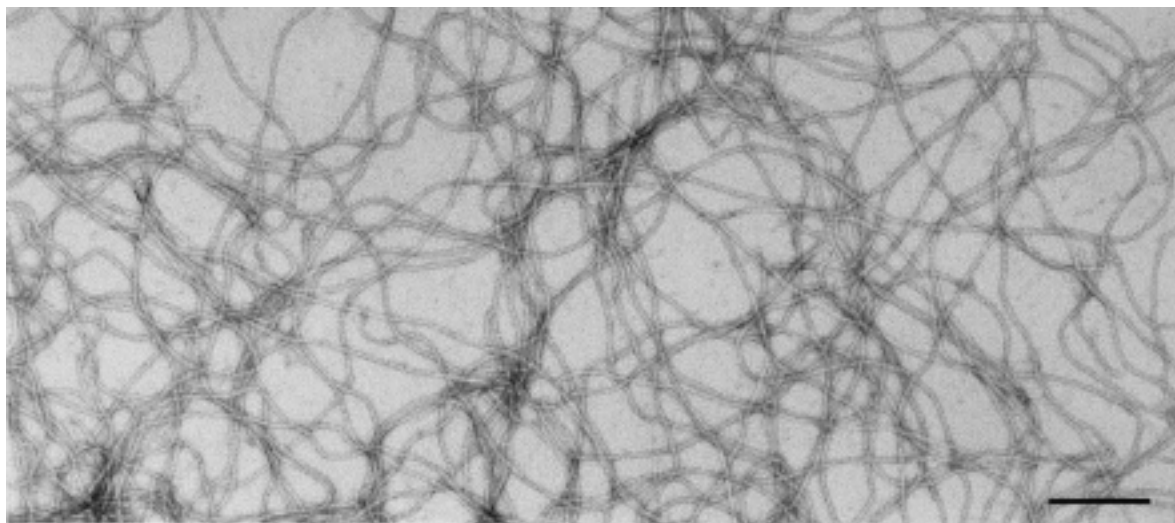


Fig. 7. Electron micrograph of IF formed by the cysteine-linked heterodimer between *S. clava* D and *Xenopus* keratin 8. Purified

heterodimers (see Fig. 6, Lanes 2 to 4; left panel) were dialysed to filament buffer without reducing agent. Bar indicates 200 nm.

- Dodemont, H., Riemer, D., Ledger, N., 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.
- Erber, A., Riemer, D., Bovenschulte, M., Weber, K. (1998): Molecular phylogeny of metazoan intermediate filament proteins. *J. Mol. Evol.* **47**, 751–762.
- Erber, A., Riemer, D., Hofemeister, H., Bovenschulte, M., Stick, R., Panopoulou, G., Lehrach, H., Weber, K. (1999): Characterization of the *Hydra* lamin and its gene; a molecular phylogeny of metazoan lamins. *J. Mol. Evol.* **49**, 260–271.
- Fuchs, E., Cleveland, D. W. (1998): A structural scaffolding of intermediate filaments in health and disease. *Science* **279**, 514–519.
- Fuchs, E., Weber, K. (1994): Intermediate filaments: Structure, dynamics, function and disease. *Annu. Rev. Biochem.* **63**, 345–382.
- Hatzfeld, M., 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**, 1199–1210.
- Karabinos, A., Riemer, D., Erber, A., Weber, K. (1998): Homologues of vertebrate type I, II and III intermediate filament (IF) proteins in an invertebrate; the IF multigene family of the cephalochordate *Branchiostoma*. *FEBS Lett.* **437**, 15–18.
- Karabinos, A., Riemer, D., Panopoulou, G., Lehrach, H., Weber, K. (2000): Characterisation and tissue-specific expression of the two keratin subfamilies of intermediate filament proteins in the cephalochordate *Branchiostoma*. *Eur. J. Cell Biol.* **79**, 17–26.
- Lehrer, S. S. (1975): Intramolecular crosslinking of tropomyosin via disulfide bond formation: evidence for chain register. *Proc. Natl. Acad. Sci. USA* **72**, 3377–3381.
- Luke, G. N., Holland, P. W. H. (1999): An amphioxus type I cDNA and the evolution of intermediate filament genes. *J. Exp. Zool.* **285**, 50–56.
- Markl, J., Schechter, N. (1998): Fish intermediate filament proteins in structure, evolution and function. In: H. Herrmann, J. R. Harris (eds.): *Subcellular Biochemistry*. Plenum Press, New York, Vol. 31, pp. 1–33.
- McLachlan, A. D., Stewart, M. (1975): Tropomyosin coiled-coil interactions; evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293–304.
- Parry, D. A. D., Steinert, P. M. (1995): *Intermediate filament structure*. Springer, New York.
- Riemer, D., Dodemont, H., 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., Karabinos, A., 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.
- Riemer, D., Weber, K. (1998): 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. *J. Cell Sci.* **111**, 2967–2975.
- Steinert, P. M. (1990): The two chain coiled coil molecule of native epidermal keratin filaments is a type I-type II heterodimer. *J. Biol. Chem.* **265**, 8766–8774.
- Terakado, K. (1988): The pattern of organisation of intermediate filaments and their asymmetrical association with dense bodies in smooth muscle of an ascidian, *Holocynthia roretzi*. *Cell Tiss. Res.* **252**, 23–32.
- Turbeville, J. M., Schulz, J. R., Raff, R. A. (1994): Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Mol. Biol. Evol.* **11**, 648–655.
- Weber, K., Plessmann, U., Ulrich, W. (1989): Cytoplasmic intermediate 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**, 3221–3227.