# Homologues of vertebrate type I, II and III intermediate filament (IF) proteins in an invertebrate: the IF multigene family of the cephalochordate *Branchiostoma*

Anton Karabinos, Dieter Riemer, Andreas Erber, Klaus Weber*

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

**Abstract** We searched for functional homologues of the four subfamilies of vertebrate cytoplasmic intermediate filament (IF) proteins in the cephalochordate *Branchiostoma*. The epidermis contains in addition to IF proteins C2 and D1 two novel IF proteins E1 and E2. Both sequence comparisons as well as the obligatory heteropolymer formation by the recombinant proteins identify E1 as a type I keratin and E2 and D1 as type II keratins. In contrast the non-epidermal B1 forms as type III homologue homopolymeric IF. We propose that type I–III diversification of IF proteins is a property of the chordate branch of metazoa and discuss a possible origin of type IV neurofilaments.
© 1998 Federation of European Biochemical Societies.

*Key words:* Intermediate filament; Keratin; Vimentin; Chordate; *Amphioxus*; *Branchiostoma*

## 1. Introduction

Characterization of the keratins in certain human skin diseases shows that one function of intermediate filaments (IF) is related to cellular resistance against mechanical stress [1]. All IF proteins show a central alpha helical rod domain of coiled coil forming ability flanked by variable head and tail domains. The rod is divided into 4 subdomains (coils 1a, 1b, 2a and 2b) connected by short linkers. The true consensus sequences at the ends of the rod are involved in filament assembly. Protein sequences, biochemical properties, cell and tissue specific expression patterns and the organization of the corresponding genes allow a division of the 50 members of the mammalian IF proteins into 5 subfamilies. The two keratin types necessary to form the obligatory heteropolymeric keratin filaments of the various epithelia form types I and II. Homopolymeric IF are observed with type III which includes vimentin, desmin and their relatives. Type IV covers the various neurofilament proteins. Finally type V describes nuclear IF proteins, the lamins. Nuclear lamins differ from vertebrate cytoplasmic IF proteins by an extra 42 residues (6 heptads) in their coil 1b subdomain and by unique tail domains which display a nuclear localization signal and in most cases a terminal CaaX box [2,3].

The four subfamilies of cytoplasmic IF proteins extend from mammals to fish [4]. In contrast to vertebrates, protostomic phyla display cytoplasmic IF proteins, which are more closely related to nuclear lamins. They also have an extra 42 residues in their coil 1b subdomain and in most cases harbor a lamin homology segment of some 120 residues in their tail domain. These features, which were originally documented for molluscs and nematodes [5–11], have meanwhile also emerged in nine additional protostomic phyla [12]. Parallel studies showed that the short coil 1b version present in all vertebrate cytoplasmic IF proteins is also present in eight cytoplasmic IF proteins of *Branchiostoma* [13,14] and two IF proteins from the tunicate *Styela* [12,15]. Thus the short coil 1b version is shared by vertebrates, cephalochordates and urochordates, which form the chordate branch of the deuterostomia. 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 [12]. An important additional question is whether the early chordates possess identifiable homologues of the type I–IV subfamilies established for vertebrates.

Previously we found it difficult to relate the eight *Branchiostoma* IF proteins unambiguously to particular type I–IV IF families of vertebrates by sequence identity values, gene organization and tissue specific expression patterns. Although the IF protein D1 could be considered as a keratin II homologue the previous collection of IF proteins lacked an obvious candidate for a type I keratin homologue [14]. We have now taken a different approach. We determined the complement of IF proteins in isolated epidermis of *Branchiostoma* by two dimensional gel electrophoresis and microsequencing and show, using in vitro filament formation as an assay, that proteins E1 and D1 are homologues of keratin I and II, respectively. Additional experiments identify the non-epidermal B1 protein as a type III homologue.

## 2. Materials and methods

### 2.1. Animals

Adult *Branchiostoma lanceolatum* (*Amphioxus*) collected from the island of Helgoland in the North Sea were used to prepare a λZap II library essentially as described [12]. Adult animals collected at Roscoff (France) were used to prepare poly(A)+ RNA and to dissect epidermis [14].

### 2.2. Epidermis cytoskeleton, two dimensional gels and peptide sequences

Epidermis was extracted with high and low salt buffer [6] and the cytoskeletal residue was dissolved in 9 M urea containing 5% NP40 and 2% ampholines (pH 3.5–10). Separation in the first dimension by non-equilibrium pH gradient electrophoresis [16] was followed by SDS-PAGE in 10% acrylamide gels. Some gels were used in immunoblotting with rabbit antibodies to IF proteins D1 and C2 and monoclonal antibody IFA [14]. Ten gels were stained with Coomassie brilliant blue. The major spots were excised, washed with water and frozen at −20°C until use. In situ proteolysis of pooled spots with endoproteinase LysC (Boehringer Mannheim, Germany), separation

of the digests by reverse phase HPLC and automated Edman degra-
dation of isolated peptides followed standard protocols.

### 2.3. Oligonucleotides, polymerase chain reaction and cloning of new IF proteins

RT-PCR reactions used degenerate oligonucleotides based on the
established peptide sequence information. In the case of E1 the sense
primer K197 (5′-GCI(CT)TIGA(AG)GT(AGTC)GA(AG)CA(AG)-
TGG) and the antisense primer K196 (5′-TTCATCAT(AG)T-
C(AG)TT(AG)TA) were based on the peptide sequences ALEVEQW
and YNDMMK, respectively. Single strand cDNA synthesis was with
Superscript II RNase H⁻ Reverse Transcriptase (Gibco, Eggenstein,
Germany) essentially as described [12]. The amplified PCR product
was cloned into the pCR 2.1 vector, transformed into competent *E.
coli* TOP F′-cells (Invitrogen, San Diego, CA, USA) and the RT-PCR
fragment E1 was identified by sequencing. It was used as a probe for
the isolation of the full length E1 cDNA from a λZap II library. The
cDNA was completely sequenced on both strands. In the case of E2
the sense primer K198 (5′-GA(AG)GA(AG)CA(AG)AC(AGT-
C)(AC)GIAA(CT)GA(AG)GG) and the antisense primer K211 (5′-
CC(GAT)AT(CT)TCIAC(AG)TC(AGTC)AG(AGTC)GCCAT) were
based on the peptide sequences EEQTRNEG and MALDVEIG, re-
spectively. Amplification of the PCR product of E2 was as above.

### 2.4. Expression and purification of recombinant proteins

Coding sequences were amplified by PCR and ligated into bacterial
expression vectors. For B1 and D1 vector pKK388-1 (Clontech, Hei-
delberg, Germany) was used and *E. coli* JM109 was transformed with
the plasmid constructs. For E1 vector pET23 (Novagen, Madison,
WI, USA) and *E. coli* BL21 (DE3) pLysS were used. Inclusion bodies,
highly enriched in recombinant proteins B1 and D1, were solubilized
in 8 M urea containing 10 mM $Na_2HPO_4$, 1 mM ME, pH 6.6, and the
IF proteins were purified by ion exchange chromatography on Mono
S in 8 M urea buffer using a linear gradient from 0 to 400 mM NaCl.
In the case of E1 9 M urea containing 10 mM Tris-HCl, 2 mM
EDTA, 1 mM ME, pH 8.6 and Mono Q chromatography with a
salt gradient from 0 to 200 mM NaCl were used. Purification was
monitored by SDS-PAGE.

### 2.5. In vitro filament formation

Aliquots of the recombinant proteins (40 µl; about 0.2 mg/ml)
either alone or as stoichiometric mixtures were dialyzed at room tem-
perature for three or more hours on dialysis filters (Millipore, Es-
chborn, Germany) against filament buffer, usually 50 mM Tris-HCl,
pH 7.5, 1 mM ME. Negative staining was with 2% uranylacetate.
Electron micrographs were taken with a Philips CM12 electron
microscope.

## 3. Results

The cytoskeletal residue from dissected epidermis of *Bran-
chiostoma* (*Amphioxus*) was separated by two dimensional gels
(Fig. 1). Corresponding spots from 10 gels were pooled and
the polypeptides were characterized by interior peptide se-
quences obtained by proteolysis with endoproteinase LysC.
The major spot D1 and the series of C2 spots were identified
as the previously cloned IF proteins D1 and C2 in line with
immunoblotting and immunofluorescence results with appro-
priate antibodies [14]. Peptide sequence information from the
major spot E1 was used to synthesize degenerate oligonucleo-
tides for RT-PCR. The cDNA fragment obtained served as
probe to isolate a full length E1 cDNA clone from a λZap II
library. Fig. 2 shows that the E1 protein is an IF protein with
a very short tail domain (5 residues) somewhat reminiscent of
keratin 19 in simple epithelia of mammals [2,3]. The N-termi-
nal head domain contains, like the terminal domains of sev-
eral mammalian epidermal keratins, several glycine loops
flanked by hydrophobic residues [3]. Over the rod domain
E1 shares up to 36% sequence identity with vertebrate type
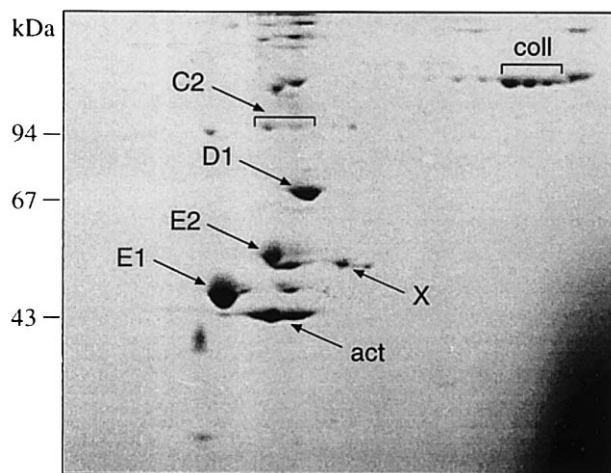I keratins and only up to 30% with type II proteins. Since we



Fig. 1. Two dimensional gel pattern of the cytoskeletal residue from
the epidermis of *Branchiostoma l.* Protein spots were identified by
microsequencing of peptides obtained by in situ digestion with en-
doproteinase LysC. The IF proteins D1 and C2 were previously
cloned [14]. Peptides from spots E1 and E2 allowed the cloning of
the new IF proteins E1 and E2 (see Fig. 2 and text). Act is actin
and coll is collagen. The spot labelled X provided only a single
pure peptide. Its sequence is compatible with either a novel cyto-
plasmic IF protein or a nuclear lamin. An approximate molecular
mass standard in kDa is given at the left.

previously speculated that the epidermal IF protein D1 could
be a candidate for a type II homologue [14] both E1 and D1
were expressed in *E. coli*. The recombinant proteins were pu-
rified in 8 or 9 M urea and their IF forming ability was
monitored by electron microscopy after removal of urea by
dialysis. E1 alone and D1 alone formed only aggregated ma-
terial. In contrast the stoichiometric mixture of E1 and D1
resulted in long filaments (Fig. 3). Thus E1 and D1 share
obligatory heteropolymer formation with vertebrate type I
and II keratins, respectively. In addition we analyzed the pol-
ymer forming ability of the IF protein B1 which was previ-
ously cloned [14]. It shows over the rod domain a slightly
higher sequence identity with vertebrate type III proteins
(up to 45%) than with type I and II keratins (up to 38%)
[14]. Fig. 3 shows that recombinant B1 forms, like vertebrate
type III proteins, long homopolymeric IF.

The cDNA sequence of the epidermal E2 protein is incom-
plete (Fig. 2). Over the 306 amino acid residues of the rod
domain E2 shares 53% identity with D1 and only 27–35%
identity with the other 8 *Branchiostoma* IF proteins. Thus
E2 is recognized as a keratin type II homologue.

Fig. 4 shows an evolutionary tree generated from the se-
quences of the rod domains of the eight previously cloned
*Branchiostoma* IF proteins [14], the new IF proteins E1 and
E2 and a variety of human representatives of vertebrate type
I–IV proteins. In agreement with the obligatory heteropoly-
mer formation E1 and D1/E2 relate to type I and II keratins,
respectively. The B1 protein and its relatives (A1, A2, A3 and
B2) form the sister group of vertebrate type III plus type IV
proteins (see Section 4). The unique C1 and C2 proteins [14]
form a separate branch.

## 4. Discussion

This study extends the number of *Branchiostoma* cDNAs
coding for cytoplasmic IF proteins from 8 to 10. Together

**E1**

```
CGTCTAATCGTCCTGAACCTGCGTCGAACTAACTCGTCGGCCATCATGAGTCGTGGTACCGGTAATATCCTGAGTAGCGCCGACATCCTGCTCAGCAATGTGTCGGGCTTGTCCTGGCAGCAAACCAGCCGCTCGGGCGGCGGGGGCATG   150
                                               M  S  R  G  T  G  N  I  L  S  S  A  D  I  L  L  S  N  V  S  G  L  S  W  Q  Q  T  S  R  S  G  G  G  M    35

ATCAGCAGCAGCTACTCCCTCGGGGGCGGTGGTGGCTTCGGAGGCGGCGGCGGCGGCGGTGGCGGCCGGTGGCGGCCGGTGGCGGGGGAGGCCAGCATGGCCCTGGTTCCCGCCGGCCGTGCCTCGGCGTCGTTTTCGTCTTCGTCCGCTAGCTTCGGGGGTGGC   300
I  S  S  S  Y  S  L  G  G  G  G  G  F  G  G  G  G  G  G  G  G  G  R  G  G  G  A  S  M  A  L  V  P  A  G  R  A  S  A  S  F  S  S  S  S  A  S  F  G  G  G    85

GGCGGCGGCGGCTTCGGAGGCGGCGGCGGCGGCGGCGGTGGTATCAGCGGCATGTGGACGGAAGAGAAGCCGACCATGAGGGGGCTGAACGACCGTCTCAGCTCCTACCTGGCCCGCGTGCGCGCACTGGAGCAGGCCAACGCCGCGCTG   450
G  G  G  G  F  G  G  G  G  G  G  G  G  I  S  G  M  W  T  E  E  K  P  T  M  R  G  L  N  D  R  L  S  S  Y  L  A  R  V  R  A  L  E  Q  A  N  A  A  L    135

CAAGCGCAGATCAACGCCCGCTCGGGGGTAGGCGGCGACGAGGACGCCTACGACTGGCAACCCGACCTGGACGCGGCCAGGGAGGCCCTCCTCAAGGCCAACCTGGAGCGAGCCCGCGTAGAGATCGAGAGGGACTCCTACGCCCTCGAG   600
Q  A  Q  I  N  A  A  S  G  V  G  G  D  E  D  A  Y  D  W  Q  P  D  L  D  A  A  R  E  A  L  L  K  A  N  L  E  R  A  R  V  E  I  E  R  D  S  Y  A  L  E    185

GTCGAGCAATGGAGGGGCAAGCTGGAGAGGGAGATGGACATGCGTGCTGACTTGGAGGCCGACATCAACGCCCTGAAGAGGGAGATGGACGAGGCTAACATGGCCAAGGTCGATCTGGAGGGCCAGATTGAGGGCGCCAAGAGCGAGCTG   750
V  E  Q  W  R  G  K  L  E  R  E  M  D  M  R  A  D  L  E  A  D  I  N  A  L  K  R  E  M  D  E  A  N  M  A  K  V  D  L  E  G  Q  I  E  G  A  K  S  E  L    235

GAGTTCATGAAGTCTGTTCATGAGCAGGAAGTGAAGGATCTGCGTGACAGGATCGGCGCTGGCGGGACGATGGACATCCAGATGGGCCAGGGCAGCTCGCAGGACCTGGTGGCCGCCGTGAAGGCCATCCGTGAGGAGTACGAGGAGATC   900
E  F  M  K  S  V  H  E  Q  E  V  K  D  L  R  D  R  I  G  A  G  G  T  M  D  I  Q  M  G  Q  G  S  S  Q  D  L  V  A  A  L  K  A  I  R  E  E  Y  E  E  I    285

GCCCAGGAAGAACAAGGAGGACGTCGAGAGGCAGTTCGAGAAGAAGGCCGGAGACTGTGAAGCAGGAGGCGTCACAGAACGTCGAGGCGGCTTTCCATGGCGAAGTCTGAGGTGAAGGAGATGAAGAGCCAGGTGCAGGGACTGATGGCCGAG   1050
A  R  N  K  E  D  V  E  R  Q  F  E  K  K  A  E  T  V  K  Q  E  A  S  Q  N  V  E  A  A  S  M  A  K  S  E  V  K  E  M  K  S  Q  V  Q  G  L  M  A  E    335

CTGGAGGCACTCAAGGCTATGCAACGCTCACTTGAGGAGCAGATCGCCGAAGCCGAGCGCCAACAACGCCGAGGCTCTGGAGGGCAAGTCGATGATCCTCGAGCAGCTCAAGCAGGAGATCGCCCGTCTGAAGGGGAGAGATGTCCGCCACC   1200
L  E  A  L  K  A  M  Q  R  S  L  E  E  Q  I  A  E  A  E  R  N  N  A  E  A  L  E  G  K  S  M  I  L  E  Q  L  K  Q  E  I  A  R  L  K  G  E  M  S  A  T    370

CTCAAGAGTTACAACGACATGATGAAGACCAAGTTGGCTCTGGAGGAAGAGATCGGCATCTACAACAACCTCCTGCAAGGCGAGGAAGGCCGTTTCTCTAGCATTGGCGAGTGATTTACATCTACAAAGAGCCGCAGTTTCGACGGGAAA   1350
L  K  S  Y  N  D  M  M  K  T  K  L  A  L  E  E  E  I  G  I  Y  N  N  L  L  Q  G  E  E  G  R  F  S  S  I  G  E  *    422

TGAGAAATTTTCCGTCAAGATGGCGGCTGCTCGACGCCACAAGATGGCGGACCGGCCAAACAAAATACCCCGGATGAAAGGAGCCAATCGCAGTAGAGCATCACCCTCTCATCTCATCTCAGCAAATATTCACATTCTTTCACAGTGTAC   1500
GTCTTTTCTACACGATCTCGTAGTACTTAAAACACTCGTTCTTTCAAACATTTATATCTCGGTTTTTGAAGTAAATGATGCGTTTGTCTTTGGTACGGTTTGATGTGAAAAGTAGAGCATGTAAAGCGTTGTTTTCAGATCAAAGGTTAT   1650
TGAATGTAGCTGACTAAAAAAAGCGAACAGACCGGTTAAATGCATAGTAACGGGCCGTTGCCATGTAATTTTGAATCGGTATAAATGTCGTCAGATGGGCGTGGTACAATCTTTAACGAGGTAAGGCCCGGAGCGAAGCGTTTGTGTTTG   1800
ACAGTAGAAAGACTTGTGATATGTGTTATCATTTATAACTTTTTTCATTCTTAGTACTTCTTGCGAATAAATTGGCCTCCTTTTCACGAAAAAAAAAAAAAAAAAAAAA   1909
```

**E2**

```
...GCCTGGGAGAAGTCCGCGTGGATCGGTCCGGTGACAAGGACGAGCTGGCCGGGCTGAACGACCGATTCGCCAGCTTCATCAACAAGGTGCGCTATTTGGAAGAGATGAATCGTAAGCTTACCTTGCAACTGGAGATGGTACTGAAGA   147
       L  G  E  V  R  V  D  R  S  G  D  K  D  E  L  A  G  L  N  D  R  F  A  S  F  I  N  K  V  R  Y  L  E  E  M  N  R  K  L  T  L  Q  L  E  M  V  L  K    48

AGTCGGGCGCAGGAGCACCGGACATGGGCAAAGAGTTGGAGGCAGAGCTGAACAACATCAGGAAGCTGATCGAGGTGGTCAACAACGAAAAGAACGCCATGAACTCCGAGAAAGACGGGCTACAGGGAGAGGCTGCTACCCTTAAAGCAA   297
K  S  G  A  G  A  P  D  M  G  K  E  L  E  A  E  L  N  N  I  R  K  L  I  E  V  V  N  N  E  K  N  A  M  N  S  E  K  D  G  L  Q  G  E  A  A  T  L  K  A    98

GCTATGAGGAGGAGCAGACTCGGAATGAGGGCCTGCGCGATGAGATCACAGCCCTCGAGGCCCGACGTGGATAATGTGAGCGTTGAACGTGTTGACCTGGAGGCCCGTCTGGACACCATCAAGGCAGAGATCGACTTCCTCAAGGAGGTGT   447
S  Y  E  E  E  Q  T  R  N  E  G  L  R  D  E  I  T  A  L  R  P  D  V  D  N  V  S  V  E  R  V  D  L  E  A  R  L  D  T  I  K  A  E  I  D  F  L  K  E  V    148

ACGCCGCGGAGCTGGATGCCCTCCGAGGACAAATCATTGACGATGGTCCTGGTGGGGTTACGACAATTGACGTCGGCGGCGCCCCAGACATCGACTTCAACAGCATCTTGGAGGAAGTGCGCGCCCAGTATGACGCTCTGGCCACAGCAG   597
Y  A  A  E  L  D  A  L  R  G  Q  I  I  D  D  G  P  G  G  V  T  T  I  D  V  G  G  A  P  D  I  D  F  N  S  I  L  E  E  V  R  A  Q  Y  D  A  L  A  Q  Q    198

CTAAGATGGAGTCGCAGGGCTGGTATGACACGAAGGTTCAGGGGCTGACCCAGCAGCAGGACCAGGCCGGCCGCGGATTGGACGACGCACGCGCGGAGTTGAGTAAGATGACAAGCGAGCTGAGGAAGGCTCCAGGCACAGATTGAAGCTG   747
S  K  M  E  S  Q  G  W  Y  D  T  K  V  Q  G  L  T  Q  Q  Q  D  Q  A  G  R  G  L  D  D  A  R  A  E  L  S  K  M  T  S  E  L  R  R  L  Q  A  Q  I  E  A    248

CCAAGGCCCAGGAATGCCCAGCTCCAGACCAGCTGACTGCTGTTGAGGAACGCGGCGTGAAGCAACTTGAAGCTAAGCAGGAAGAGATCACCAAGCTGGAGGAAGAGCTGACCACCATTCGCAGCAAGATCAAGGAGCAGATGGTAGAGT   897
A  K  A  R  N  A  Q  L  Q  D  Q  L  T  A  V  E  E  R  G  V  K  Q  L  E  A  K  Q  E  E  I  T  K  L  E  E  E  L  T  T  I  R  S  K  I  K  E  Q  M  V  E    298

ACCAGGCACTGATGGCAGTTAAG...   920
Y  Q  A  L  M  A  V  K    306
```

Fig. 2. Nucleotide and predicted amino acid sequence of novel cDNAs for cytoplasmic IF proteins from *Branchiostoma l.* The translational stop codon is marked by an asterisk. A: cDNA for epidermal IF protein E1. The predicted molecular weight and isoelectric point are 48 891 and 4.59, respectively. B: Incomplete cDNA for epidermal IF protein E2. Peptide sequences obtained by microsequencing are underlined.

with 2 partial clones (our unpublished results) the multigene family of the lancelet covers at least 12 distinct cytoplasmic proteins. This is one of the largest groups of related proteins currently established in *Branchiostoma*, which is considered to be the closest relative of the vertebrates. To identify those IF proteins which are true homologues of vertebrate keratins due to obligatory heteropolymer formation we have analyzed a defined epithelium. In *Branchiostoma* the epidermis is a simple epithelium, rich in IF [17] and easy to dissect [14]. Two dimensional gels identified 3 major (E1, E2, D1) and one moderately abundant (C2) IF protein (Fig. 1). Use of purified recombinant proteins showed that only the stoichiometric mixture of E1 and D1 was able to form 10-nm filaments in vitro (Fig. 3). Thus parallel to sequence relations (see Section 3) E1 and D1 are identified as true homologues of type I and II keratins, respectively. The strong sequence homology of E2 and D1 (see Section 3) identifies also the E2 protein as a type II keratin. We have extended our analysis also to the B/A subfamily of *Branchiostoma* IF proteins [14]. Recombinant B1 formed homopolymeric 10-nm filaments and thus is identified as a type III homologue (Fig. 3). Since one of the two IF proteins established in the urochordate (tunicate) *Styela* is a muscle type III homologue [15] we suggest that at least type III proteins such as desmin and most likely also keratin type I and II proteins are common to the entire chordate branch of metazoa. Future experiments have to formally document the keratin homologues in *Styela* epidermis and analyze the neuronal IF complement in both *Branchiostoma* and *Styela*. This approach should decide whether the early chordates already possess type IV homologues or whether this type, as modern neurofilaments, is restricted to vertebrates. Finally, assembly
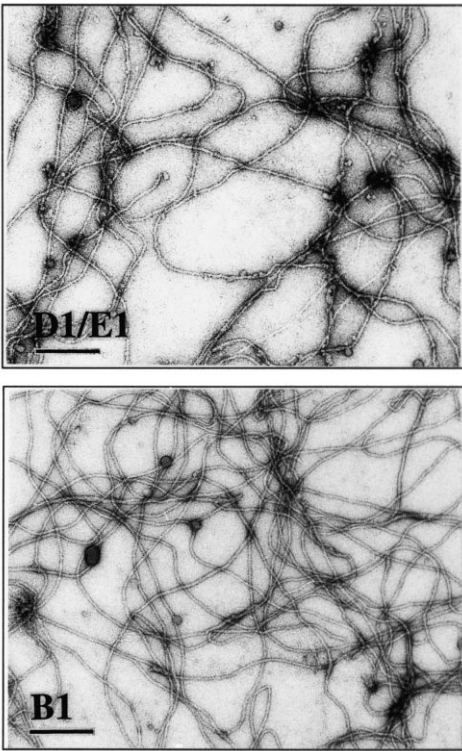


Fig. 3. Electron micrographs of *Branchiostoma* IF assembled in vitro. Stoichiometric mixture of recombinant IF proteins E1 and D1 (upper panel). Recombinant IF protein B1 (lower panel). Bar 0.2 μm.
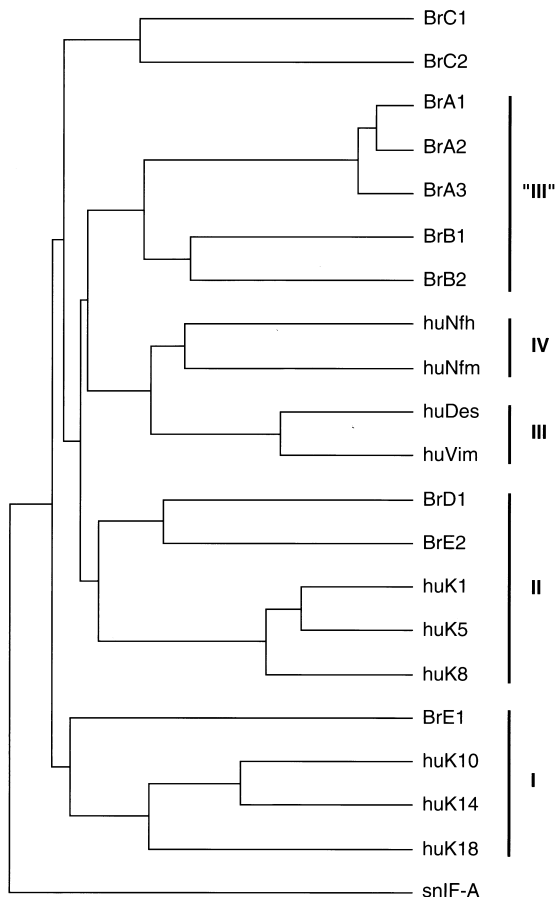
Fig. 4. Reconstruction of an evolutionary tree formed by IF proteins from the cephalochordate *Branchiostoma l.* and vertebrates. Alignment of rod domains and tree construction were performed with the Wisconsin GCG software package (version 9.0) using the UPGMA method [20]. Except for E1 and E2 (this study; Fig. 2) the other 8 *Branchiostoma* proteins were reported previously [14]. The various human IF proteins and their accession numbers (in parentheses) are: type II: keratin 1 (huK1; P04264), keratin 5 (huK5; P13647), keratin 8 (huK8; P05787); type I: keratin 10 (huK10; P13645), keratin 14 (huK14; P02533), keratin 18 (huK18; P05783); type III proteins: desmin (huDes; P17661), vimentin (huVim; P08670) and type IV proteins: NF-M (huNfm; P07197) and NF-H (huNfh; P12036). *Branchiostoma* proteins are abbreviated as Br. The IF protein of the snail *Helix aspersa* (snIF-A; P16275) represents the protostomic IF proteins (see Section 1) and served as outgroup. The four subfamilies I–IV of vertebrate IF proteins are indicated. Note the inclusion of the *Branchiostoma* proteins D1 and E2 in the keratin II group and the inclusion of the *Branchiostoma* E1 protein in the keratin I group. The *Branchiostoma* IF protein group A1, A2, A3, B1 and B2 [14] is marked 'III' (for details see text). Note also that the *Branchiostoma* IF proteins C1 and C2 form a separate branch.

keratins, agrees well with the evolutionary tree calculated from the rod sequences of human and *Branchiostoma* IF proteins (Fig. 4). An interesting aspect of the evolutionary tree concerns the relative positions of vertebrate type III and IV proteins versus the *Branchiostoma* B/A group, which is recognized as type III due to the homopolymer formation documented for B1. This part of the tree would indicate that type IV neurofilament proteins arose from a type III precursor after the separation of the vertebrate and cephalochordate lineages (Fig. 4). Although previous work on the organization of IF genes raised the possibility of such a derivation and proposed an mRNA transposition event as the origin of type IV genes, the relative time point in metazoan evolution remained unknown [12,18,19]. A precise evaluation of the origin of type IV neurofilament proteins can be expected once the catalogue of IF proteins of *Branchiostoma* and a tunicate like *Styela* is completed.

## References

[1] Fuchs, E. and Cleveland, D.W. (1998) Science 279, 514–519.
[2] Fuchs, E. and Weber, K. (1994) Annu. Rev. Biochem. 63, 345–382.
[3] Parry, D.A.D. and Steinert, P.M. (1995) Intermediate Filament Structure, Springer, New York, NY.
[4] Markl, J. and Schechter, N. (1998) In: Subcellular Biochemistry, Vol. 31, Intermediate Filaments (Herrmann, H. and Harris, J.R., Eds.) pp. 1–33, Plenum Press, New York, NY.
[5] Weber, K., Plessmann, U., Dodemont, H. and Kossmagk-Stephan, K. (1988) EMBO J. 7, 2995–3001.
[6] Weber, K., Plessmann, U. and Ulrich, W. (1989) EMBO J. 11, 3221–3327.
[7] Dodemont, H., Riemer, D. and Weber, K. (1990) EMBO J. 9, 4083–4094.
[8] Szaro, B.G., Pant, H.C., Way, J. and Battey, J. (1991) J. Biol. Chem. 266, 15035–15041.
[9] Way, J., Hellmich, M.R., Jaffe, H., Szaro, B., Pant, H.C., Gainer, H. and Battey, J. (1992) Proc. Natl. Acad. Sci. USA 89, 6963–6967.
[10] Tomarev, S.I., Zinovievam, R.D. and Piatigorsky, J. (1993) Biochim. Biophys. Acta 1216, 245–254.
[11] Dodemont, H., Riemer, D., Ledger, N. and Weber, K. (1994) EMBO J. 13, 2625–2638.
[12] Erber, A., Riemer, D., Bovenschulte, M. and Weber, K. (1998) J. Mol. Evol., in press.
[13] Riemer, D., Dodemont, H. and Weber, K. (1992) Eur. J. Cell Biol. 58, 128–135.
[14] Riemer, D., Karabinos, A. and Weber, K. (1998) Gene 211, 361–373.
[15] Riemer, D. and Weber, K. (1998) J. Cell Sci., in press
[16] O'Farrell, P.H. (1975) J. Biol. Chem. 250, 4007–4021.
[17] Bartnik, E. and Weber, K. (1989) Eur. J. Cell Biol. 50, 17–33.
[18] Lewis, S.A. and Cowan, N.J. (1986) Mol. Cell. Biol. 6, 1529–1534.
[19] Klymkowsky, M.W. (1995) Curr. Opin. Cell Biol. 7, 46–54.
[20] Sneath, P.H.A. and Sokal, R.R. (1973) Numerical Taxonomy, Freeman, San Francisco, CA.

studies may also help to understand the structural relation of the *Branchiostoma* C1 and C2 proteins, which have unique tail domains of coiled coil forming ability [14], to other IF proteins.

The identification of E1 as type I, and D1 and E2 as type II