doi:10.1006/jmbi.2001.5298 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 316, 127–137

JMB

Tissue-Specific Co-expression and *in vitro* Heteropolymer Formation of the Two Small *Branchiostoma* Intermediate Filament Proteins A3 and B2

Anton Karabinos[1], Jürgen Schünemann[1], David A. D. Parry[2] and Klaus Weber[1]*

[1]*Max Planck Institute for Biophysical Chemistry Department of Biochemistry Am Fassberg 11 37077 Goettingen, Germany*

[2]*Institute of Fundamental Sciences, Massey University Palmerston North, New Zealand*

**Corresponding author*

The two small intermediate filament (IF) proteins A3 and B2 of the cephalochordate Amphioxus were investigated. Blot overlays indicated a heterotypic interaction pattern of the recombinant proteins. While the individual proteins formed only aggregates, the stoichiometric mixture formed obligatory heteropolymeric filaments. Mutant proteins with a single cysteine residue in equivalent positions gave rise to filaments that oxidize to the disulfide-linked heterodimer, which can again form IF. Thus the A3/B2 filaments, which are expressed in the intestinal epithelium, are based on a hetero coiled coil. This keratin-like assembly process of A3 plus B2 was unexpected, since previous evolutionary tree calculations performed by two laboratories on the various Amphioxus IF proteins identified keratin I and II orthologs but left the A/B group as a separate branch. We discuss obvious evolutionary aspects of the Amphioxus IF multigene family, including the previously made observation that B1, the closest relative of B2, forms homopolymeric IF *in vitro* and is, like vertebrate type III proteins, expressed in mesodermally derived tissues.

Keywords: Amphioxus; Branchiostoma; cephalochordate; coiled coil; developmental expression

Introduction

A variety of epidermal fragility syndromes are caused by mutations in human keratin genes, suggesting that an important function of cytoplasmic intermediate filaments (IF) is to provide resistance against mechanical stress.[1] Most of the close to 60 members of the mammalian multigene family of IF proteins fall into one of the five major subgroups.[2,3] Type I and type II keratins are the largest subfamilies and give rise to the epithelial keratin filaments that are based on obligatory heteromeric double-stranded coiled coils formed by a type I and a type II keratin. Type III covers four mesenchymally expressed proteins, which, at least in part, are able to form homopolymeric IF. Type IV spans five neuronal IF proteins. Two large molecular mass proteins (synemin and paranemin), and the two IF proteins of the eye lens (phakinin and filensin) fall outside these subfamilies. Finally, type V covers the nuclear lamins.

Lamins differ from vertebrate cytoplasmic IF proteins by an extra 42 residues in the coil 1b subdomain of the central rod domain and by unique carboxy-terminal domains that harbor a nuclear localisation signal and end, in most cases, with a CaaX box.[4,5] The short coil 1b domain, first defined in vertebrate cytoplasmic IF proteins, extends to the cephalochordates and urochordates,[6–9] while cytoplasmic IF proteins from the protostomia show the long coil 1b version of the nuclear lamins and, in most cases, display a lamin homology segment in their tail domains.[10,11]

Recent studies on the lancelet, the cephalochordate Branchiostoma (Amphioxus),[6,12–14] and the urochordate Styela[8,9] document that the early chordates have among their cytoplasmic IF proteins

Abbreviations used: IF, intermediate filament; PVDF, poly(vinylidene fluoride).

E-mail address of the corresponding author: akarabil@gwdg.de

0022-2836/02/010127–11 $35.00/0

orthologs of the type I to III proteins originally defined in the vertebrates.

We previously cloned 13 cytoplasmic IF proteins from Amphioxus. Evolutionary tree calculations indicated three type I and two type II keratins. This assignment was confirmed by the obligatory heteropolymeric filament formation of the recombinant proteins. Any stoichiometric mixture of type I (k1, Y1 and E1) and type II (D1 and E2) proteins provided IF. In addition, two of the lancelet type I keratins formed chimeric IF when mixed with human keratin 8, a type II keratin.[6] Three keratins (k1, Y1 and D1) and protein X1 are the only IF proteins expressed in the gastrula, which is reminiscent of the expression of two type I keratins (K18 and K19) and one type II keratin (K8) in the murine gastrula.[15] The number of lancelet IF proteins increases at the neurula and early larval stages to seven and 11, respectively. However, 13 different proteins are found in the adult. The keratins are the major IF proteins in the Amphioxus nerve cord. Proteins X1 and C2 were found, by immunoelectron microscopy, to be integrated into the epidermal and nerve cord IF, which are built from keratins. The latter two proteins and the C1 protein seem to be lancelet-specific IF proteins.[13] The five remaining IF proteins form the A/B group in the evolutionary tree previously calculated (Figure 4 of Karabinos *et al.*[6]). So far, only the B1 protein has been characterized. It can form homopolymeric IF *in vitro* and is expressed in the mesodermally derived muscle tails and in coelomic epithelia. These properties indicate some type III character for the B1 protein.[12,13] Here, we show that A3 and B2, which are co-expressed in the intestinal epithelial cells, form obligatory heteropolymeric IF based on a hetero coiled coil consisting of one A3 and one B2 polypeptide chain.

Results

Branchiostoma IF proteins A3 and B2 essentially lack a tail domain

Cloning and recombinant expression of A3 and B2 have been described.[6,7] During the current study we noted that the apparent molecular masses of both proteins observed by SDS-PAGE (A3, about 37 kDa; B2, about 39 kDa) were smaller than the masses calculated from the protein sequences predicted from the cDNAs (A3, 41,679 Da; B2, 49,326 Da). Automated sequencing of the proteins blotted on to a poly(vinylidene fluoride) (PVDF) membrane provided the predicted N-terminal sequences. The experimentally inferred masses derived by mass spectrometry (A3, 37,278 Da; B2, 39,694 Da) were lower than the values calculated from the predicted sequences. We therefore subjected the coding fragments to a detailed sequence analysis and found that the original analysis[7] had overlooked an earlier stop codon. Thus, the open reading frames of A3 and B2 were shortened by 39 and 92 codons, respectively (the GenBank accession numbers of the revised sequences are AJ223575 and AJ223576). The molecular masses calculated from the revised sequences are in good agreement with those obtained experimentally.

The alignment in Figure 1 shows that IF proteins A3 and B2 essentially lack a C-terminal tail domain. Only two (B2) and four (A3) residues follow the well-conserved end of the rod domain. Unusually short tail domains are known for the Amphioxus keratin E1 and the human keratin 19 ([6,12] and references therein).

Blot overlays reveal heterotypic interactions between A3 and B2

The ten currently available recombinant IF proteins of the lancelet were analysed in blot overlays using biotinylated A3 and B2 proteins. After extensive washing steps, the biotinylated proteins were detected by the streptavidin-peroxidase system. Figure 2 (middle panel) shows that biotinylated A3 binds strongly to B2. Except for a very weak reaction on B1, no other IF protein was detected by the A3 probe. The biotinylated B2 protein reacted strongly with the A3 protein and more weakly with the protein B1 (Figure 2, lower panel). These results show a strong heterotypic interaction between A3 and B2, and indicate a lack of homotypic interactions, i.e. A3 with A3 or B2 with B2, in the overlay system. These observations raised the question of whether A3 and B2 could form heteropolymeric IF.

Equimolar amounts of A3 and B2 form heteropolymeric IF

Purified recombinant proteins A3 and B2 were dialysed either alone or in equimolar mixture to remove the urea, and the self-assembly products were analysed by electron microscopy after negative staining. While the individual proteins yielded only aggregated material, the mixture of A3 and B2 formed long IF (Figure 3). Thus, A3 and B2 lack the ability to form homopolymeric IF but participate in obligatory heteropolymeric IF assembly.

The structural unit of A3/B2 filaments is the heterodimer

To decide whether the A3/B2 filaments are based on a A3/B2 heterodimer or the A3 and B2 homodimers, we used the approach previously taken to demonstrate the heterodimeric nature of keratin filaments.[9,16] It is based on the observation that in parallel and unstaggered α helices of a double-stranded coiled coil, a cysteine residue present in an *a* or *d* position of the heptads can form a disulfide bond crosslink, provided both α helices have the cysteine residue in equivalent positions. Thus we aimed at mutants of A3 and B2 that have a single cysteine residue, which is located in the same position as in the previously studied keratins.[9,16] This position is exactly 40 residues from

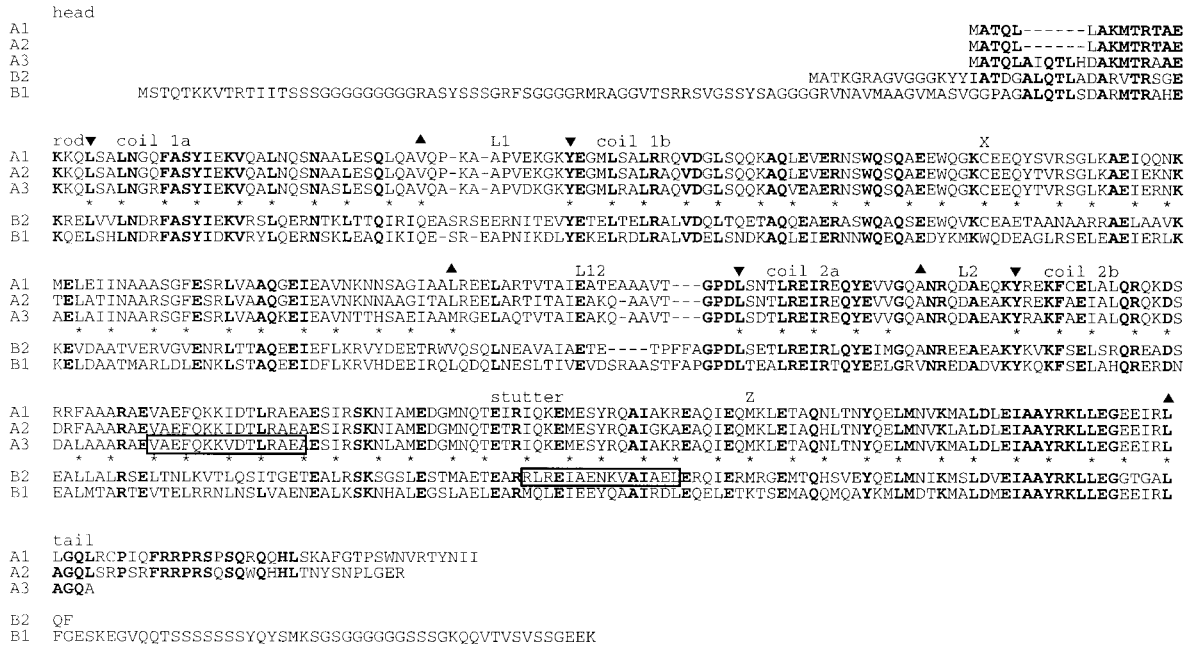

Figure 1. Alignment of the amino acid sequences of the five related Branchiostoma IF proteins A1, A2, A3, B1 and B2. The typical structural organisation consisting of head, rod and tail domains is indicated. Arrowheads pointing down or up mark the start and end of the four subdomains of the helical rod domain (coils 1a, b, 2a and 2b), which are connected by three non-helical linkers (L1, L12 and L2). Asterisks (*) between the rod domains of A3 and B2 mark the *a* and *d* positions of the heptad repeat pattern with the stutter in coil 2b (for details on IF structure, see Fuchs & Weber[2] and Parry & Steinert[3]). Dashes are used to optimize the sequence alignment. Bold letters highlight similarities in head or tail domains and indicate residues in the rod domains that are identical in all five proteins. Sequences of two synthetic peptides used for antibody production are boxed. The capital letters X and Z shown above the sequences indicate positions that were changed by site-specific mutagenesis. The cysteine codons TGC of the A3 and B2 sequences were changed to the serine codons AGC (marked X). The methionine codons ATG of the A3 and B2 Cys mutants were changed to the cysteine codons TGC (see Results for details). The resulting mutant A3 and B2 proteins contained a single cysteine residue (position Z). The GenBank accession numbers of the A3 and B2 sequences are AJ223575 and AJ223576, respectively.

the end of the rod and involves a heptad *d* position (Figure 1, marked Z). In the first round of site-specific mutagenesis, we changed the single cysteine residue of A3 and B2 (marked X in Figure 1) to serine. In the second step, the methionine residue marked Z in Figure 1 was changed to cysteine. The mutations were confirmed by DNA sequencing and the mutant recombinant proteins were purified.

Filaments were formed by the A3 and B2 mutant proteins in the presence of 2-mercaptoethanol. Subsequent extensive dialysis removed the reducing agent and allowed cystine formation. One and two-dimensional gels run under non reducing and reducing conditions (Figure 4) identified as major components the A3/B2 disulfide-linked heterodimer as well as the A3 and B2 monomers. Dimers of A3 and B2 were not observed (Figure 4(b)). The oxidized filaments were dissolved in 8 M urea without reducing agent. Ion-exchange chromatography on Mono Q using the same solvent yielded the purified heterodimer with a small amount of contaminating monomeric B2 (Figure 5(a)). The assignment of the A3/B2 heterodimer and the B2 contaminant was verified by automated sequencing from a protein blot onto a membrane. Figure 5(b) shows that the disulfide-linked A3/B2 heterodimer forms very long IF in the absence of reducing agent. Thus, the A3/B2 filaments are built from double-stranded hetero coiled coils. This result is consistent with the strong heterotypic interaction pattern of A3 and B2 proteins in the blot overlays (Figure 2).

Tissue-specific co-expression of B2 and A proteins in the adult and the early larvae

To determine expression patterns of the A3 and B2 proteins, we raised two peptide antibodies, which were antigen affinity-purified. Figure 6 shows the results of immunoblotting on ten recombinant lancelet IF proteins. The B2 antibody recognized only B2, while A3 antibody detected only A3. The A3 antibody, however, detects also A1 (Figure 6(b)) and A2 (data not shown). This was expected from the high level of sequence homology of three A proteins. For instance, the sequences of A1 and A2 corresponding to the A3 peptide used as antigen differ only by a single valine to isoleucine exchange (Figure 1). We therefore refer to this

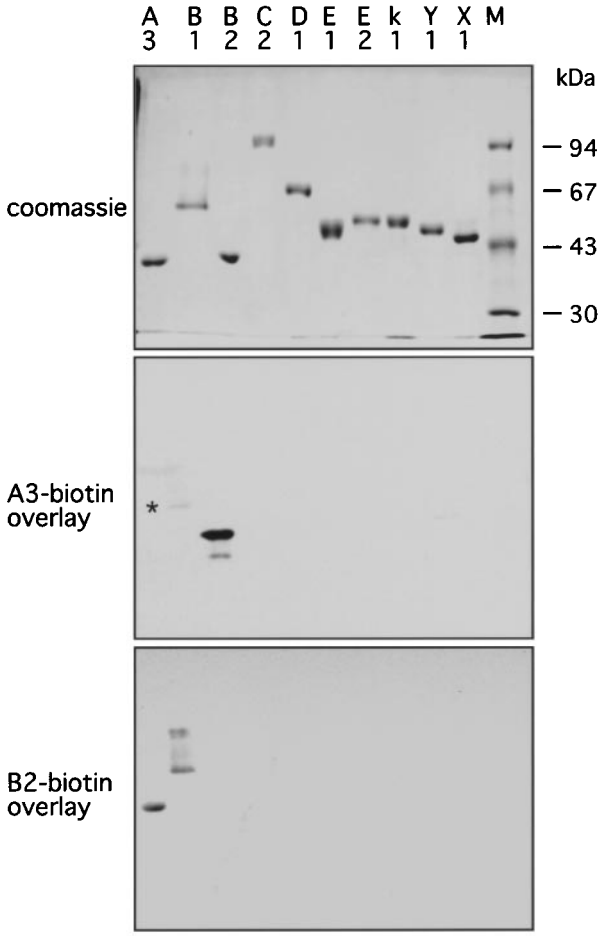

Figure 2. Blot overlay search for interactions of the biotin-labeled A3 and B2 proteins with blotted immobilized Branchiostoma IF proteins. Equal amounts of purified recombinant lancelet proteins A3, B1, B2, C2, D1, E1, E2, k1, Y1 and X1 were separated by SDS/10% PAGE and stained with Coomassie brilliant blue (upper panel) or transferred to nitrocellulose membranes. The middle and lower panels show the blots after incubation and staining with biotinylated A3 and B2 proteins, respectively (marked as A3-biotin and B2-biotin on the left). The biotinylated A3 (middle panel) strongly decorates the B2 protein and a proteolyzed derivative. There is a very weak reaction on B1 (marked by the asterisk (*)) but all other proteins were negative. The biotinylated B2 protein (lower panel) recognizes the protein A3, and B1 plus some aggregate of it. All other proteins were negative with biotinylated B2. Marker proteins (M) and an approximate molecular mass standard (in kDa) are given at the right of the upper panel.

antibody as the A-specific antibody. Used in immunofluorescence microscopy on frozen sections of adult *Branchiostoma floridae*, the antibodies demonstrated co-expression of A and B2 proteins in the intestine (Figure 7(a)-(d)). The staining was mostly concentrated into the apical part of the simple columnar epithelium of the intestine.[17] In addition, both antibodies provided a regular staining of small areas surrounded by the two connecting myomers and the epidermis (marked with arrows in Figure 7(e)-(h)). In the early (one to two days) *B. floridae* larva, the intestinal cells were the only structures stained by the two antibodies (Figure 8).

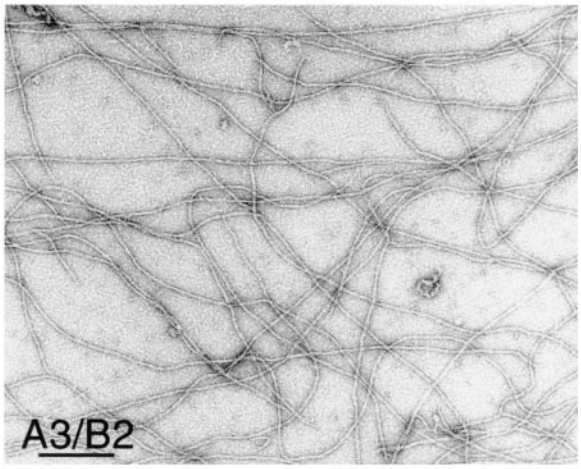

Figure 3. Electron micrograph of Branchiostoma IF assembled *in vitro* from equal amounts of recombinant proteins A3 and B2. The scale bar represents 0.2 μm.

Calculation of interchain ionic interaction between the A and B proteins

The ionic interactions that could occur in principle between the chains that constitute a coiled coil molecule are readily calculated.[18,19] Such interactions are known to play an important role in spe-

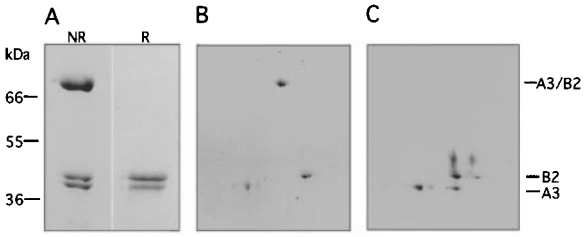

Figure 4. Identification of cysteine-linked heterodimers of mutant recombinant A3 and B2 proteins. Filaments produced in the presence of 2-mercaptoethanol in filament buffer were dialysed extensively against the same buffer in the absence of reducing agent to allow oxidation. Filaments were harvested by centrifugation. An aliquot was solubilized in 8 M urea and analysed by non-reducing and reducing SDS-PAGE (lanes R and NR in (a)) and by two-dimensional gel-electrophoresis under non-reducing conditions in the first dimension and (b) non-reducing or (c) reducing conditions in the second dimension. Positions of the A3 monomer, the B2 monomer and the A3/B2 cross-linked dimer are marked on the right as A3, B2 and A3/B2, respectively. Note the high yield of the disulfide-linked heterodimer A3/B2 and the absence of homodimers of A3 and B2.

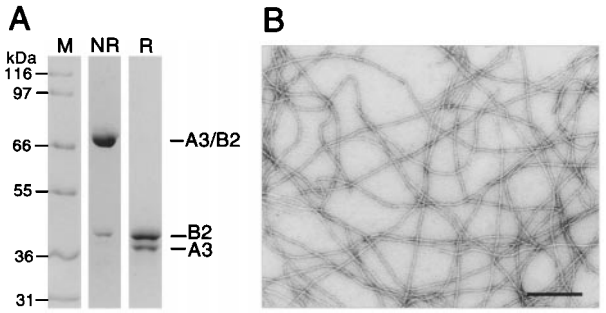

Figure 5. Purification of disulfide-linked A3/B2 heterodimer and its ability to form IF *in vitro*. The disulfide-linked heterodimer was purified from the harvested oxidized filaments (see Figure 4). Filaments were dissolved in 8 M urea buffer without reducing agent and the protein was subjected to ion-exchange chromatography on Mono Q, which removed the A3 and B2 monomers. An aliquot of the Mono Q-purified heterodimer was analysed by non-reducing and reducing SDS-PAGE (lanes R and NR in (a), respectively) and was found to contain a small amount of monomeric B2. The nature of all polypeptides was confirmed by automated sequencing after blotting onto a PVDF membrane. Marker proteins (M) and an approximate molecular mass standard (in kDa) are given at the left. (b) The ability of the disulfide-linked A3/B2 heterodimer to form long IF in the absence of reducing agent. The scale bar represents 0.2 μm.

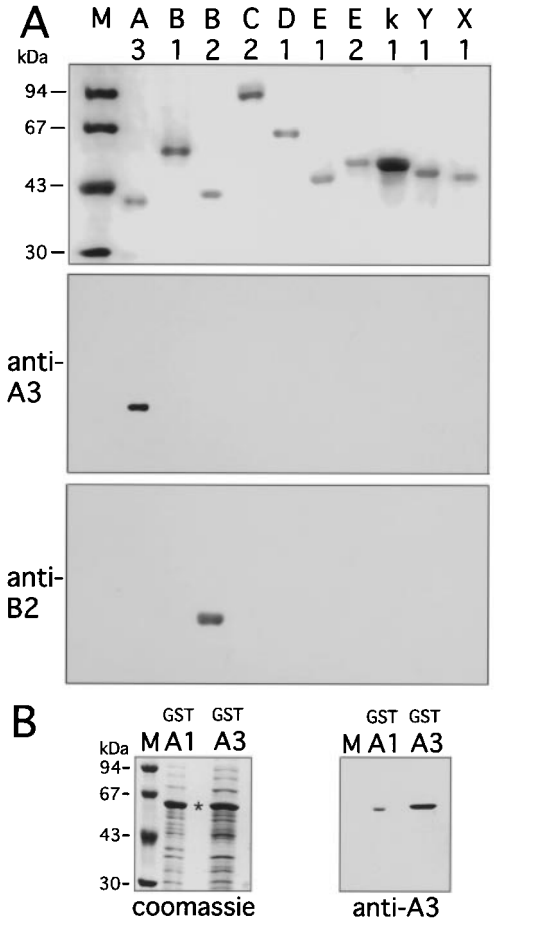

Figure 6. Immunoblot analysis of Branchiostoma recombinant proteins. (a) Equal amounts of purified proteins A3, B1, B3, C2, D1, E1, E2, k1, Y1 and X1 were separated by SDS/10% PAGE and stained with Coomassie brilliant blue (upper panel) or transferred to nitrocellulose membranes. Blots were incubated with affinity-purified antibodies to Branchiostoma A3 and B2 proteins as indicated on the left. Both antibodies recognised exclusively the corresponding protein and no cross-reactivity with other IF proteins was observed. Marker proteins (M) and an approximate molecular mass standard (in kDa) are given at the left of the upper panel. (b) Bacterial lysates containing equal amounts of the recombinant A1 and A3 GST-fusion proteins (marked by asterisk (*)) were separated by SDS/10% PAGE and stained with Coomassie brilliant blue (left panel) or immunoblotted with affinity-purified antibody A3. The antibody strongly stained the A1 and the A3 GST-fusion protein bands (right panel). Marker proteins (M) and an approximate molecular mass standard (in kDa) are given at the left.

cifying chain assembly and, in particular, the relative chain stagger and polarity. For IF proteins, the preferred calculated arrangement has been without exception for the chains to lie parallel with one another and in axial register. Experimental observations have confirmed these calculations.[3] In the case of Branchiostoma, the interchain ionic interactions have been determined for a total of 15 chain combinations involving three members of the A family and the two members of the B family. In all cases, the preferred arrangement is for the chains to lie parallel with one another and to be in axial register. All of the A/B heterodimers score more highly than the A/A homodimers and A/A heterodimers (Table 1). This is not surprising, since keratin heterodimers are the norm. However, the (joint) highest calculated scores are for the B1 homodimer and the B1/B2 heterodimer (score +16). This is consistent with the observation of B1 homodimers[12] but suggests, in theory, that B1/B2 heterodimers might also be viable. As regards A3, its preferred partners are B1 and B2 (score +14). Overall, there is good but not perfect correlation between the experimental observations and the theoretical calculations. However, X-ray crystallographic studies on synthetic polypeptides that adopt a coiled coil conformation indicate that only some of the possible interchain ionic interactions are actually made *in vivo*. The calculations must thus be seen as indicative rather than definitive.

Designation of chain type of the 13 *Branchiostoma* IF proteins

Each of the 13 Branchiostoma chains (A1, A2, A3, B1, B2, C1, C2, D1, E1, E2, k1, X1 and Y1) was compared with the consensus sequences unique to the type I and type II keratin chains (Figure 9;[20] but updated in line with the additional data now

Table 1. Calculation of interchain ionic interactions of individual rod homo- and heterodimers lying parallel and in axial register

IF chain	A1	A2	A3	B1	B2
A1	8	7	9	13	12
A2	7	10	8	12	12
A3	9	8	10	14	14
B1	13	12	14	16	16
B2	12	12	14	16	12

The numbers show potential ionic interactions calculated for charged residue pairs in positions 2e′-1g, 1;g′-2e, 2a′-1g, 1g′-2a, 1e′-1d, and 1d′-1e of the two chains. Attractive interactions between oppositely charged residues were scored +1 and interactions between residues that are either both positively or both negatively charged were scored −1. All other interactions were scored 0.[21]

available). In eight cases, the analysis (Table 2) was such that the chain type became immediately apparent (type I, B1, B2, C2, E1, k1 and Y1; type II, D1 and E2). In three instances, the designation of chain type, while less definitive, clearly favoured one option (type II, A1, A2 and X1). Two chains, however, were more difficult to type unequivocally (A3 and C1). For the A3 chain, the data pertaining to the entire sequence slightly favoured type I designation. However, the analysis relating to the residues in the non-*a* and non-*d* positions firmly favoured the chain being type II (see Materials and Methods). For this reason, A3 was specified a type II chain. The situation was even more difficult for the C1 chain, where the data only slightly favoured type II designation on the basis of the consensus sequence in all positions of the heptad repeat but proved effectively indeterminate in selecting chain type using the non-*a* and non-*d* position data. The decision to designate C1 a type II chain on the basis of sequence homology was thus a marginal one.

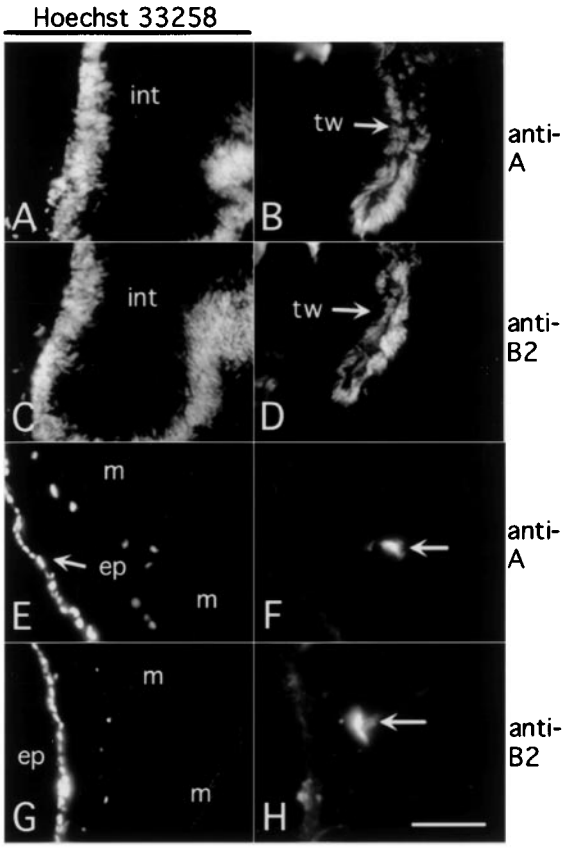

Figure 7. Tissue-specific co-expression of A and B2 IF proteins in adult Branchiostoma. Frozen cross-sections were stained with Hoechst 33258 (DNA staining; (a), (c), (e) and (g)) to facilitate the identification of cells and tissues. Indirect immunofluorescence microscopy of parallel sections labeled with anti-A and anti-B2 antibodies as indicated on the right. (b) and (d) The terminal web (tw) of the intestinal cells (int) is stained by both antibodies. Arrows in (f) and (h) indicate the area between two myomeres (m) and the epidermis (ep), which shows significant expression of both proteins. The scale bar represents 20 μm.

Evolution of the *Branchiostoma* IF proteins

Bootstrap maximum parsimony analysis of the rod sequences from all 13 Branchiostoma IF proteins, various IF proteins representing vertebrate

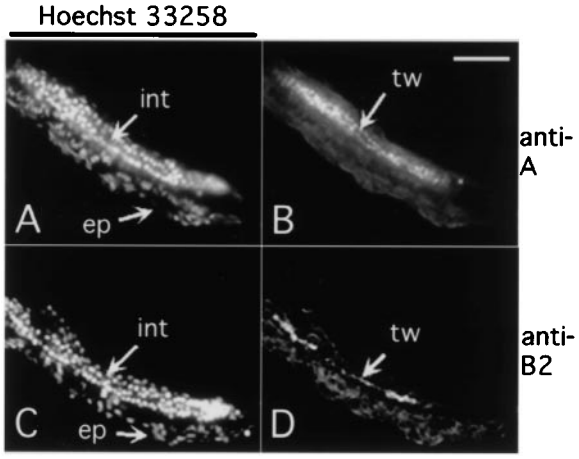

Figure 8. Tissue-specific co-expression of A and B2 IF proteins in the Branchiostoma early larva. Frozen longitudinal sections were labeled with (b) A and (d) B2 antibodies and (a) and (c) simultaneously with Hoechst 33258 as indicated. (b) and (d) Expression of A and B2 proteins in the terminal web (tw) of the intestinal simple epithelia (int) is indicated. The epidermis (ep) and other body parts were negative. The scale bar represents 20 μm.

```
coil 1a:                                    RD
type I    K.TMQNXXDRLXXYLXXXXAXXE.X.EXEVXIKEW
type II   REQIK.XXNKFXXFIXXXXFXXQQXKVX.TXW.LL
                                        L

coil 1b:              R                     R
type I    YS.YF.TX.DXK..I..AT.DNA...LQIDNAKLAADXFR.XFX.X..LX..VXADINGX
typeII    DPLFE.YX..XRR..D....D.GRL..ELR..QDLVEXYRKXYXDXINKXT.AXNDFV.X
          E                     E         K         FKN    E          E

                  R
type I    RRVLXELTL.K.DXE..IE.X.EXLAYXKKNHEEX....R.
type II   KKDVX.AYM.K.KKEAK.D.X..XIDFXR.LY..XLSQ.Q.
                       Q          N

coil 2a:        N
type I    DX..XLSXMR.XXE...E.
type II   DXDSXI.XVKAXXD.IA.R
          N             E

linker L2:  R
type I    NRKDXX.W
type II   SRAEXX.W
           K     L

coil 2b:                                  R
type I    F....EEX..EV..N.E..QS..SXXTELKXT.X.XEIXLQSQL..K.SXE..LEETX.R
type II   YQSKY..X..TAGRHGD..RNSK.XX.D.NX.IXRXRSXID..KKQ...XQA.IA.AXQ.
            T F       V K      K T      E         TA  E          EQ

                     G V       L  D D
type I    YC.QX.QIQ..IS.LXEQXAQIR.E.E.QN.XX..XLD.XTRXE.XXXXXXRXXXXX....
type II   GE.AXKDA..KL..LXDAXQ.AKQDMARLLRXXQEXMN.XLAXD.XXXXXX.XXXXXE.R.
                         E         L  Q K
```

Figure 9. Consensus sequences using the one-letter code for type I and type II keratin intermediate filament chains for rod domain segments 1a, 1b, 2a, L2 and 2b. These are based on the earlier work by Conway & Parry[20] but are updated in line with the additional data now available. Where there is no clear consensus or where three or more different types of residue are seen in a particular position, then that position is indicated by a dot. The symbol X implies that the same consensus residue is observed in both the type I and type II chains, and as it is not a distinguishing feature of either chain type it plays no further part in designation of chain type.

type I to type IV and several additional invertebrate IF proteins resulted in two equally parsimonious phylograms (Figure 10(a)). The majority-rule consensus of these trees clearly demonstrates the presence of the type I (E1, k1 and Y1) and type II (D1 and E2) keratin orthologues in Branchiostoma. This supports a direct common ancestry of all these sequences and the vertebrate keratins from archetypal type I and type II genes. On the other hand, the A and B genes, and the C and X1 genes of Branchiostoma appear on the separate A/B and C/X branches, of which only the former had strong bootstrap support (99 %; Figure 10(a)). We[6,12] and Luke & Holland[14] reported similar relationships within the Branchiostoma sequences, and between Branchiostoma and vertebrate sequences using the unweighted pair-group method with arithmetic mean (UPGMA) and the bootstrap neighbor-joining method, respectively. Thus, the A, B, C and X1 molecules seem to have no orthologue in either vertebrates or urochordates.[9] This raises the possibility that they evolved only after the divergence of the cephalochordate lineage. On the other hand, since these sequences possess some type I or type II character (Table 2), a possible common ancestry with the five lancelet kera-

Table 2. Theoretical designation of chain type in the 13 Branchiostoma IF proteins

IF Chain	Species	Accession number	All[a]	Non-*a* and non-*d*[a]	Calculated chain type	Evidence from experiments
A1	*B. floridae*	AJ223574	−1.8	−3.9	II	Probably like A3
A2	*B. floridae*	AJ223582	−3.3	−5.3	II	Probably like A3
A3	*B. floridae*	AJ223575	+0.7	−4	II	Heteropolymeric IF with B2 (this study)
B1	*B. lanceolatum*	X64522	+12.7	+7.9	I	Forms homopoly-meric IF[b]
B2	*B. floridae*	AJ223576	+6.8	+6.5	I	Heteropolymeric IF with A3 (this study)
C1	*B. floridae*	AJ223577	−1.9	+0.1	II	Not determined
C2	*B. lanceolatum*	AJ223578	+5.1	+5.8	I	Present in keratin IF[c]
D1	*B. lanceolatum*	AJ223579	−19.7	−12.8	II	Type II[d]
E1	*B. lanceolatum*	AJ010294	+20.7	+16.9	I	Type I[d]
E2	*B. lanceolatum*	AJ010293	−15.7	−13.6	II	Type II[d]
k1	*B. lanceolatum*	AJ245426	+12	+3.8	I	Type I[d]
X1	*B. lanceolatum*	AJ245427	−1.6	−5.2	II	Present in keratin IF[c]
Y1	*B. lanceolatum*	AJ245428	+10.2	+2.4	I	Type I[d]

[a] The numbers of identities between the Branchiostoma chains and the type I and type II consensus sequences were calculated both for the entire sequence (All) and for those residues falling in the non-*a* and non-*d* positions of the heptad repeat (Non-*a* and non-*d*).

[b] See in Karabinos *et al.*[12]

[c] Immunoelectromicroscopical results (Karabinos *et al.*[6]

[d] Obligatory heteropolymeric filament assembly, E1 and Y1 form IF with human keratin 8, a type II keratin (Karabinos *et al.*[13]). Note that there is only one disagreement between the calculated chain type and the experimental evidence. B1 although calculated as type I chain is able to form homopolymeric IF *in vivo* (see the text).

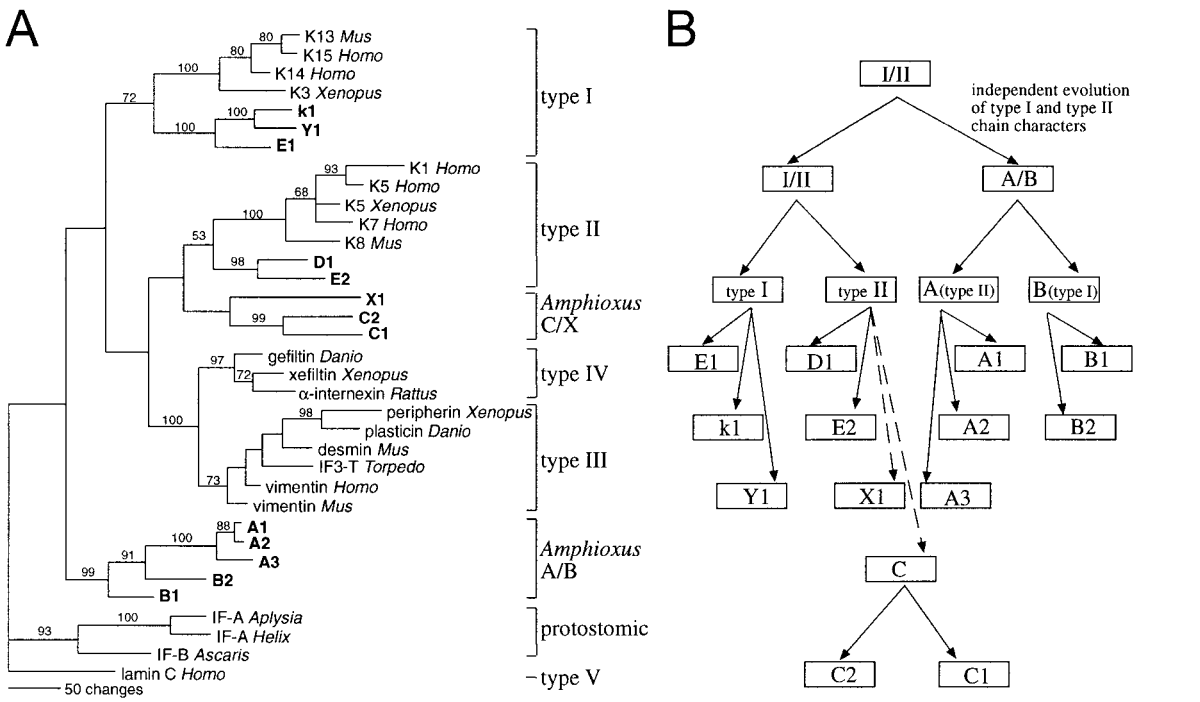

Figure 10. (a) Phylogenetic analysis of Branchiostoma and other IF rod sequences from invertebrates and vertebrates. Majority-rule consensus phylogram of weighted bootstrap (1000 replications) maximum parsimony analysis[22] of individual IF sequences. The bootstrap values are shown above the internal nodes. The five subfamilies I-V of vertebrate IF proteins, the A/B and C/X subfamilies of Amphioxus IF proteins as well as the protostomic IF branch are indicated. (b) Hypothetical scheme for the evolution of the 13 Branchiostoma IF genes. The indicated relation of C and X1 to the type II gene cluster is interpreted as provisional result (see Results). This scheme suggests that the more or less significant keratin characters of the A and B genes (see above and Table 2) has probably been the result of an independent convergent evolution (see the text).

tins is indicated. In addition, immunoelectron microscopy on tissue sections shows that C2 and X1 occur along IF built from keratin orthologues.[13] They may therefore reflect some kind of keratin-related or keratin-derived molecules that have evolved more rapidly than the keratin orthologues. Based on these considerations, we propose a hypothetical scheme of the evolution of IF protein in Branchiostoma (Figure 10(b)). The single archetypal ancestor of both keratin groups (typeI/typeII gene) was duplicated into typeI/typeII and A/B genes. This was followed by additional rounds of gene duplications leading to the three type I genes (E1, k1 and Y1), the five type II genes (D1, E2, C1, C2 and X1), the three A genes (A1, A2 and A3) and the two B genes (B1 and B2). The scheme, which is in line with the present and the previous phylogenetic analyses (see above), indicates that the more or less significant keratin character of the A and B chains (see above and Table 2) was probably the result of independent convergent evolution.

Discussion

The cephalochordate Branchiostoma (Amphioxus) is thought to be a key organism for understanding the origin and evolution of the chordates ([14] and references therein). Using phylogenetic tree analysis we previously showed that five of the 13 lancelet cytoplasmic IF proteins are orthologues of vertebrate keratins. Three type I (E1, k1 and Y1) and two type II (D1 and E2) keratin orthologues clustered with the corresponding human or vertebrate type I and type II keratins.[6,12,14] More importantly, the corresponding five recombinant proteins showed the obligatory heteropolymeric IF assembly based on a hetero coiled coil containing a keratin type I and a keratin type II polypeptide. Independent evidence for this assignment comes from the chimeric filament formation observed for two lancelet type I keratins with human keratin 8, a type II keratin. Just as with mammalian keratins, any lancelet type I protein formed IF when mixed with a lancelet type II protein.[6,12] Keratin expression in the lancelet occurs in a tissue-specific manner in epidermal and internal epithelia and, in contrast to vertebrates, in the nerve cord.[6,13]

Here, we identified a second obligatory heteropolymeric IF system of Branchiostoma. It is based on A3 as a representative of the three closely related A proteins (sequence identity about 90%) and the B2 protein. Recombinant A3 plus B2 proteins form long, obligatory heteropolymeric IF, while each individual protein yielded only aggregated material. Mutant proteins with a single

cysteine residue located to the same heptad *d* position in coil 2b formed IF, which oxidized readily. Gel electrophoresis under non-reducing conditions identified the disulfide-containing heterodimer A3/B2, and the two monomers A3 and B2. No crosslinked homodimers were observed. Since the purified A3/B2 heterodimer with the cystine crosslink formed readily IF the A3/B2 filaments are based on a double-stranded hetero coiled coil. The A3/B2 keratin-like IF differ from the true keratin filaments described above by a strict preference at the dimeric stage, which is seen in gel overlay assays. While normal keratins show a broad promiscuity, in that any type I binds to any type II protein (for mammalian keratins see Fuchs & Weber,[2] Parry & Steinert[3] and Hatzfeld & Franke[16]) the A3 probe detected only B2 and B1, and the B2 probe detected only A3 in a collection of ten different lancelet IF proteins (Figure 2; for a weak reactivity of A3 on B1, see below). This strong preference at the dimer level is indicated also by the tissue-specific expression patterns. B2 and A proteins are expressed in the intestinal epithelium (Figures 7 and 8), in which no normal keratin has been detected.

In addition to the heteropolymeric assembly pathways of true keratin orthologues and the newly defined keratin-like A/B2 molecules, there is a third IF assembly mechanism in Branchiostoma. IF protein B1, which shares 54 % identity in the rod domain with B2,[7] is the only lancelet protein able to form homopolymeric IF *in vitro*.[12] *In vivo*, it is the only IF protein restricted to mesodermally derived tissues[13] and thus can be functionally considered as a type III orthologue.

Interchain ionic interactions, which are known to be an important factor in specifying assembly partners in coiled coil molecules, have been calculated and shown to be consistent with the observed preference of A3/B2 heterodimers and B1 homodimers (Table 1). However, the calculations show that a small number of other chain combinations are also favoured. Since the role of the head and tail domains on IF assembly is not predictable at the residue level, and because not all of the possible ionic interactions are likely to be made *in vivo*, further discrimination between chain combinations is not possible at present.

Currently we do not know why the A/B branch of the evolutionary trees calculated by two laboratories and using both the distance and the maximum-parsimony methods (Figure 10(a); [6,12,14]) covers several functional and structural distinct proteins: the A group, which shows theoretical preference for type II chains, the functional type III protein B1 and the related protein B2, which both were theoretically designated to have a type I structure (Table 2). One possible explanation is given in Figure 10(b). This hypothetical view of Branchiostoma IF evolution proposes (a) that the more or less defined keratin-like character of the A and B chains resulted from evolutionary processes independent from those that drove the diversification of the two types of *bona fide* keratins in Amphioxus and (b) that the gene duplication event of the B-gene ancestor may have been followed by the rapid acquisition of functional differences within a remaining context of similar B1 and B2 sequences. Finally, we note, that the current (Figure 10(a)) and all previous phylogenetic trees (see above) indicated the lack of a sequence orthologue of vertebrate type III IF proteins within the currently known *Branchiostoma* IF proteins. Such proteins are, however, present in urochordates,[8,9] and can therefore be expected for cephalochordates.

Materials and Methods

Animal specimens

The collection of *B. floridae* adult and early larvae was as described.[13] Specimens were frozen in liquid nitrogen and kept at −80 °C.

Nucleic acid and protein techniques

Site-directed mutagenesis of the *B. floridae* cDNAs for proteins A3 and B2, cloned into the pET23a expression vector,[12] was performed using the QuiaChange mutagenesis kit (Stratagene).

Expression of the rod domains of A1 and A3 as GST fusion proteins were as described.[7] Expression, purification and *in vitro* filament assembly of recombinant A3 and B2 proteins were done as described.[6] The filament buffer was 10 mM Tris-HCl (pH 7.5).

Biotinylation of recombinant proteins was achieved with the EZ-Link™ Sulfo-NHS-LC-Biotinylation kit from Pierce (Rockford, IL) in 20 mM sodium borate (pH 8.5) containing 4 M urea and 1 mM 2-mercaptoethanol. The overlay experiments were performed on nitrocellulose membranes containing the ten immobilized, non-modified lancelet IF proteins A3, B1, B2, C2, D1, E1, E2, k1, Y1 and X1.[6] Membranes were blocked by incubation for one hour at room temperature in 5 % (w/v) skim milk in TBS (20 mM Tris-HCl (pH 7.4), 150 mM NaCl), 0.05 % (v/v) Tween 20 (TBST) and then incubated for two hours at room temperature with the biotinylated A3 or B2 probes in 20 mM Tris-HCl buffer (pH 7.5) containing 4 M urea. After washing in TBST, in TBST, 1 % (v/v) Triton, in TBST, 0.5 % (w/v) NaCl and again in TBST, each for five minutes, the membranes were treated with horse-radish peroxidase-conjugated streptavidin (Pierce) used at a concentration of 0.2 μg/ml. The blots were finally developed using the ECL chemiluminescence kit from Amersham according to the manufacturer's instructions.

Isolation of the polymerisation-competent heterodimer of *B. floridae* A3 and B2 containing a single disulfide bridge was essentially as described.[9] Briefly, equimolar amounts of both recombinant proteins were dialysed for three hours at room temperature against 2 mM Tris-HCl (pH 9), 1 mM 2-mercaptoethanol and then overnight against 10 mM Tris-HCl (pH 7.5) filament buffer in reducing agent. An aliquot was removed to monitor filament formation by electron microscopy. Filaments were then dialysed for 20 hours against the filament buffer without reducing agent 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 hours. An aliquot was removed to monitor filament formation by electron microscopy and to characterize the composition of oxidized filaments on SDS/10% (w/v) polyacrylamide gels run under non-reducing as well as reducing conditions. Filament were dissolved in 10 mM Tris-HCl (pH 8), containing 8.5 M urea without reducing agent, followed by Mono Q chromatography using the same buffer and a gradient formed with 0.4 M NaCl in the same buffer. An aliquot of the major peak eluted from the Mono Q column was subjected to one-dimensional (SDS/10% polyacrylamide) and two-dimenstional gel electrophoresis using the IPGphor system from Pharmacia (pH 3-10). Gels were stained with coomassie brilliant blue. Another aliquot was dialysed for four hours against 10 mM Tris-HCl (pH 7.5) without reducing agent to monitor filament assembly of the crosslinked heterodimer by electron microscopy.

Antibody preparations and immunofluorescence analyses

The A3 rabbit and B2 guinea pig antisera were obtained using the synthetic A3 peptide CVAEFQKKVDTLRAEA (amino acid residues 236-250 of the A3 protein in Figure 1) and B2 peptide CRLREIAENKVAIAEL (residues 288-302 of the B2 protein in Figure 1), respectively. Peptides were coupled *via* their extra N-terminal cysteine residue to hemocyanin. Both antisera were affinity-purified and, if necessary, preabsorbed on other recombinant proteins coupled to cyanogen bromide-activated Sepharose (Pharmacia). The specificity of each antibody was verified by immunoblotting on purified recombinant IF proteins as described.[6]

Frozen sections of *B. floridae* early larvae and adults were processed for indirect immunofluorescence microscopy as described.[6]

Sequence analyses

Potential interchain ionic interactions between charged residue pairs in positions 2e′-1 g, 1 g′-2e, 2a′-1 g, 1 g′-2a, 1e′-1d, and 1d′-1e were calculated as described.[21]

The following methodology was used to designate the chain type. Consensus sequences were derived for rod domain segments 1a, 1b, 2a, L2 and 2b in type I and type II mammalian keratins. Variability in lengths and character precluded a similar consensus being obtained for segments L1 and L12. For some positions in the sequences, considerable variability was observed, and these sites have been represented by a dot (Figure 9). In addition, where the same consensus residue was present in both the type I and the type II chains, both residues were designated X, and were thus not used further in attempts to assess the typeI/II character of Branchiostoma chains. The numbers of identities between the Branchiostoma chains and the type I and type II consensus sequences were calculated both for the entire sequence and for those residues falling in the non-*a* and non-*d* positions of the heptad repeat. Finally, the results were scaled to allow for the fact that the numbers of conserved residues unique to the type I and type II consensus sequences differed (144 and 157 respectively in all positions, and 105 and 120 respectively in non-*a* and non-*d* positions). The differences between the numbers of identities for the type I and type II consensus sequences is listed in Table 2. Positive and negative values indicate a preference respectively for type I and type II chains. Bigger values indicate a stronger preference for the particular chain type.

Only 259 amino acid long sequences of the rod domains (the linker and other regions of questionable homology were excluded) of all 13 Branchiostoma IF proteins and of 22 other vertebrate and invertebrate IF sequences were aligned and submitted to weighted (rescaled consistency index over an interval of 1-1000) maximum parsimony phylogenetic analysis.[22] Random additions (ten cycles) of the data were analysed with a heuristic method with a branch-swapping algorithm (tree bisection-reconnection, TBR). A bootstrap analysis (1000 replications) was used to infer a majority rule consensus phylogram that was rooted with the *Homo sapiens* lamin C as outgroup. Accession numbers of the Branchiostoma sequences used in this study are presented in Table 2; other sequences were reported elsewhere.[14]

Acknowledgements

We thank Wolfgang Berning-Koch, Uwe Plessmann and Susanne Brandfass for expert technical assistence.

References

1. McLean, W. H. & Lane, E. B. (1995). Intermediate filaments in disease. *Curr. Opin. Cell Biol.* **7**, 118-125.
2. Fuchs, E. & Weber, K. (1994). Intermediate filaments: structure, dynamics, function and disease. *Annu. Rev. Biochem.* **63**, 345-382.
3. Parry, D. A. D. & Steinert, P. M. (1995). *Intermediate Filament Structure*, Springer, New York.
4. Erber, A., Riemer, D., Hofemeister, H., Bovenschulte, M., Stick, R., Panopoulou, G. *et al.* (1999). Characterisation of the Hydra lamin and its gene; a molecular phylogeny of metazoan lamins. *J. Mol. Evol.* **49**, 260-271.
5. Stuurman, N., Heins, S. & Aebi, U. (1998). Nuclear lamins: their structure, assembly, and interactions. *J. Struct. Biol.* **122**, 42-66.
6. 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**, 1-10.
7. 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.
8. 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.
9. Wang, J., Karabinos, A., Schünemann, J., Riemer, D. & Weber, K. (2000). 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. *Eur. J. Cell. Biol.* **79**, 478-487.
10. Erber, A., Riemer, D., Bovenschulte, M. & Weber, K. (1998). Molecular phylogeny of metazoan intermediate filament proteins. *J. Mol. Evol.* **47**, 751-762.

11. 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.
12. 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 Letters,* **437**, 15-18.
13. Karabinos, A., Wang, J., Wenzel, D., Panopoulou, G., Lehrach, H. & Weber, K. (2001). Developmentally controlled expression pattern of intermediate filament proteins in the cephalochordate *Branchiostoma*. *Mech. Dev.* **101**, 283-288.
14. Luke, G. N. & Holland, P. W. (1999). Amphioxus type I keratin cDNA and the evolution of intermediate filament genes. *J. Exp. Zool.* **285**, 50-56.
15. Hesse, M., Franz, T., Tamai, Y., Taketo, M. M. & Magin, T. M. (2000). Targeted deletion of keratin 18 and 19 leads to trophoblast fragility and early embryonic lethality. *EMBO J.* **19**, 5060-5070.
16. Hatzfeld, M. & Franke, W. W. (1985). Pair formation and promiscuity of cytokeratins: formation in vitro of heterotypic complexes and intermediate-sized filaments by homologous and heterologous recombinations of purified polypeptides. *J. Cell. Biol.* **101**, 1826-1841.
17. Ruppert, E. E. (1997). Cephalochordata (Acrania). In *Microscopic Anatomy of Invertebrates*, vol. 15, pp. 349-504, Wiley-Liss, New York.
18. McLachlan, A. D. & Stewart, M. (1975). Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293-304.
19. Parry, D. A. D., Crewther, W. G., Fraser, R. D. B. & MacRae, T. P. (1977). Structure of α-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. *J. Mol. Biol.* **113**, 449-454.
20. Conway, J. F. & Parry, D. A. D. (1988). Intermediate filament structure. 3. Analysis of sequence homologies. *Int. J. Biol. Macromol.* **10**, 79-98.
21. Citi, S., D'Atri, F. & Parry, D. A. D. (2000). Human and *Xenopus* cingulin a modular organization of the coiled-coil rod domain: predictions for intra- and intermolecular assembly. *J. Struct. Biol.* **131**, 135-145.
22. Swofford, D. L. (1998). *PAUP*4.0-Phylogenetic Analyses using Parsimony (* and Other Methods)*, Sinauer Assoc., Sunderland, MA.

Edited by W. Baumeister

(Received 5 September 2001; received in revised form 13 November 2001; accepted 13 November 2001)