

Research article

Open Access

Molecular evolution of type VI intermediate filament proteins

Dominique Guérette¹, Paul A Khan², Pierre E Savard³ and Michel Vincent*¹

Address: ¹CREFSIP and Département de médecine, Pavillon Charles-Eugène-Marchand, Université Laval, Québec, G1K 7P4, Canada, ²Unité de recherche en pédiatrie, Centre de recherche du CHUL, Université Laval, Québec, G1V 4G2, Canada and ³Unité de recherche en Neurosciences, Centre de recherche du CHUL, Université Laval, Québec, G1V 4G2, Canada

Email: Dominique Guérette - doguerette@gmail.com; Paul A Khan - paul_allan47@yahoo.ca; Pierre E Savard - pierre.savard@crchul.ulaval.ca; Michel Vincent* - Michel.Vincent@rsvs.ulaval.ca

* Corresponding author

Published: 13 September 2007

Received: 28 May 2007

BMC Evolutionary Biology 2007, **7**:164 doi:10.1186/1471-2148-7-164

Accepted: 13 September 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/164>

© 2007 Guérette et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tanabin, transitin and nestin are type VI intermediate filament (IF) proteins that are developmentally regulated in frogs, birds and mammals, respectively. Tanabin is expressed in the growth cones of embryonic vertebrate neurons, whereas transitin and nestin are found in myogenic and neurogenic cells. Another type VI IF protein, synemin, is expressed in undifferentiated and mature muscle cells of birds and mammals. In addition to an IF-typical α -helical core domain, type VI IF proteins are characterized by a long C-terminal tail often containing distinct repeated motifs. The molecular evolution of type VI IF proteins remains poorly studied.

Results: To examine the evolutionary history of type VI IF proteins, sequence comparisons, BLAST searches, synteny studies and phylogenetic analyses were performed. This study provides new evidence that tanabin, transitin and nestin are indeed orthologous type VI IF proteins. It demonstrates that tanabin, transitin and nestin genes share intron positions and sequence identities, have a similar chromosomal context and display closely related positions in phylogenetic analyses. Despite this homology, fast evolution rates of their C-terminal extremity have caused the appearance of repeated motifs with distinct biological activities. In particular, our *in silico* and *in vitro* analyses of their tail domain have shown that (avian) transitin, but not (mammalian) nestin, contains a repeat domain displaying nucleotide hydrolysis activity.

Conclusion: These analyses of the evolutionary history of the IF proteins fit with a model in which type VI IFs form a branch distinct from NF proteins and are composed of two major proteins: synemin and nestin orthologs. Rapid evolution of the C-terminal extremity of nestin orthologs could be responsible for their divergent functions.

Background

The intermediate filament (IF) family is composed of more than 70 genes that are expressed in a tissue- and developmental stage- specific manner in metazoan cells [1-3]. All IF proteins exhibit a tripartite structure comprising a central α -helical core domain flanked by globular head and tail regions [4]. Members of the IF family are

grouped together in a class of nuclear proteins (lamins: type V) along with four or five classes of cytoplasmic proteins (types I-IV, VI) depending on the criteria used for their classification [3-6]. Keratins represent the first two classes (types I and II) of IF proteins and they are obligatory heteropolymers. Keratin genes are the most abundant IF family members. In humans, they are clustered on chro-

mosomes 17q21 (type I) and 12q13 (type II) [7]. Vimentin, desmin, peripherin and GFAP form type III IF proteins that can assemble in filaments on their own, or in combination with type IV and type VI IF proteins. Neuronal IF proteins comprise NF-L (light), NF-M (medium), NF-H (heavy) neurofilament protein subunits that along with α -internexin are classified as type IV IF proteins.

Upon its identification in 1990, nestin was designated as the prototype of a new IF protein group (type VI) because it did not fall clearly into any of the previously described types [6]. Some debate arose on this classification since nestin gene structure is closely related to the neurofilament branch in having two of its three intron positions in common with NF genes [8]. Accordingly, it had been proposed to re-classify nestin as a type IV IF protein [9]. However, the low level of sequence similarity of the α -helical region of nestin and NF proteins as well as the presence of a third intron in the nestin gene constitute strong arguments in favor of its classification as a distinct type [6,8,10]. Furthermore, the discovery of tanabin in *Xenopus laevis* a few years later led to the proposal that this tanabin protein could be the prototype of a different IF type (type VII) because of the lack of significant sequence similarities with other IF proteins [11]. Shortly after nestin and tanabin were sequenced, the gene structures of synemin [12,13] and transitin [14] were also described. According to their sequence similarities, tanabin was then grouped with transitin, paranemin (a splice variant of transitin), synemin and nestin as type VI IF proteins. All these proteins are distinguished by a long C-terminal extremity and by the fact that they cannot self-form into filaments. Rather, they need other IF proteins to build filamentous structures [10,15].

Tanabin is specific to amphibians, transitin to birds and nestin to mammals. Tanabin is expressed during neurulation of *X. laevis* and its function is not well understood [11]. Transitin and nestin are transiently expressed in myogenic and neurogenic cells of birds and mammals, respectively [6,16-20]. Chicken transitin is co-expressed with vimentin in proliferating myoblasts and is associated for a short period of time with desmin at the Z line during muscle differentiation [17]. Transitin expression persists in the smooth muscle cells of elastic arteries and in Purkinje fibers where it is expressed in association with vimentin [21]. Its expression is also induced in activated Müller glial cells of chicken retinas following acute retinal damage [22]. Paranemin, a splice variant of transitin [14], is important for the formation of an extended IF network when co-transfected with desmin in SW13 cells [23]. Recently, transitin has been shown to play an important role in determining the intracellular localization of Numb in mitotic neuroepithelial cells [24]. In mammals, nestin expression is induced in certain tumors as well as in regen-

erating skeletal muscles [25,26]. In addition, nestin is implicated in vimentin intermediate filament disassembly during mitosis [27] and is a survival determinant through cdk5 regulation in oxidant-induced cell death [28]. All these observations suggest that both transitin and nestin have important and distinct functions in various cell types during embryonic development and in tissue regeneration in adults.

Despite a low level of sequence identity, the large tail domains of nestin and transitin display some similarities including the presence of highly charged glutamate-rich stretches and of repeated motifs prone to α -helicity. The tail domain of transitin contains a motif comprised of more than 50 leucine zipper-like heptad repeats (HR domain) of the consensus sequence LQVEHGD [29] whereas that of nestin features an 11-amino acid repeat motif whose number of repetitions varies according to the species [6]. Synemin is another type VI IF protein expressed in developing and adult skeletal muscles of both birds and mammals [30-32]. Different studies report that interaction of the long C-terminal tail of synemin with other cytoskeletal components may be a key component linking myofibrillar Z lines to costameres in skeletal muscle cells [33,34].

The molecular evolution of type VI IF proteins remains poorly studied. As already mentioned, the common denominator shared by all IF proteins is the presence of an α -helical region involved in filament assembly. Two prototypes of cytoplasmic IF proteins, defined by the presence or absence of a long lamin-like coil 1b within the α -helical domain, seem to parallel metazoan phylogeny. The first prototype has the long coil 1b subdomain and often a lamin homology segment in its tail domain. It has been documented for 12 protostomic phyla [35,36] and an hemichordate, although the "long" coil 1b is shortened by 11 residues in the latter [36]. The second prototype, restricted to the chordates, contains a coil 1b shortened by 42 residues and lack a lamin homology segment. Following a 42 residue deletion that occurred at the origin of the chordate branch, type I-III- IF proteins were established by duplication events and sequence drift. The genes encoding type IV NF proteins have different intron positions than do type I-III genes. They were proposed to be derived from retrotransposition of an intron-less intermediate followed by the acquisition of new introns [37] but this hypothesis has been recently challenged by the documentation of a fish gene combining type I-III intron positions with type IV intron positions [38].

To analyze the evolution of genes encoding type VI IF proteins, sequence comparison, BLAST searching, synteny studies and phylogenetic analysis were performed with members of this group. This study provides new evidence

that tanabin, transitin and nestin are indeed orthologous type VI IF proteins. These proteins possess significant diversity in composition of their long C-terminal tails that likely provides them with different, but specific functions in myogenic and neurogenic cells of developing vertebrate systems. In particular, *in silico* and *in vitro* analyses provide evidence that the C-terminal extremity of (avian) transitin, but not that of (mammalian) nestin, contains a repeat domain displaying nucleotide hydrolysis activity.

Results and discussion

1-Overview of type VI IF proteins

The cDNA sequence of tanabin from *X. laevis* has been published [11], but the organization of its gene structure is not known. In order to analyze the evolution of type VI IFs, we first determined the exon/intron structure of the tanabin gene of *X. tropicalis* using JGI portal v.4.1. As tanabin was first described in *X. laevis* [11], a BLASTp search was made to determine the putative ortholog of tanabin in *X. tropicalis*. The protein sequence of tanabin from *X. laevis* (tanabin-xl) was used as the query sequence in BLASTp searching carried out at the JGI portal against the *X. tropicalis* genome assembly v4.1. A sequence named fgenes1_pg.C of 1868 bp was found to be 67% identical to tanabin-xl. Tanabin from *X. tropicalis* (tanabin-xt) is 1970 amino acids (aa) long and possesses the typical α -helical rod domain of IF proteins and a long C-terminal tail of more than 1400 aa. Tanabin-xl and tanabin-xt share more than 90% sequence similarity in their IF rod domain (data not shown).

The mRNA sequence encoding for tanabin-xt was then compared with the genomic sequence of *X. tropicalis* (assembly v4.1.) to determine intron boundaries (Fig. 1a). The genomic sequence of tanabin-xt revealed the presence of 5 exons of 804 bp, 125 bp, 65 bp, 4848 bp and 71 bp as well as 4 introns of 4610 bp, 3715 bp, 499 bp and 6593 bp, respectively. The first three exons encode for the IF core domain and the last two for the C-terminal domain. Comparison of the gene structure of tanabin with that of the other type VI family members (nestin, transitin, synemin) is illustrated in Figure 1b. All genes share 3 identical intron positions within the 5' portion of the gene and, as already observed [8,39], the positions of the two most upstream introns are also shared by the NF genes. The IF rod domain of type VI IF proteins is thus encoded by three exons. Our analysis also shows that tanabin has a supplementary intron located at the 3'end of the gene (Fig. 1A).

The protein sequences of type VI IF proteins were compared using BL2SEQ on the NCBI web site to verify the level of similarity in their IF rod domain and their C-terminal tail (Table 1). Tanabin-xt and transitin share 51% sequence identity in their rod domain, which is the most

Table 1: Protein sequence identity of the alpha helical rod domain and the C-terminal tail of type VI Ifs. Protein sequences deduced from available sequenced genomes present in NCBI and JGI genome assembly v4.1. were analyzed using BL2SEQ searching on the NCBI server. Sequence identities are presented on the table and --- means that no sequence identity was found in these regions of the proteins. Chicken transitin: X80877, human nestin: NM_006617, tanabin-xt: 186291 and human synemin: CAC83859

	Sequence identity(%) Rod domain	Sequence identity(%)C-terminal
Transitin vs Tanabin-xt	51	19
Transitin vs Nestin	44	24
Tanabin-xt vs Nestin	26	19
Transitin vs Synemin	31	---
Tanabin-xt vs Synemin	24	---
Nestin vs Synemin	30	---

important level found among type VI IF proteins. Transitin exhibits 44% identity with human nestin in this domain, a little lower than the usual 50% observed in IF proteins of the same group. The overall level of sequence identity in the carboxy-terminal domain of the proteins is lower at around 20%.

A well conserved sequence among IF chains is the helix termination motif at the end of segment 2B of the α -heli-

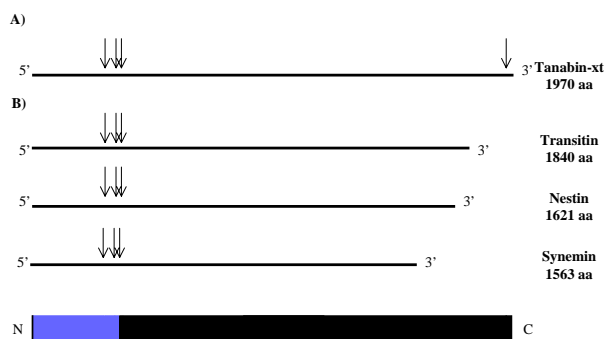


Figure 1
Schematic representation of the position of introns in type VI intermediate filament genes. A) The arrows along the tanabin cDNA diagram indicate the predicted positions of introns in the tanabin gene deduced from the genomic sequence of *X. tropicalis* (v4.1). **B)** The arrows indicate the location of introns in nestin [6], transitin [14] and synemin [13]. An IF protein diagram is illustrated at the bottom of the figure for orientation purposes. The blue box represents the α -helical rod domain.

cal rod domain. The consensus sequence for this motif is: E-Y-Q-X-L-L-D/N-V-K-X-R/A-L-D/E-X-E-I-A-T-Y-R-K/R-L-L-E-G-E-E/D-X-R-L/N/I [4]. Multiple alignments using type VI IF proteins as well as desmin and NF-M sequences show that the helix termination motifs of type VI IF chains diverge from the consensus sequence at specific sites: D/G/E-R/D/G/Y-Q-X-L-A/M/L-H/Q-L/V-K-X-S/G-L-S-X-E-V-A-T-Y-R-T/S/A-L-L-E-A/G-E-X-R-L/I/Q/E (Fig. 2A). The helix termination motif contains key residues that are important for structure and assembly of IFs and this region represents one of the two mutation hotspots in IF proteins [40]. The observed divergences suggest that type VI IF proteins evolved in a branch distinct from NF proteins and that they gained new residues at the same sites of segment 2B: Q95, H98, S102, S104 and A116 (asterisks in Fig. 2A). Those residues may be implicated in specific functions of type VI IF proteins.

To investigate potential relationships among type VI IF proteins, a phylogenetic tree was constructed using a multiple alignment of the entire sequence of known type VI proteins. The sequence of NF-M proteins was also used to examine how closely related type VI IF proteins are to NF proteins. As seen in Fig. 2B, such phylogenetic analysis shows that synemin, transitin, tanabin and nestin are all part of a group that is distinct from the NF-M protein. This correlates with the idea that type VI IF genes are the evolutionary result of a duplication event of an ancestral NF gene and that type VI IF genes then evolved independently from NFs. Transitin and tanabin are probably more closely related to each other than to other type VI IF proteins as suggested by their close phylogenetic proximity. In summary, our analysis indicates that tanabin, transitin and nestin form a branch distinct from NF proteins and they are more closely related to each other than to synemin.

2-Tanabin, transitin and nestin are orthologous proteins

To investigate whether tanabin, transitin and nestin are orthologous proteins, a synteny analysis was performed using the NCBI mapped genomic scaffolds of human, mouse, rat and chicken along with the JGI genomic scaffolds of *X. tropicalis*. As seen in figure 3, regions of strong synteny conservation exist around both the nestin gene in all three mammalian species and the transitin gene in chicken. Nestin and transitin genes both locate between BCAN and PRCC genes and the order of the neighboring genes is conserved. Tanabin was located close to the BCAN gene on *X. tropicalis* Scaffold_790. The PRCC, SH2D2A and ARHGEF11 genes were grouped on a different scaffold (not shown). This indicates that the genomic location of amphibian tanabin, avian transitin and mammalian nestin has been conserved. These observations, combined with the fact that these genes have the same

Table 2: ATPase score motif associated with the C-terminal tail of tanabin, transitin and nestin of different species by Conserved domain database analysis on the NCBI server. An ATPase motif was detected in tanabin, transitin and nestin in comparison with the domain architectures available in the Conserved domain database on the NCBI server. The score shown in the table is calculated using PSSM scoring matrix based on protein alignment. The expected value is also given for each alignment and significant E-values for this database searching are considered to be > 10⁻⁵. Protein sequences for this analysis were obtained from sequenced genomes available on NCBI and JGI genome assembly v4.1. Tanabin-xt: 186291, chicken transitin: X80877, dog predicted-nestin: XP_547531.2, mouse nestin: NP_057910.3, rat nestin: NP_037119.1 and human nestin: NM_006617

Query sequence	E-value (IF domain)	E-value (ATPase domain)
Tanabin (<i>Xenopus laevis</i>) gi 549051	2e-15	0.0003 ¹
Transitin (<i>Gallus gallus</i>) gi 45384298	5e-14	5e-15 ²
Nestin ³ (<i>Canis familiaris</i>) gi 73961547	6e-8	2e-8 ¹
Nestin ³ (<i>Monodelphis domestica</i>) gi 126307847	3e-19	0.000009 ¹
Nestin ³ (<i>Bos taurus</i>) gi 76612380	9e-13	No hits
Nestin (<i>Rattus norvegicus</i>) gi 6981262	7e-11	0.008 ¹
Nestin (<i>Mus musculus</i>) gi 50363232	0.000002	0.003 ²
Nestin ³ (<i>Macaca mulatta</i>) gi 109017378	9e-13	No hits
Nestin (<i>Homo sapiens</i>) gi 1346682	9e-7	0.00004 ²

¹ NCBI imported model COG5271, MDN1, AAA ATPase containing von Willebrand factor type A domain

² NCBI imported model COG1196, Smc, Chromosome segregation ATPases

³ Predicted sequence

intron distribution strongly suggest that tanabin, transitin and nestin originate from the same ancestral gene.

3-ATPase and GTPase activity of the transitin HR domain

We examined the domain architecture of the type VI IF proteins through the CDD database (Table 2). As expected, this analysis confirmed the presence of a canonical IF domain in the N-terminal moiety of tanabin, transitin and nestin but a new domain was revealed in their C-terminal extremity, corresponding to an ATPase motif of either the SMC type (S_{tr}uctural M_{ai}ntenance of C_hromosomes superfamily of proteins) or the AAA+ family (A_TPases A_ssociated with various cellular A_ctivities). The highest score was obtained for the transitin tail and corresponded to its HR domain. Lower scores were obtained for tanabin and nestin C-terminal extremities from different species (Table 2). The exact function of the transitin HR domain is unknown. Paranemin, a splice variant of transitin, has been shown to be involved in the formation

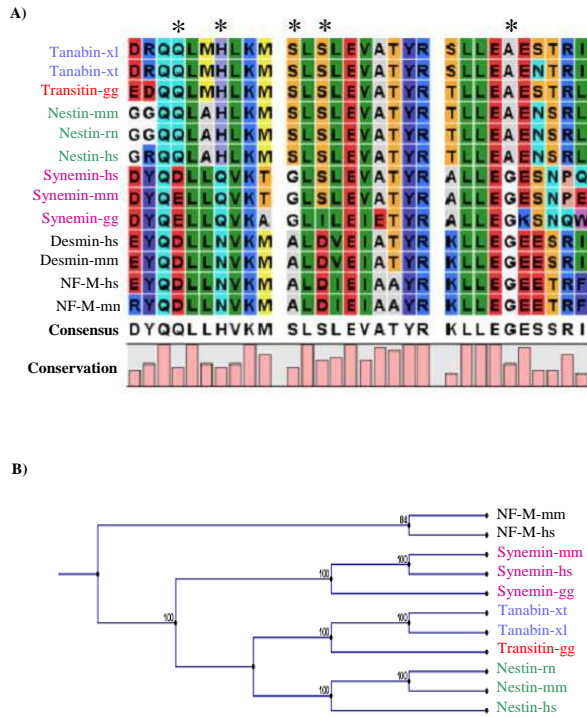


Figure 2
Evolution of type VI IF proteins. A) Amino acid sequence alignment of the helix termination motif at the end of segment 2B of the α -helical rod domain of type VI IF proteins, desmin and NF-M from different vertebrate species. Asterisks represent amino acid residues specifically conserved in tanabin, transitin and nestin. The consensus sequence deduced from this multiple alignment is indicated at the bottom of the figure. The histogram indicates the level of residue conservation (from 0 to 100%) at each position. **B)** A phylogenetic tree was constructed based on the protein sequences of type IV NF-M and of type VI tanabin, transitin, nestin and synemin from different species using neighbor-joining methods (CLCbio software; 100 bootstraps). A closer evolutionary relationship is noticed between tanabin, transitin and nestin. The bootstrap values are shown at the nodes. xl, *Xenopus laevis*; xt, *Xenopus tropicalis*; gg, *Gallus gallus*; rn, *Rattus norvegicus*; mm, *Mus musculus*; hs, *Homo sapiens*.

of an extended IF network when co-transfected with desmin in IF-free cells [23]. The HR domain may be implicated directly in this mechanism since this protein motif has been shown to interact with type III IF proteins vimentin and desmin during myogenesis (Gu erette *et al.*, submitted). The low sequence identity between the repeated motifs of transitin and nestin carboxy-terminal extremities and the lowest score for an ATPase motif in the nestin tail may be the consequence of the rapid evolution of this part of the gene. As previously shown, the carboxy-termi-

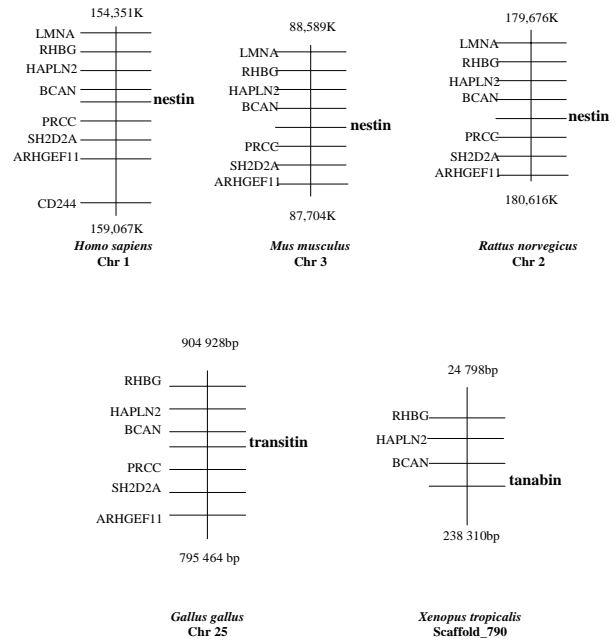


Figure 3
Syntenic relationship among frog tanabin, avian transitin and nestin from human, mouse and rat. Physical maps of human, mouse, rat and chicken genomic sequences along with an anuran scaffold (*X. tropicalis* scaffold_790) were used to identify genes neighboring nestin, transitin and tanabin. Each diagram represents a chromosomal region whose position is defined by the number of base pairs from the telomere of the short arm of the chromosome (top and bottom of each diagram). Each horizontal line represents a gene position on the chromosome. Nestin, transitin and tanabin genes are shown in bold for clarity. Diagrams are arranged in such a way that the gene positions are visualized in the same orientation. LMNA, lamin A/C; RHGB, Rhesus blood group, B glycoprotein; HAPLN2, hyaluronan and proteoglycan linked protein 2; BCAN, brevican; PRCC, papillary renal cell carcinoma; SH2D2A, SH2 domain protein 2A; ARHGGEF11, Rho guanine nucleotide exchange factor (GEF) 11 and CD244, CD244 natural killer cell receptor 2B4.

nal region of nestin contains a repeated domain subjected to size fluctuations among rodents and human that could be linked to its higher evolutionary rates, compared to the IF rod domain [8]. Incidentally, better E-values were obtained for an ATPase motif using sequences from dog, dolphin and human rather than from rodents (Table 2).

We verified by *in vitro* means whether the C-terminal repeated domains of both chicken transitin and mouse nestin could have either ATPase or GTPase activity. To do so, fusion proteins representing either the entire HR domain of transitin (HR) or the entire C-terminal

repeated domain of nestin (CTR) were expressed in the same bacterial strain. These fusion proteins were purified and ATP and GTP hydrolysis activities were assayed for phosphate release using the malachite green colorimetric method for phosphate determination (Fig. 4) [41]. The HR domain of transitin has ATPase (Fig. 4a) as well as GTPase (Fig. 4b) activity as shown by an OD augmentation in our assay. Using such an assay, it has not been possible to detect any ATPase or GTPase activity in the mouse nestin CTR. These experimental results are concordant with our *in silico* analysis suggesting that transitin ATPase activity could have been lost during the rapid evolution of nestin C-terminal repeat domain in mammals.

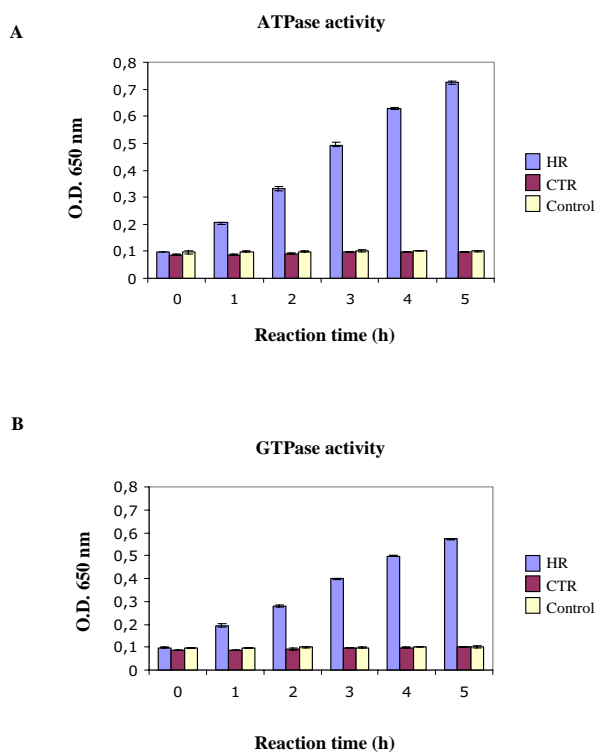


Figure 4
In vitro ATPase and GTPase activity of the HR domain. The HR domain of chicken transitin and the CTR domain of mouse nestin were expressed in bacteria and purified. Their ATPase/GTPase activities were tested individually using the malachite green colorimetric method for phosphate release determination. Phosphate release was determined at 25°C for each recombinant protein incubated with either **A**) ATP or **B**) GTP. The optical density at 650 nm was determined in triplicate at regularly timed intervals during the course of the enzymatic reaction and the mean values are presented with standard deviation error bars. Control: Buffer with the HR recombinant protein without NTP.

The SMC and AAA+ proteins contain two nucleotide-binding modules, the Walker A and Walker B motifs [42], defining a broad superfamily of nucleotide-binding proteins including many ATPases, myosin and numerous kinases [43]. In these proteins, the ATP-binding module is activated by the formation of an oligomeric assembly and drives conformational changes affecting target substrates. The Walker A and B motifs of SMC proteins, which show the highest score with transitin HR domain, are located in the N- and C-terminal extremities of these proteins and are separated by a central domain composed of a hinge sequence flanked by two long coiled-coil motifs [44]. As the bulk of transitin HR domain is predicted to have a coiled-coil structure [45], it may be anticipated that the HR domain presents ATPase activity at one or both ends. The Walker B signature motif, as described by Walker [42], corresponds to R/KX₃GX₃Lh₄D (h = hydrophobic and X = any residue) and could loosely match with the sequence RDLQEGHGDLQVEHED located at the N-terminal extremity of the HR domain. In fact, hydrolytic activity has been detected in our ATPase assay using fragment HR_{1-4'}, which contains the first 4 repeats of the HR domain. In addition, this domain has been shown to completely disassemble the IF network when overexpressed in avian myoblasts (Guérette *et al.*, submitted). Site-directed mutagenesis is now under way to identify the most important residues for ATP binding and hydrolysis.

4-Conservation of the HR domain in some species

Since the HR domain is a feature unique to transitin in that it possesses both ATPase and GTPase activities *in vitro*, we focused on the evolution of this particular domain. Assuming that transitin and nestin are orthologous proteins and that nestin CTR does not show significant sequence identity with transitin HR domain, we looked for the presence of HR domain-like sequences in other genomic regions. BLASTn searches demonstrated that the transitin HR domain has sequence similarities to two genomic clones from human and mouse origins. The human clone RP1 155d22 (gi: 2827470) is located on chromosome 6q27 and is 82% identical over 118 nt to the HR domain of chicken transitin (E-value = 4e-7). This clone does not correspond to the human nestin gene, located on chromosome 1. The mouse genomic clone RP2389A3 (gi: 106520665) is located on chromosome 17 and is 79% identical over 353 nt to the HR domain of chicken transitin (E-value of 5e-16). An hypothetical protein is also predicted (gi: 94403328) with an E-value of 4e-13. This protein is encoded in part by an RP2389A3 clone that likely corresponds to the mouse version of the HR domain. The human and mouse versions of the HR domain are not part of any currently known gene. Moreover, these regions are devoid of any IF core feature in the 5' regions suggesting that the putative HR domain found in human (HRH) and mouse (HRM) may be part of pro-

teins that are not members of the IF family. The HR domain of chicken transitin was used as a query for a BL2SEQ alignment with the HRH and HRM sequences. Both have 50% protein sequence identity with the chicken transitin HR domain. Consensus repeated sequences found in HRH and HRM are LQVEEGS and MQVEHGD respectively, compared with the consensus sequence of LQVEHGD in the chicken transitin HR domain. Furthermore, monoclonal antibody VAP-5 directed against a repeated epitope of the HR domain of chick transitin [45] targets a similar epitope on a synthetic HRM protein (Fig. 5). The HRM sequence has been subcloned in a pET-based expression vector and a His-tag-HRM fusion protein bacterially produced under IPTG induction. In induced bacterial cultures, an immunoreactive band of the expected Mr value was observed in Western blots using the anti-His-tag antibody and VAP-5. This protein was not detected in uninduced cultures (Fig. 5: without IPTG).

A synteny search was conducted using the physical maps of human, mouse and chicken genomes to determine whether the genes encoding the HR domain in human and mouse were conserved among the same set of neighboring genes. The "mammalian" HR domain genomic segments were syntenic in human (chromosome 6q27) and mouse (chromosome 17) (Fig. 6a) and correspond to a well-conserved gene cluster on chicken chromosome 3 from which the "avian" HR domain genomic segment was absent (Fig. 6b) as it is located on chromosome 25 as part of the transitin gene (Fig. 6a). These observations suggest that the nucleotide sequence corresponding to the conserved HR domain in human and mouse likely originated

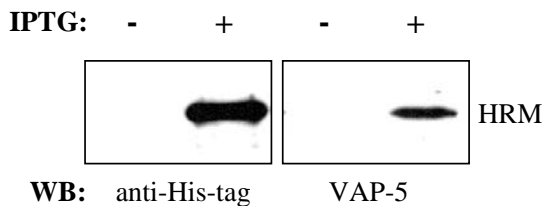


Figure 5
Reactivity of the HR-specific monoclonal antibody VAP-5 with the "hypothetical" mouse protein HRM.
 A fragment of mouse genomic clone RP2389A3 (gi: 106520665) showing the highest sequence identity to chick transitin HR domain has been subcloned in a pET-based expression vector and a His-tag-HRM fusion protein bacterially produced under IPTG induction. A co-migrating immunoreactive band was observed in Western blots using the anti-His-tag antibody and VAP-5 in extracts of bacteria submitted to IPTG induction but was not detected in the uninduced culture.

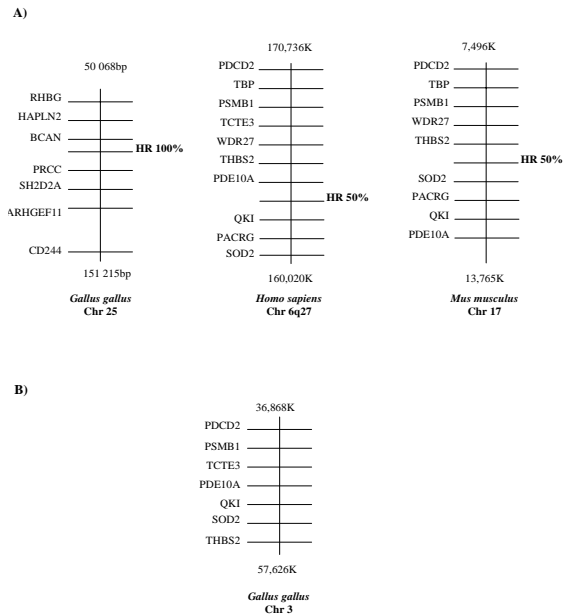


Figure 6
Syntenic relationships of the conserved HR domain in birds, humans and mice. Physical maps of human, mouse, rat and chicken genomic sequences were used to identify genes neighboring the conserved HR domains. Each diagram represents a chromosomal region whose position is defined by the number of base pairs from the telomere of the short arm of the chromosome (top and bottom of each diagram). Each horizontal line represents the position of a gene and is placed in physical order on the chromosome. Diagrams are depicted in such a way that genes are oriented in the same way. **A)** The HR domains were syntenic in human (chromosome 6q27) and mouse (chromosome 17). They share 50% identity with the chicken HR domain of transitin located on chromosome 25 (see Fig. 3). **B)** A corresponding gene cluster was found on chicken chromosome 3 from which an "avian" HR domain genomic segment was absent. RHGB, Rhesus blood group, B glycoprotein; HAPLN2, hyaluronan and proteoglycan link protein 2; BCAN, brevican; PRCC, papillary renal cell carcinoma; SH2D2A, SH2 domain protein 2A; ARHGEF11, Rho guanine nucleotide exchange factor (GEF) 11 and CD244, CD244 natural killer cell receptor 2B4; PDCD2, programmed cell death 2; PSMB1, proteasome (prosome, macropain) subunit, beta type, 1; TCTE3, t-complex-associated-testis-expressed 3; PDE10A, phosphodiesterase 10A; QKI, quaking homolog; SOD2, superoxide dismutase 2; THBS2, thrombospondin 2, PACRG, PARK2 co-regulated and WDR27, WD repeat domain 27.

from the exon coding for the HR domain in a chicken transitin ancestral gene.

To establish whether sequences similar to the HR domain of chicken transitin could be found in species located in

upstream branches of vertebrate evolution, a tBLASTn search of *Takifugu rubripes* (pufferfish) was made using the entire transinin protein sequence. Genome sequence analysis of *T. rubripes* did not reveal the presence of a nestin homolog [38]. This may explain why the rod domain of desmin was found by tBLASTn analysis to be the strongest hit to the rod domain of transinin (27% identity). On the other hand, the HR domain of transinin is 22% identical to a retinitis pigmentosa GTPase regulator-like (RPGR-like) protein spanning 555 nt with an E-value of $2e-28$. As the protein sequence of tanabin is closely related to transinin, tanabin was used to conduct a second tBLASTn search. Once again, the RPGR-like protein emerged (22% identical) spanning 580 nt with an E-value of $2e-21$ to the C-terminal of tanabin. The observations suggest that tanabin and transinin tail domains have evolved from an RPGR-like protein.

5-Model for the evolution of type VI IF proteins

Type VI IF proteins have two of the three intron positions in common with type IV NF genes but the level of similarity in the α -helical regions is only 20% compared with 50% observed among the NF genes. It has been postulated that a type VI IF gene ancestor branched off before the split into the three NF genes where the ancestor later gained a third intron [8]. We propose a model to explain the evolution of type VI IF proteins (Fig. 7). This model suggests that the first type VI IF gene arose evolutionarily as the result of incorporation of an RPGR-like cassette into the 3' extremity of an ancestral NF gene. From tanabin in amphibians, the history of type VI IF genes may have included loss of the supplementary intron in the 3' part of the gene followed by fast evolution of the C-terminal RPGR-like cassette giving rise to the HR domain as it is found in avian transinin. In mammals, this domain must have evolved quickly as some similarity is found between the chicken HR domain and the dog CTR domain but none exists between the chicken HR domain and the CTR domain of mouse and human nestin. According to this model, the tail domain of human nestin would not have resulted from the loss of the RPGR-like cassette and its substitution by the CTR domain but rather from a fast evolution rate leading to loss of its nucleotidase activity. In addition, the HR domain could have been duplicated since it is found in a non-IF hypothetical gene which has been conserved in a syntenic way among mice and humans.

Conclusion

Many lines of evidence based on sequence identity, gene structure, synteny comparison and phylogeny searching point to the conclusion that frog tanabin, chicken transinin and mammalian nestin are orthologous members of the type VI IF proteins. The C-terminal domains of both tanabin and transinin were predicted to have nucleotide

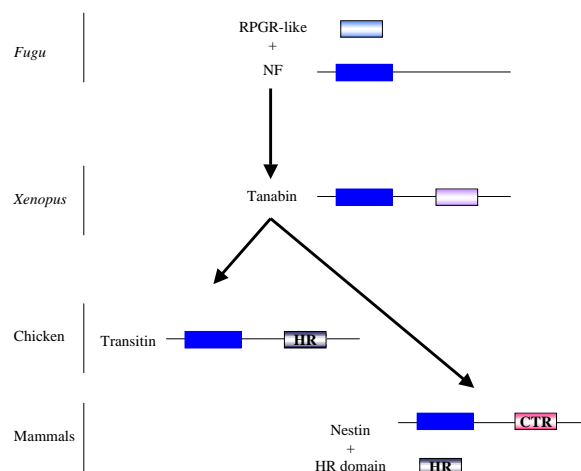


Figure 7

Evolutionary model for type VI IF proteins. *Fugu* lacks obvious nestin homologs [38] and NF proteins may well represent ancestral type VI IF proteins [8]. Our model proposes that the first type VI IF protein resulted from the fusion of a RPGR-like cassette C-terminal to a NF gene. The C-terminal domain of a tanabin-like protein may have evolved giving rise to the HR domain displaying nucleotide hydrolysis activity in birds whereas this activity was progressively lost in mammalian nestins. In addition, the HR domain may have been duplicated and conserved in a mammalian gene that is not part of the IF family.

hydrolysis activity *in silico*, and indeed, ATPase activity was measured *in vitro* within the HR domain of transinin. This domain apparently experienced a fast evolution rate that could have resulted in loss of ATPase activity in mammals.

Methods

Known type VI protein sequences

Type VI IF protein sequences used in this study were: chicken transinin [GenBank: [X80877](#)], human nestin [GenBank: [NM_006617](#)], tanabin-xl [GenBank: [M99387](#)], tanabin-xt [JGI: 186291], human synemin [GenBank: [CAC83859](#)]. Other protein sequences used in this study were: human NF-M [GenBank: [CAA68276](#)], mouse NF-M [GenBank: [CAA29127](#)], human desmin [GenBank: [NP_001918](#)], mouse desmin [GenBank: [NP_034173](#)], and RPGR-like [GenBank: [AAG00554](#)].

BLAST, multiple sequence alignment and phylogenetic tree analysis

BLASTn, BLASTp, tBLASTn and BL2SEQ searches were performed in NCBI or JGI (in the case of *X. tropicalis*) databases. Default parameters were used to conduct the

searches. Multiple sequence alignments were prepared using CLC workbench version 3.0.1. These multiple sequence alignments were used to create a phylogenetic tree with neighbor-joining methods with 100 bootstrap analysis using CLC workbench version 3.0.1 software.

Syntenic relationship identification

Chromosomal locations of nestin and transitin genes were identified using physically-mapped human (build 36.1), mouse (build 35.1), rat (build v3.4) and chicken (build 1.1) genomes. For each species, syntenic genes were located using mapviewer (NCBI). In the case of transitin, as this gene was not located on any chicken chromosome (build 1.1) at the onset of our study, the contig [NW_094723.1](#) containing the transitin gene was used to identify neighboring genes. These genes were later located to chicken chromosome 25 (build 2.1) and then compared with physically mapped human, mouse and rat genomes. To analyze synteny relationships of the transitin HR domain, the contigs containing the human and mouse HR sequences were located on human and mouse chromosomes and the genes surrounding these sequences were positioned and compared with the chicken genome.

Assay of ATPase and GTPase activity

Different fusion proteins were cloned as described previously [45]. Fusion proteins were purified by T7-tag affinity purification (Novagen) or by His-tag Sepharose (Amersham) according to the instructions of the manufacturers. ATPase and GTPase activities were determined using "malachite green phosphate assay kits" from BioAssay Systems. After purification, fusion proteins were dialyzed against Buffer A (50 mM Tris-HCL pH 7.5, 10 mM MgCl₂, 100 mM NaCl, 20 mM KCl and 1 mM β-mercaptoethanol) overnight at 4 °C and concentrated using Centricon YM-10 (Millipore). Protein concentrations were determined with a Micro BCA Protein Assay kit (Pierce). Fusion proteins (0.04 μg/μl) were incubated in Buffer A with 1 mM ATP or 1 mM GTP at 25 °C. Aliquots were taken at different times and mixed with malachite green buffers as described by the manufacturer (BioAssay Systems). After 15 min of incubation, the O.D. at 650 nm was determined using a Multiskan Spectrum spectrophotometer (Thermo-LabSystems). A standard curve with free phosphate was produced according to the instructions of the manufacturer.

SDS-PAGE and Western blots

A fragment of mouse genomic clone RP2389A3 (gi: 106520665) showing the highest sequence identity to chick transitin HR domain has been amplified by PCR and subcloned in a His-tag containing pET30 expression vector to transform BL21(DE3)pLysS bacteria. The expression of the His-tag-HRM fusion protein was induced by 0.1 mM IPTG and bacterial pellets directly solubilized in

electrophoresis sample buffer. The protein samples were resolved by SDS-PAGE and transferred to nitrocellulose membranes as described [46]. For Western blots, the membranes were saturated for 1 hour at room temperature using 1% blocking reagent (Roche Diagnostics). The primary antibody (anti His-tag or mAb VAP-5) was incubated for 1 hour and the secondary antibody for 45 minutes, both at room temperature. The proteins were detected using the BM chemiluminescence kit (Roche Diagnostics).

Authors' contributions

Bioinformatic and experimental analysis were performed by DG. All authors contributed to designing experiments, analyzing data and writing the manuscript. They all accepted the final version.

Acknowledgements

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR; FRN 72199). D.G. is recipient of a CIHR studentship as part of a Strategic Training Program Grant in genomics (STP-53894) We are grateful to Ms Karine Blais and Mr Julien Trépanier for technical assistance and to Dr Sébastien Michaud (Laval University) for critical reading of the manuscript.

References

- Hesse M, Magin TM, Weber K: **Genes for intermediate filament proteins and the draft sequence of the human genome: novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18.** *J Cell Sci* 2001, **114**(Pt 14):2569-2575.
- Lazarides E: **Intermediate filaments: a chemically heterogeneous, developmentally regulated class of proteins.** *Annu Rev Biochem* 1982, **51**:219-250.
- Coulombe PA, Wong P: **Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds.** *Nat Cell Biol* 2004, **6**(8):699-706.
- Parry DA: **Microdissection of the sequence and structure of intermediate filament chains.** *Adv Protein Chem* 2005, **70**:113-142.
- Toivola DM, Tao GZ, Habtezion A, Liao J, Omary MB: **Cellular integrity plus: organelle-related and protein-targeting functions of intermediate filaments.** *Trends Cell Biol* 2005, **15**(11):608-617.
- Lendahl U, Zimmerman LB, McKay RD: **CNS stem cells express a new class of intermediate filament protein.** *Cell* 1990, **60**(4):585-595.
- Waseem A, Gough AC, Spurr NK, Lane EB: **Localization of the gene for human simple epithelial keratin 18 to chromosome 12 using polymerase chain reaction.** *Genomics* 1990, **7**(2):188-194.
- Dahlstrand J, Zimmerman LB, McKay RD, Lendahl U: **Characterization of the human nestin gene reveals a close evolutionary relationship to neurofilaments.** *J Cell Sci* 1992, **103**(Pt 2):589-597.
- Herrmann H, Aebi U: **Intermediate filaments and their associates: multi-talented structural elements specifying cytoarchitecture and cytodynamics.** *Curr Opin Cell Biol* 2000, **12**(1):79-90.
- Steinert PM, Chou YH, Prahlad V, Parry DA, Marekov LN, Wu KC, Jang SI, Goldman RD: **A high molecular weight intermediate filament-associated protein in BHK-21 cells is nestin, a type VI intermediate filament protein. Limited co-assembly in vitro to form heteropolymers with type III vimentin and type IV alpha-internexin.** *J Biol Chem* 1999, **274**(14):9881-9890.
- Hemmati-Brivanlou A, Mann RW, Harland RM: **A protein expressed in the growth cones of embryonic vertebrate neu-**

- rons defines a new class of intermediate filament protein. *Neuron* 1992, **9(3)**:417-428.
12. Titeux M, Brocheriou V, Xue Z, Gao J, Pellissier JF, Guichenev P, Paulin D, Li Z: **Human synemin gene generates splice variants encoding two distinct intermediate filament proteins.** *Eur J Biochem* 2001, **268(24)**:6435-6449.
 13. Xue ZG, Cheraud Y, Brocheriou V, Izmiryani A, Titeux M, Paulin D, Li Z: **The mouse synemin gene encodes three intermediate filament proteins generated by alternative exon usage and different open reading frames.** *Exp Cell Res* 2004, **298(2)**:431-444.
 14. Napier A, Yuan A, Cole GJ: **Characterization of the chicken transitin gene reveals a strong relationship to the nestin intermediate filament class.** *J Mol Neurosci* 1999, **12(1)**:11-22.
 15. Herrmann H, Hesse M, Reichenzeller M, Aebi U, Magin TM: **Functional complexity of intermediate filament cytoskeletons: from structure to assembly to gene ablation.** *Int Rev Cytol* 2003, **223**:83-175.
 16. Chabot P, Vincent M: **Transient expression of an intermediate filament-associated protein (IFAPa-400) during in vivo and in vitro differentiation of chick embryonic cells derived from neuroectoderm.** *Brain Res Dev Brain Res* 1990, **54(2)**:195-204.
 17. Cossette LJ, Vincent M: **Expression of a developmentally regulated cross-linking intermediate filament-associated protein (IFAPa-400) during the replacement of vimentin for desmin in muscle cell differentiation.** *J Cell Sci* 1991, **98(Pt 2)**:251-260.
 18. Simard JL, Cossette LJ, Rong PM, Martinoli MG, Pelletier G, Vincent M: **Isolation of IFAPa-400 cDNAs: evidence for a transient cytostructural gene activity common to the precursor cells of the myogenic and the neurogenic cell lineages.** *Brain Res Dev Brain Res* 1992, **70(2)**:173-180.
 19. Sejersen T, Lendahl U: **Transient expression of the intermediate filament nestin during skeletal muscle development.** *J Cell Sci* 1993, **106(Pt 4)**:1291-1300.
 20. Zimmerman L, Parr B, Lendahl U, Cunningham M, McKay R, Gavin B, Mann J, Vassileva G, McMahon A: **Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors.** *Neuron* 1994, **12(1)**:11-24.
 21. Vincent M, Levasseur S, Currie RW, Rogers PA: **Persistence of an embryonic intermediate filament-associated protein in the smooth muscle cells of elastic arteries and in Purkinje fibres.** *J Mol Cell Cardiol* 1991, **23(7)**:873-882.
 22. Fischer AJ, Omar G: **Transitin, a nestin-related intermediate filament, is expressed by neural progenitors and can be induced in Muller glia in the chicken retina.** *J Comp Neurol* 2005, **484(1)**:1-14.
 23. Schweitzer SC, Klymkowsky MW, Bellin RM, Robson RM, Capetanaki Y, Evans RM: **Paranemin and the organization of desmin filament networks.** *J Cell Sci* 2001, **114(Pt 6)**:1079-1089.
 24. Wakamatsu Y, Nakamura N, Lee JA, Cole GJ, Osumi N: **Transitin, a nestin-like intermediate filament protein, mediates cortical localization and the lateral transport of Numb in mitotic avian neuroepithelial cells.** *Development* 2007, **134(13)**:2425-2433.
 25. Dahlstrand J, Collins VP, Lendahl U: **Expression of the class VI intermediate filament nestin in human central nervous system tumors.** *Cancer Res* 1992, **52(19)**:5334-5341.
 26. Vaitinen S, Lukka R, Sahlgren C, Hurme T, Rantanen J, Lendahl U, Eriksson JE, Kalimo H: **The expression of intermediate filament protein nestin as related to vimentin and desmin in regenerating skeletal muscle.** *J Neuropathol Exp Neurol* 2001, **60(6)**:588-597.
 27. Chou YH, Khuon S, Herrmann H, Goldman RD: **Nestin promotes the phosphorylation-dependent disassembly of vimentin intermediate filaments during mitosis.** *Mol Biol Cell* 2003, **14(4)**:1468-1478.
 28. Sahlgren CM, Mikhailov A, Vaitinen S, Pallari HM, Kalimo H, Pant HC, Eriksson JE: **Cdk5 regulates the organization of Nestin and its association with p35.** *Mol Cell Biol* 2003, **23(14)**:5090-5106.
 29. Yuan Y, Lee JA, Napier A, Cole GJ: **Molecular Cloning of a New Intermediate Filament Protein Expressed by Radial Glia and Demonstration of Alternative Splicing in a Novel Heptad Repeat Region Located in the Carboxy-Terminal Tail Domain.** *Mol Cell Neurosci* 1997, **10(1/2)**:71-86.
 30. Granger BL, Lazarides E: **Synemin: a new high molecular weight protein associated with desmin and vimentin filaments in muscle.** *Cell* 1980, **22(3)**:727-738.
 31. Sandoval IV, Colaco CA, Lazarides E: **Purification of the intermediate filament-associated protein, synemin, from chicken smooth muscle. Studies on its physicochemical properties, interaction with desmin, and phosphorylation.** *J Biol Chem* 1983, **258(4)**:2568-2576.
 32. Price MG, Lazarides E: **Expression of intermediate filament-associated proteins paranemin and synemin in chicken development.** *J Cell Biol* 1983, **97(6)**:1860-1874.
 33. Bellin RM, Sernett SW, Becker B, Ip W, Huiatt TW, Robson RM: **Molecular characteristics and interactions of the intermediate filament protein synemin. Interactions with alpha-actinin may anchor synemin-containing heterofilaments.** *J Biol Chem* 1999, **274(41)**:29493-29499.
 34. Bellin RM, Huiatt TW, Critchley DR, Robson RM: **Synemin may function to directly link muscle cell intermediate filaments to both myofibrillar Z-lines and costameres.** *J Biol Chem* 2001, **276(34)**:32330-32337.
 35. Erber A, Riemer D, Bovenschulte M, Weber K: **Molecular phylogeny of metazoan intermediate filament proteins.** *J Mol Evol* 1998, **47(6)**:751-762.
 36. Zimek A, Weber K: **The gene for a cytoplasmic intermediate filament (IF) protein of the hemichordate *Saccoglossus kowalevskii*; definition of the unique features of chordate IF proteins.** *Gene* 2002, **288(1-2)**:187-193.
 37. Lewis SA, Cowan NJ: **Anomalous placement of introns in a member of the intermediate filament multigene family: an evolutionary conundrum.** *Mol Cell Biol* 1986, **6(5)**:1529-1534.
 38. Zimek A, Stick R, Weber K: **Genes coding for intermediate filament proteins: common features and unexpected differences in the genomes of humans and the teleost fish *Fugu rubripes*.** *J Cell Sci* 2003, **116(Pt 11)**:2295-2302.
 39. Dodemont H, Riemer D, Weber K: **Structure of an invertebrate gene encoding cytoplasmic intermediate filament (IF) proteins: implications for the origin and the diversification of IF proteins.** *Embo J* 1990, **9(12)**:4083-4094.
 40. Parry DA, Steinert PM: **Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism.** *Q Rev Biophys* 1999, **32(2)**:99-187.
 41. Lanzetta PA, Alvarez LJ, Reinach PS, Candia OA: **An improved assay for nanomole amounts of inorganic phosphate.** *Anal Biochem* 1979, **100(1)**:95-97.
 42. Walker JE, Saraste M, Runswick MJ, Gay NJ: **Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.** *Embo J* 1982, **1(8)**:945-951.
 43. Erzberger JP, Berger JM: **Evolutionary relationships and structural mechanisms of AAA+ proteins.** *Annu Rev Biophys Biomol Struct* 2006, **35**:93-114.
 44. Hirano T: **At the heart of the chromosome: SMC proteins in action.** *Nat Rev Mol Cell Biol* 2006, **7(5)**:311-322.
 45. Darenfed H, Ma X, Davis L, Juge N, Savard PE, Cole GJ, Vincent M: **Molecular polymorphism of the intermediate filament protein transitin.** *Histochem Cell Biol* 2001, **116(5)**:397-409.
 46. Duval M, Ma X, Valet JP, Vincent M: **Purification of developmentally regulated avian 400-kDa intermediate filament associated protein. Molecular interactions with intermediate filament proteins and other cytoskeleton components.** *Biochem Cell Biol* 1995, **73(9-10)**:651-657.