Origin of Nogo-A by Domain Shuffling in an Early Jawed Vertebrate

Aleksandra Shypitsyna, Edward Málaga-Trillo,*1 Alexander Reuter,1 and Claudia A.O. Stuermer*1
1Department of Biology, University of Konstanz, Konstanz, Germany
*Corresponding author: E-mail: edward.malaga@uni-konstanz.de; claudia.stuermer@uni-konstanz.de.
Associate editor: Claudia Kappen

Abstract

Unlike mammals, fish are able to regenerate axons in their central nervous system. This difference has been partly attributed to the loss/acquisition of inhibitory proteins during evolution. Nogo-A—the longest isoform of the reticulon4 (rtn4) gene product—is commonly found in mammalian myelin where it acts as a potent inhibitor of axonal regeneration. Interestingly, fish RTN4 isoforms were previously reported to lack the most inhibitory Nogo-A–specific region (NSR). Nevertheless, fish axons collapse on contact with mammalian NSR, suggesting that fish possess a functional Nogo-A receptor but not its ligand. To reconcile these findings, we revisited the early evolution of rtn4. Mining of current genome databases established the unequivocal presence of NSR-coding sequences in fish rtn4 paralogues. Further comparative analyses indicate that the common ancestor of fish and tetrapods had an NSR-coding rtn4 gene, which underwent duplication and divergent evolution in bony fish. Our genomic survey also revealed that the cephalochordate Branchiostoma floridae contains a single rtn gene lacking the NSR. Hence, Nogo-A most probably arose independently in the rtn4 gene of a gnathostome ancestor before the split of the fish and tetrapod lineages. Close examination of the NSR uncovered clusters of structural and sequential similarities with neurocan (NCAN), an inhibitory proteoglycan of the glial scar. Notably, the shared presence of transposable elements in ncan and rtn4 genes suggests that Nogo-A originated via insertion of an ncan-like sequence into the rtn4 gene of an early jawed vertebrate with myelinated axons.

Key words: Nogo-A, reticulon, neurocan, axon regeneration, neurite outgrowth inhibitors.

Introduction

The central nervous system (CNS) of mammals contains several proteins able to inhibit process extension in neurons and block axon regeneration. One of these, Nogo-A, is associated with myelin and oligodendrocytes in the CNS and acts as a powerful inhibitor of axon growth (Chen et al. 2000; GrandPre et al. 2000; Prinjha et al. 2000). Nogo-A is the longest of three major isoforms encoded by the mammalian reticulon4 (rtn4) gene. It contains an ∼200 aa–long N-terminal stretch also present in the Nogo-B isoform, followed by a unique variable region of ∼1000 aa known as the Nogo-A–specific region (NSR) and the ∼200 aa–long reticulon homology domain (RHD), which is common to all RTN4 isoforms and highly conserved among eukaryotes (Oertle, Klinger, et al. 2003). In particular, the NSR and a 66-aa extracellular loop within the RHD (known as Nogo-66) have been demonstrated to be responsible for the inhibitory effect of Nogo-A on neurite outgrowth (GrandPre et al. 2000; Fournier et al. 2001; Oertle, van der Haar, et al. 2003; Hu and Strittmatter 2008).

In contrast to mammals, fish appear to lack inhibitory proteins in their myelin and are capable of regenerating CNS axons (Bastmeyer et al. 1991; Wanner et al. 1995). For example, optic nerve and spinal cord axons can regenerate and establish functional connections (Gaze 1970; Becker CG and Becker T 2008). Accordingly, previous bio-informatic analyses failed to identify Nogo-A isoforms among the duplicated rtn4 paralogues of fish (rtn4a and b, formerly known as rtn4 and 6, respectively) (Diekmann et al. 2005). Nevertheless, we also have shown that delta-20, the most inhibitory region within mammalian Nogo-A, can block the growth of fish axons (Diekmann et al. 2005; Abdesselem et al. 2009). Taken together, these data would imply that fish possess a functional receptor for mammalian delta-20 despite their own ligand being either absent or not inhibitory.

On the other hand, zebrafish rtn4 paralogues contain the highly conserved and potentially inhibitory RHD. Recently, however, we reported that the Nogo-66 domain of zebrafish RTN4a blocks growth cone elongation neither in fish nor in mammalian neurons (Abdesselem et al. 2009) and that it is unable to activate the cellular signals required to inhibit axon growth in mammals (Hsieh et al. 2006). These experiments suggest that, during evolution, the inhibitory properties of Nogo-66 were lost in fish or, alternatively, acquired in land vertebrates. Characterizing the early evolution of rtn4 is crucial to clarify these issues. Moreover, because information on NSR-binding partners is very limited, structural and sequence analysis of vertebrate RTN4s would help identify putative Nogo-A–specific receptor(s).

Apart from Nogo and other myelin-associated molecules, strong inhibitors of axonal regeneration are also present in the mammalian glial scar. Among these, chondroitin sulfate proteoglycans (CSPGs) can effectively block axon regrowth (Fawcett and Asher 1999). In the fish optic nerve, a scar is also formed on lesion (Hirsch et al. 1995), but it apparently contains neither CSPGs (Becker CG and Becker 2000).
T 2002) nor other components with repulsive influence on retinal axons. It would thus seem that the major inhibitors of axon regeneration are either absent or inactive in the fish CNS. Nevertheless, the existence of a fish receptor able to recognize mammalian Nogo-A and cause collapse challenges this notion.

In this study, we took advantage of improved bioinformatic resources to revisit the molecular origin and early evolution of RTN4. Our analyses reveal the presence of Nogo-A isoforms in fish, rejecting the notion that this inhibitor was differentially lost/acquired in vertebrates. In addition, we present novel evidence suggesting that the NSR was derived from an evolutionarily unrelated CSPG, an inhibitory molecule from the gial scar.

Materials and Methods

Data

All protein and gene sequences are referred to by their NCBI/GenBank, ENSEMBL and JGI accession numbers (supplementary tables S1–S4, Supplementary Material online). The protein and gene sequences in this study correspond to 33 chordate and invertebrate taxa. Common names used in alignments and phylograms correspond to the following scientific names: nematode, Caenorhabditis elegans; tunicate, Ciona intestinalis; lancelet, Branchiostoma floridae; human, Homo sapiens; mouse, Mus musculus; rat, Rattus norvegicus; opossum, Monodelphis domestica; platypus, Ornithorhyncus anatinus; dog, Canis lupus familiaris; frog, Xenopus tropicalis; chicken, Gallus gallus; zebra finch, Taeniopygia guttata; carp, Cyprinus carpio; zebrafish, Danio rerio; goldfish, Carassius auratus; fugu, Takifugu rubripes; Tetraodon, Tetraodon nigroviridis; medaka, Oryzias latipes; salmon, Salmo salar; trout, Oncorhynchus mykiss; little skate, Leucoraja erinacea; and electric ray, Torpedo californica. We used the protein sequences of previously characterized reticulon genes to screen online public databases at NCBI (http://www.ncbi.nlm.nih.gov), ENSEMBL (http://www.ensembl.org), and JGI (http://www.jgi.doe.gov) using the BLAST algorithm (Altschul et al. 1997). Searches were carried out using default values with the low-complexity filter on.

Sequence Analysis

Identity plots and multiple protein alignments were performed using Geneious (http://www.geneious.com/) with the implementation of the MUSCLE algorithm (Edgar 2004). The resulting alignments were optimized manually. Final formatting of multiple sequence alignments was made with ALINE (Bond and Schuttelkopf 2009). Distant homologs were detected by searching chordate or complete nonredundant protein databases with PSI-BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Screening was carried out using default values with the low-complexity filter on. As queries, we used all previously characterized or newly found NSR sequences as well as consensus sequences generated from tetrapod or vertebrate aligned sets. All retrieved candidates were also globally aligned with the queries in order to eliminate cases where the observed similarity was based only on domain sharing.

Phylogenetic Analysis

Phylogenetic trees were reconstructed using the phylogenetic maximum likelihood method (Guindon and Gascuel 2003) in the SEAVIEW program, with implementation of the Dayhoff amino acid exchange matrix (Galtier et al. 1996). Reliability of the tree topologies was evaluated via the aLRT statistical approach (Anisimova and Gascuel 2006) as indicated by the numbers at the main nodes in figures 1A and 3A and supplementary figures S3B, S6, and S7 (Supplementary Material online). Hidden Markov models (HMMs) were constructed from manually edited multiple alignments using the Ammon HMM builder in the Indonesia package (Johansson 2001), which is especially intended for sequence classification.

Structural Analysis

Tertiary structure analysis was performed using the PSI-BLAST algorithm and homology structure modeling at the LOMETS (Wu and Zhang 2007) and ROBETTA (Kim et al. 2004) servers. When using PSI-BLAST for distant homology screening, the complex domain architecture of the studied protein is an obvious disadvantage. Nevertheless, this method can detect the presence of certain domains even in cases when low sequence conservation rates prevent the use of specialized domain/motif algorithms. Thus, we used all previously characterized or newly found NSR sequences to predict the domain architecture of this region. Structural annotations of vertebrate NCAN proteins were retrieved from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein/).

Results

In order to clarify the molecular origin of Nogo-A, we reassessed the evolutionary history of RTN4 proteins. Analysis of all 11 available fully annotated RTN4 sequences of tetrapods shows that in contrast to the RHD, which displays 80% average sequence identity across vertebrate taxa, the NSR is considerably variable in length and sequence (~32% identity between human and any of the three frog sequences analyzed) (supplementary fig. S1, Supplementary Material online). However, several short stretches are conserved within the NSR, including four tripeptide motifs (Asp-Leu/Ile-Val/Leu/Ile) that form a “molecular signature” (fig. 1A). It is of note that three of these motifs (M2, M3, and M4) are located within the inhibitory delta-20 region (fig. 18). Interestingly, peptides containing similar motifs are sufficient to bind β1-integrin and mediate cell adhesion in vitro (Kim et al. 2000).

Taking this into account, we ran reciprocal PSI-BLAST searches using all tetrapod NSRs as queries to look for related sequences in teleost fish. Positive matches were confirmed in zebrafish as well as in the pufferfish species Tak. rubripes and Tet. nigroviridis. As expected from distant PSI-BLAST searches, the fish NSR sequences are only ~18% identical to those of mammals and ~200 aa shorter than them (supplementary figs. S2 and S3, Supplementary Material online). However, they share most of the conserved...
NSR stretches, including all four putative \(\beta_1\)-integrin-binding motifs (fig. 1A). Notably, the NSR of zebrafish is not encoded in the \(rtn4a\) gene but in its paralogue \(rtn4b\) (fig. 2). Conversely, in both pufferfish species—which have lost the \(rtn4b\) gene (Diekmann et al. 2005)—the NSR is encoded in \(rtn4a\) (fig. 2). This is in agreement with the notion that a genome duplication in a bony fish ancestor led to the independent loss of gene duplicates in modern fish lineages (Amores et al. 1998).

Further analysis via PSI-BLAST, optimized sequence alignment, and comparison of exon–intron organization revealed that the fish RTN4a isoforms previously described as \(L\) and \(M\) (Diekmann et al. 2005) are in fact homologous to the mammalian Nogo-B and -C isoforms, respectively (supplementary figs. S4 and S5, Supplementary Material online). Therefore, we conclude that all major RTN4 isoforms of mammals (Nogo-A, -B, and -C) were present in the last common ancestor of bony fish and tetrapods, although it remains unclear whether these ancestral proteins had inhibitory properties.

The four members of the reticulon gene family, \(rtn1–rtn4\), are thought to have arisen from the quadruplication of a single \(rtn\) gene during early vertebrate evolution (Oertle, Klinger, et al. 2003). The absence of NSR-like sequences in RTN1, RTN2, and RTN3 suggests that this region was not present in the ancestral RTN of early chordates but acquired independently in the RTN4 of early vertebrates. To trace back the origin of the NSR, we surveyed the recently completed genome of the cephalochordate \(Branchiostoma floridae\) (lancelet), widely regarded as the closest living relative to the chordate ancestor (Delsuc et al. 2006). Using vertebrate RHD sequences as queries in BLAST searches, we identified a single \(rtn\) homolog in the lancelet. The RHD of the lancelet RTN is encoded by the same six exons as in

\[\text{FIG. 1} \quad \text{Nogo-A in vertebrates. (A) Left, phylogram showing the evolutionary relationships among vertebrate NSR amino acid sequences. The scale bar represents the number of exchanges per site. Right, multiple alignment of selected conserved clusters (in brackets) containing putative } \beta_1\text{-integrin–binding motifs (M1–M4). The degree of sequence similarity is highlighted using the cyan-to-red color code. The location of the clusters along the Nogo-A polypeptide is indicated in (B). (B) Schematic representation of the entire Nogo-A alignment. Within the central NSR (light blue), the relative positions of delta-20 (dark blue) and the putative } \beta_1\text{-integrin–binding motifs (red) are indicated. The percent identity plot above the scheme shows positions with identity scores higher than or lower than 50% in red and green, respectively.}\]

\[\text{FIG. 2} \quad \text{Gene structure of vertebrate } \text{rtn4} \text{ genes. Exons are drawn in scale and aligned in color-shaded columns. Numbers indicate intron phases. Gray boxes show } 5'\text{- and } 3'\text{-UTRs where this information was available. The NSR is absent in zebrafish } \text{rtn4a} \text{ and lancelet } \text{rtn} \text{ (included for comparison). Dotted lines in zebrafish } \text{rtn4b} \text{ represent gaps in database records.}\]
Vertebrate Rtns (fig. 2). In addition, secondary structure predictions and hydrophobic profiles indicate that the protein architecture of the RHD is strongly conserved from the lancelet to vertebrates (data not shown).

Next, we examined the evolutionary relationship between the lancelet and vertebrate RTNs at the RHD using the sequences of the nematode Cae. elegans and the urochordate Ci. intestinalis (tunicates) as outgroups. The topology of the resulting phylogram groups the lancelet RTN sequence with the RTN4 clade (fig. 3A and supplementary fig. S6, Supplementary Material online). Because the relationships among vertebrate RTNs cannot be reliably determined using distant outgroups (Li 1997), we repeated the analysis with only the lancelet RTN as an outgroup and found RTN4 to be the most basal group among vertebrate RTNs (supplementary fig. S7, Supplementary Material online). HMM analysis confirms that the lancelet sequence is more closely related to RTN4 than to other RTNs (fig. 3B). Furthermore, we also identified lancelet homologs of RTN4R (Nogo receptor) and the RTN4-interacting protein 1 (RTN4IP1) (supplementary figs. S8 and S9, Supplementary Material online).

Characterization of Expressed Sequence Tags (ESTs) revealed that the lancelet rtn gene produces at least three transcripts (fig. 3C). Two of them (RTN-B1 and RTN-B2) show remarkable similarities in sequence and exon–intron organization to vertebrate Nogo-B (supplementary fig. S10, Supplementary Material online). The variable part of the third transcript (RTN-C) bears no obvious resemblance to known reticulon proteins. Matching of the corresponding genomic regions confirmed the absence of putative NSR-containing exons in the lancelet rtn gene. Similar results were obtained for the single rtn gene of Ci. intestinalis (supplementary fig. S10, Supplementary Material online).

Altogether, these data strongly suggest that early chordates possessed only a Nogo-B–like isoform and that the Nogo-A isoform probably arose independently in the rtn4 gene of early vertebrates. If this event took place in a common ancestor of all jawed vertebrates, then Nogo-A isoforms would also be expected in cartilaginous fish (Chondrichthyys). We searched the limited raw sequence data presently available for these organisms but found only partial RTN4 fragments. Although the N-terminal sequences were too preliminary to

**Fig. 3** The rtn gene of the lancelet. (A) Phylogenetic relationships between RHD amino acid sequences of the lancelet and 69 vertebrate RTNs using tunicate and nematode proteins as outgroups. For simplicity, RTN1–RTN3 taxa are condensed into shaded triangles. (B) HMM profiling data for pairwise comparisons between the RHDs of lancelet, tunicate, and nematode RTNs and those of vertebrates (as displayed in the tree). Higher (positive) scores reflect shorter evolutionary distances. (C) Exon–intron organization of the lancelet rtn gene and its major transcripts. Exons and introns are drawn in scale. The conserved organization of RHD exons between lancelet and vertebrate rtns is highlighted within a frame. Nogo-B–like exons are shown in violet.
unambiguously identify Nogo isoforms, their C-terminal counterparts allowed for phylogenetic analysis of the RHD, showing that chondrichthyan RTN4s cluster among other jawed vertebrate sequences (fig. 3A).

The fact that the NSR of RTN4 is unique among vertebrate RTNs suggested to us that it may have been acquired by transposition of a foreign DNA sequence. To identify potential donor genes, we performed PSI-BLAST searches using vertebrate NSR sequences as queries. We identified at least 12 clusters of sequence similarity between the NSR and the protein core of NCAN, a member of the lec-tican family of CSPGs (fig. 4A and supplementary fig. S11, Supplementary Material online). Our analysis shows that the NSR and the matching sequence in NCAN have similar protein architecture consisting of an N-terminal immunoglobulin-like V-set domain and a mucin-like region (Retzler et al. 1996; Zander et al. 2007), with multiple putative $\beta_1$-integrin–binding motifs (fig. 4B). Most notably, transposable elements flank the corresponding exons of all rtn4 and ncan homologs examined (supplementary fig. S12, Supplementary Material online). These findings support the notion that Nogo-A originated from the insertion of an NCAN-like DNA sequence into the rtn4 gene of an early jawed vertebrate (fig. 5).

**Discussion**

In this study, we reconstructed the evolutionary history of RTN4 and its inhibitory domains from its ancestral form in early chordates to its differential evolution in vertebrates. Although it had previously been difficult to find correspondence between the fish and mammalian isoforms of RTN4 (Diekmann et al. 2005), the present analysis using improved algorithms and expanded databases shows that they are indeed homologous. Based on their protein domains and gene structure, the zebrafish RTN4a isoforms formerly described as L and M can now be matched to mammalian RTN4-B/Nogo-B and RTN4-C/Nogo-C isoforms, respectively, whereas the RTN4b isoform analyzed here corresponds to mammalian RTN4-A/Nogo-A.

The identification of fish counterparts to the three major RTN4/Nogo isoforms of tetrapods has important evolutionary consequences. It had been previously assumed that the RTN4 isoforms of fish (L, M, and N) and mammals (A, B, and C) had different evolutionary origins and that the NSR had been either lost in fish or acquired in mammals (Diekmann et al. 2005). The new data presented here demonstrate that the A, B, and C isoforms are common to these two vertebrate classes and that the major inhibitory domains of RTN4 (delta-20 and Nogo-66) already existed in the last common ancestor of fish and tetrapods. These findings might explain why fish axons respond to mammalian Nogo-A, and also suggest that bony fish may have acquired the ability to regenerate CNS axons due to divergent evolution and/or gene loss. This is consistent with our recent observation that the Nogo-66 domain of zebrafish RTN4a does not inhibit the regrowth of fish or mammalian axons (Abdesselem et al. 2009). Comparable data are not yet available for the NSR of zebrafish RTN4b, but the growth-permissive nature of fish CNS myelin (Bastmeyer et al. 1991) suggests that, if inhibitory, this protein may be absent from myelin and instead play repulsive roles during axon guidance. Functional characterization of fish RTN4 homologs will be needed to clarify how inhibition mediated by the NSR was gained/lost during evolution. Interestingly, fish are not the only case where Nogo-A–mediated inhibition may have been suppressed as a result of divergent evolution. For instance, frogs, salamanders, and lizards produce Nogo-A isoforms, yet they can regenerate at least some of their CNS axons as adults (Lang et al. 1995, 1998; Ferretti et al. 2003; Klinger et al. 2004).
Characterization of the lancelet rtn, together with the previous identification of its tunicate homolog (Oertle, Klinger, et al. 2003), allowed us to establish that the genome of the ancestral chordate most likely contained a single rtn gene encoding a Nogo-B–like isoform. Functional analysis of the lancelet rtn should help clarify whether such a primitive rtn might have influenced axon growth or regeneration. The gene phylogeny in figure 3A is in agreement with the classical textbook view on the evolutionary relationships between basal chordate groups, which places tunicates as basal to cephalochordates and vertebrates. However, the sharing of identical exon–intron organization at the RHD between lancelet and vertebrates suggests that the loss of introns at the RHD of Ci. intestinalis RTN (Oertle, Klinger, et al. 2003) is a derived condition. This would be in line with more recent molecular phylogenies that provided novel genetic material for the evolution of vertebrates (Kawashima et al. 2009).

The shared presence of β1-integrin–binding motifs in NCAN and Nogo-A has interesting mechanistic implications. Produced by astrocytes in the mammalian glial scar, NCAN is a well-known inhibitor of axon regeneration (Fawcett and Asher 1999) as well as of β1-integrin–dependent cell adhesion and neurite outgrowth (Li et al. 2000). Interestingly, the inhibitory myelin-associated glycoprotein (MAG) can interact not only with RTN4R but also with β1-integrin causing activation of the focal adhesion kinase (FAK) and repulsive growth cone responses (Goh et al. 2008). Similarly, recent data from the Strittmatter laboratory indicates that the NSR can inhibit cell adhesion and axonal outgrowth by blocking various integrin complexes (including β1-integrin) and FAK activation (Hu and Strittmatter 2008). These data suggest that distinct inhibitory molecules
of myelin and the glial scar may act via similar receptor complexes that influence integrin-mediated growth. Hence, at least for NCAN and Nogo-A, sharing of protein domains as a result of shuffling may have facilitated the evolution of such functional overlap.

The crucial issue of whether mammalian and fish NSR domains can functionally interact with fish β1-integrin on regenerating retinal axons requires experimental testing. Specifically, it needs to be clarified whether these interactions can elicit growth cone collapse as this would require activation of signaling molecules not targeted by β1-integrin signaling, such as Rho GTPases (Niederost et al. 2002). Functional evolutionary studies of Nogo-A and NCAN will be crucial to ascertain whether their fish homologs possess intrinsic inhibitory/repulsive properties and why these are not deployed during axon regeneration.

**Supplementary Material**

Supplementary tables S1–S4 and figures S1–S12 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank Gonzalo Solis for critical comments and suggestions. Supported by Deutsche Forschungsgemeinschaft grants.

**References**


