Comparative Genome Analysis of the Neurexin Gene Family in *Danio rerio*: Insights into Their Functions and Evolution

Alberto Rissone,† Matteo Monopoli,‡ Monica Beltrame,§ Federico Bussolino,*† Franco Cotelli,‡[1] and Marco Arese*†[1]*

*Department of Oncological Sciences, University of Torino, Strada Provinciale 142, 10060 Candiolo, Torino, Italy; †Division of Molecular Angiogenesis, Institute for Cancer Research and Treatment, University of Torino, 10060 Candiolo, Torino, Italy; ‡Department of Biology, University of Milano, 20133 Milano, Italy; and §Department of Biomolecular Sciences and Biotechnologies, University of Milano, 20133 Milano, Italy

Neurexins constitute a family of proteins originally identified as synaptic transmembrane receptors for a spider venom toxin. In mammals, the 3 known Neurexin genes present 2 alternative promoters that drive the synthesis of a long (α) and a short (β) form and contain different sites of alternative splicing (AS) that can give rise to thousands of different transcripts. To date, very little is known about the significance of this variability, except for the modulation of binding to some of the Neurexin ligands. Although orthologs of Neurexins have been isolated in invertebrates, these genes have been studied mostly in mammals. With the aim of investigating their functions in lower vertebrates, we chose *Danio rerio* as a model because of its increasing importance in comparative biology. We have isolated 6 zebrafish homologous genes, which are highly conserved at the structural level and display a similar regulation of AS, despite about 450 Myr separating the human and zebrafish species. Our data indicate a strong selective pressure at the exonic level and on the intronic borders, in particular on the regulative intronic sequences that flank the exons subject to AS. Such a selective pressure could help conserve the regulation and consequently the function of these genes along the vertebrates evolutive tree. AS analysis during development shows that all genes are expressed and finely regulated since the earliest stages of development, but mark an increase after the 24-h stage that corresponds to the beginning of synaptogenesis. Moreover, we found that specific isoforms of a zebrafish Neurexin gene (*nrxn1a*) are expressed in the adult testis and in the earliest stages of development, before the beginning of zygotic transcription, indicating a potential delivery of paternal RNA to the embryo. Our analysis suggests the existence of possible new functions for Neurexins, serving as the basis for novel approaches to the functional studies of this complex neuronal protein family and more in general to the understanding of the AS mechanism in low vertebrates.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Introduction

Neurexins are a family of neuronal proteins originally identified as synaptic receptors for the spider venom toxin, α-latrotoxin (Ushkaryov et al. 1992). In human and rodents, the 3 known Neurexin genes contain 2 promoters that drive the synthesis of a long-form called α-Neurexin and a short form called β-Neurexin (Rowen et al. 2002; Tabuchi and Sudhof 2002). Neurexins are classical type I membrane proteins with an N-terminal extracellular modular domain and a short intracellular C-terminal tail (Ushkaryov et al. 1992). They exhibit a peculiar domain content and distribution: in α-Neurexins, from amino to carbossi terminal, the mature protein contains 3 modular repeats, each constituted by an epidermal growth factor (EGF)–like domain surrounded by 2 laminin-neurexin-sex hormone (LNS)–binding protein domains, an O-linked sugar domain, a transmembrane domain, and an intracellular tail. The β-Neurexins possess, in their extracellular portion, a short β-specific amino acid stretch followed by a sequence common to α-Neurexins: one LNS domain (that corresponds to the 6th LNS domain of α-Neurexins), the transmembrane region and the intracellular tail.

The functions of Neurexins are not completely understood. Generally, they can be described as a highly heterogeneous family of synaptic adhesion proteins with a role in synaptogenesis and a potential role in differential neuron-to-neuron recognition. Insights into their functions come from their endogenous ligands that include the Neuroligins, Dystroglycan, and the Neurexophilins. The Neuroligins are a family of synaptic transmembrane proteins that are localized at the postsynaptic side of both excitatory and inhibitory synapses of the central nervous system (CNS). They are mainly composed by an acetilcholinesterase-like extracellular domain followed by a transmembrane stretch and an intracellular C-terminal region (Ichtchenko et al. 1995). Neurexins and Neuroligins form a Ca^{2+}-dependent trans-synaptic link (Nguyen and Sudhof 1997) that is instrumental in the formation of new synapses (synaptogenesis). Dystroglycan is a large transmembrane cell surface protein constituted by 2 subunits connecting the intracellular actin cytoskeleton to the extracellular matrix proteins (Henry and Campbell 1999) whose specific functions in the CNS are unknown (Sugita et al. 2001). Neuroligins and Dystroglycan bind to both α and β-Neurexins, whereas Neurexophilins, which are neuropeptide-like soluble proteins of unknown function, only show binding to α-Neurexins (Petrenko et al. 1996).

The best-characterized functions of Neurexins are related to their binding to Neuroligins. Indeed, different studies have shown that both Neuroligins and Neurexins when overexpressed in nonneuronal cells can induce, respectively, pre- or postsynaptic specializations in contacting axons or dendrites through their reciprocal interaction (Scheiffele et al. 2000; Graf et al. 2004).

Intracellularly, through a highly conserved amino acid stretch, Neurexins bind to the PDZ-containing protein CASK (Hata et al. 1996), a protein of the MAGUK family that connects Neurexins to the exocytotic machinery (Butz et al. 1998) and to the actin cytoskeleton (Biederer and Sudhof 2001). Neuroligins bind to one of the PDZ domains of the postsynaptic density proteins PSD95 that is an organizer

[1] These authors contributed equally to the paper.

Key words: Neurexins, alternative splicing, evolution, zebrafish.

E-mail: federico.bussolino@unito.it.

Mol. Biol. Evol. 24(1):236–252. 2007
doi:10.1093/molbev/msl147
Advance Access publication October 13, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

of the postsynaptic functions (Irie et al. 1997). Loss of function approaches has been published only for α-Neurexins. Mice null for one Neurexin α-isoform are viable and fertile, whereas mice depleted of more than one isoform die perinatally of respiratory problems (Missler et al. 2003).

The transcripts of the Neurexin genes are subjected to extensive alternative splicing (AS) (Missler and Sudhof 1998). Five different AS sites (SS1–5) are known in α-Neurexins, the last 2 being present in β-Neurexins too. This extensive processing has been shown to have 2 types of consequences: the first is at the structural level. Indeed, the interactions with α-latrotoxin, Dystroglycan, and Neuroligins are all tightly regulated by AS at the site 4 (SS4) (Sugita et al. 1999, 2001; Ichtchenko et al. 1995). AS at SS2 in the 2nd LNS domain also regulates Dystroglycan binding (Sugita et al. 2001), but not the one of Neurexophlins (Missler et al. 1998). No similar functions have been reported for the other splicing sites. Although the sites of AS are at different positions in the sequences of the 2nd and 6th LNS domains, the crystal structure of the 6th LNS domain of Neurexin 1 (Rudenko et al. 1999) showed that both splice sites are located in close proximity on the surface of the folded domain, suggesting that this location is a sensitive one for ligand binding. In addition to altering protein interactions, AS of Neurexin 3, but not Neurexin 1 or 2, can generate mRNAs encoding proteins lacking a transmembrane domain that are thought to be secreted from the cell (Ushkaryov and Sudhof 1993).

High heterogeneity is the other consequence of the complicated transcript processing of the Neurexins. AS is an important mechanism for generating protein diversity from a small set of genes. Different studies have shown that the 5 alternative splice sites of Neurexins are independently regulated and can give rise to more than 3,000 different isoforms (Ushkaryov et al. 1992; Ushkaryov and Sudhof 1993; Ullrich et al. 1995; Patzke and Ernsberger 2000; Rowen et al. 2002; Tabuchi and Sudhof 2002). Intriguingly, only another neuronal protein, DSCAM, found in Drosophila, produces such a number of isoforms (for a review see [Lipscombe 2005]). Because of these features, Neurexins could be enrolled in the so-far little group of candidate protein families that act in nervous system as molecular codes for the neuron-to-neuron recognitions events and make up the complexity of the CNS (Missler and Sudhof 1998).

Although orthologs of Neurexin genes are present in invertebrates such as *Drosophila melanogaster* and *Caenorhabditis elegans* (Tabuchi and Sudhof 2002), this gene family has been studied mostly in mammals. Given the lack of data for these genes in low vertebrates, we chose *Danio rerio* as a target for our research. Indeed, zebrafish presents some experimental advantages and what is important is a quite simple and well-characterized nervous system and broad developmental and physiological similarities to humans (Amatruda et al. 2002; Chen and Ekker 2004).

Here, we describe the isolation, the structural studies, and the AS analysis of the zebrafish Neurexin genes. We identified 6 different homologs (called *nrxn1a-b*, *nrxn2a-b*, and *nrxn3a-b*) in the genome that present a high degree of conservation at exonic level, a very similar exon–intron organization and AS regulation. Our data indicate a very high selective pressure operating both at exonic and at intronic level to preserve the gene structure, the AS mechanisms, and probably the function of these genes during evolution.

Through Reverse-transcriptase polymerase chain reaction (RT-PCR)–based methods, we analyzed the expression pattern of the bona fide orthologs of mammalian Neurexins during zebrafish development, finding that α-forms are expressed since the earliest stages of development and are maternally inherited. Surprisingly, for splice site 1 of *nrxn1a*, we found specific isoforms in the cDNA of adult testis that matched those present in the early stages of development, suggesting a possible paternal origin of some Neurexin 1 transcripts.

Our data indicate new possible functions of some members of the family of Neurexins. Understanding the genomics of these genes in zebrafish provides the basis for future functional studies with the use of knockdown and knockout technologies to assess the role of these genes in vertebrates.

Materials and Methods

Zebrafish Embryos Maintenance

Zebrafish was raised and maintained under standard laboratory conditions as described in the Zebrafish Book (Westerfield 2000) and staged according to Kimmel (Kimmel et al. 1995). Beginning from 24 h post fertilization (hpf), embryos were collected and cultured in fishwater containing 0.003% 1-phenyl-2-thiouera to prevent pigmentation and 0.01% methylene blue to prevent fungal growth.

Databases and Bioinformatic Analysis of Data

The following genome assemblies were searched: *Homo sapiens* (National Center for Biotechnology Information [NCBI] build 36), *Mus musculus* (NCBI m35 December 2005), *D. rerio* (Zv5 and Zv6 assemblies dated, respectively, 27 May 2005 and 30 March 2006). The *C. elegans* genome (WS150), *Tetraodon nigroviridis* genome (version 7 of the assembly), *X. tropicalis* genome (assembly v4.1), and *C. intestinalis* genome (JGI version 2.0) were analyzed with the tools available from Ensembl. Multialignment of human and zebrafish Neurexins was performed with AlignX of Vector NTI Advance 10.1.1 (Invitrogen Corporation, Carlsbad, CA) and are presented in Boxshade 3.2 (available at http://www.ch.embnet.org/software/BOX_form.html). Structural features of amino acid sequences of human Neurexins were predicted using release 19.28 (dated 30 May 2006) of PROSITE (http://www.expasy.org/prosite/).

Phylogenetic Tree

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1 (Kumar et al. 2004). Amino acid sequences of Neurexins from organisms representing different taxa were aligned in MEGA 3.1 using ClustalW algorithm with a Gonnet protein weight matrix. A rooted phylogenetic tree was built using a Neighbor-Joining (NJ) method. Bootstrap analysis for 1,000 cycles was used to assess the strength of the topologies.

Cloning of Zebrafish Neurexins

Public databases were searched to find genes and their genomic environments: Ensembl database http://www.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 1
Sequences of Primers Used for AS Pattern Analysis

Gene Name	Splice Site	Primer Name	Primer Sequence	Annealing Temperature (°C)	Full Insert Size (bp)
nrxn1a	SS1	zNRX1alfa SS1 for	5′-cggtagcgatggagtcaacag-3′	61	235
		zNRX1alfa SS1 rev	5′-ctccgaacccttgaatgtggc-3′		
	SS2	NRXAalpha-F2N	5′-gcaatggactgatgcttcacac-3′	61	638
		NRXAalpha-R2N	5′-tccatggctgaagaggagcag-3′		
	SS3	NRXA alpha-F4	5′-gtccgactcactgtcaatctag-3′	61	210
		NRXA alpha-R3	5′-gtggaactccagctgcgtgtg-3′		
	SS4	zNRX1 SS4 for	5′-tgttcaatgtgggcaccgatg-3′	59	300
		zNRX1 SS4 rev	5′-ccttctcccagcctccaatct-3′		
	SS5	zNRX1alfa SS5 for	5′-ggcaagcagaccacaactcca-3′	61	155
		zNRX1alfa SS5 rev	5′-atcacctctgggctcgggtaa-3′		
nrxn2a	SS1	zNRX2alfa SS1 for	5′-agccagggtgtgaagaacgat-3′	59	337
		zNRX2alfa SS1 rev	5′-tttcatcggtgctgctctgga-3′		
	SS2	zNRX2alfa SS2 for	5′-cgatggcaagttcaacgacaatg-3′	61	133
		zNRX2alfa SS2 rev	5′-gaggagcccgtccactgatat-3′		
	SS3	zNRX2alfa SS3 for	5′-tggacgggggaagagtcaag-3′	59	246
		zNRX2alfa SS3 rev	5′-tgcccgtctctatgtggtgg-3′		
	SS4	zNRX2 SS4 for	5′-acgagcctgcggtcactgt-3′	57	416
		zNRX2 SS4 rev	5′-ggtggtttcagtggacagga-3′		
	SS5	NRX2beta-F3N	5′-gggagattaccttcagttgcac-3′	61	1,485
		NRX2beta-R3	5′-tctcagacgtaatactccttgtc-3′		
nrxn3a	SS1	zNRX3alfa SS1 for	5′-aacggagggacgtgctctgtt-3′	61	155
		zNRX3alfa SS1 rev	5′-gctgctctggatgggattctg-3′		
	SS2	zNRX3alfa SS2 for	5′-cttggcatgatgtgaaggtcac-3′	61	126
		zNRX3alfa SS2 rev	5′-tgtagcctgtcgtggtcaggat-3′		
	SS3	zNRX3alfa SS3 for	5′-tacgcttcggctagagctgga-3′	61	216
		zNRX3alfa SS3 rev	5′-tcgcccaccatctgaccttct-3′		
	SS4	zNRX3alfa SS4 for	5′-cacattcaacattgggacggc-3′	59	411
		zNRX3alfa SS4 rev	5′-cggacgctgccattgattttga-3′		
	SS5	NRX3-F1	5′-gccgctggttcagccagaac-3′	61	1,105
		NRX3-R2	5′-tctgtttcttgttgccgcttttg-3′		

ensembl.org/) and NCBI (http://www.ncbi.nlm.nih.gov/). The degenerate primers used for part of the cloning were previously published by Russell and Carlson (1997). We set the optimal PCR reaction conditions for the degenerate primers on rat and bovine Neurexin cDNAs (kindly provided by T. C. Sudhof) and then we used aliquots of zebrafish cDNA from different developmental stages. For 5′ RACE technique, we used 2 different kits: FirstChoice RLMRACE Kit (Ambion, Austin, TX) and 5′ RACE System for Rapid Amplification of cDNA Ends (Invitrogen Corporation) following the manufacturer's instructions. The cDNA sequences obtained with PCR methods were compared with the genomic sequences to identify the splice sites.

Syntenic Analysis

To find evidence for the conservation of synteny, we compared genomic regions neighboring the zebrafish Neurexins with the genes neighboring the human Neurexins. Putative orthologs for each zebrafish gene were located on the human map using the comparative tools recently added to the Ensembl genome database (Hubbard et al. 2005).

Reverse Transcription-Polymerase Chain Reaction

Total RNAs were prepared from zebrafish adult organs (ovary and testis), oocytes, and embryos at different developmental stages using the Totally RNA Isolation Kit (Ambion) or the RNAgents Total RNA Isolation System (Promegam, Madison, WI), treated with DNase I RNase-free (Roche, Basel, Switzerland) to avoid possible contamination from genomic DNA and then reverse transcribed using Superscript II (Invitrogen Corporation), Oligo dT primers or the ImProm-II Reverse Transcription System (Promega), and Random primers. The cDNAs were then subjected to PCR amplification using specific primers (see below) using Expand High Fidelity Taq polymerase (Roche) or Taq Platinum DNA polymerase (Invitrogen Corporation) following the manufacturer's instructions. When possible, all primer pairs have been designed on different exons to avoid the amplification of DNA contaminations eventually present in cDNA preparations. Control PCR experiments with samples prepared without reverse transcriptase were performed to ensure that genomic DNA contamination did not contribute to the PCR amplification (data not shown). The PCRs consisted of an initial denaturation of the samples at 95 °C for 3 min, followed by 35 cycles. Each cycle consisted of a denaturation step at 95 °C for 30 s, a 30-s annealing step at the temperatures specified in table 1, and an extension step at 72 °C for a time depending on fragment length. A final extension cycle of 10 min at 72 °C was added to each PCR. Products were then separated on agarose gels at different concentration (from 1% to 3% maximum, based on the fragments length), visualized by ethidium bromide staining and then scanned with a Typhoon 8600 (Molecular Dynamics, Sunnyvale, CA). Nucleotide sequences of PCR products were determined by cloning in pCRII system (Invitrogen Corporation) followed by sequencing of both strands (PRIMM, Milan,

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Italy). A fragment of zebrafish β actin cDNA was amplified by PCR (35 cycles) as an internal control for the quality of cDNA using the primers ACTF (5V-AACTGGGAYGA-CATGGAGAA-3V) and ACTR (5V-TTGAAGGTCTCA-AACATGAT-3V). For the analysis shown in figures 6 and S4 (Supplementary Material online) we used a 2nd couple of primers specific for zebrafish β actin (Argenton et al. 2004) that demonstrates the lack of genomic contamination in our RNA preparations.

Gene-Specific Primer Sequences

The sequences of primers used in AS pattern analysis are listed in table 1. Table S1 (Supplementary Material online) shows the size of the fragments amplified by each couple of primers.

Accession Numbers

The accession numbers of zebrafish Neurexin family members are as follows: *nrxn1a* alpha, DQ641424; *nrxn1a* beta, DQ641425; *nrxn1b* alpha, DQ641426; *nrxn1b* beta, DQ641427; *nrxn2a* alpha, DQ641428; *nrxn2a* beta, DQ641429; *nrxn2b* alpha, DQ641430; *nrxn2b* beta, DQ641431; *nrxn3a* alpha, DQ641432; *nrxn3a* alpha soluble form, DQ641433; *nrxn3a* beta, DQ641434; *nrxn3b* alpha, DQ641435; *nrxn3b* alpha soluble form, DQ641436; and *nrxn3b* beta, DQ641437.

The accession numbers of Neurexins from different organisms are as follows: ceNRX alpha (*C. elegans*): C29A12.4 (Wormbase Gene ID); cbNRX alpha (*C. briggsae*): CAE64902; and ciNRX alpha (*C. intestinalis*): ENSCING00000005088 (Ensembl Gene ID).

Results and Discussion

Molecular Cloning and Characterization of the Zebrafish Neurexins

To identify the human Neurexins orthologs in *D. rerio* we used different approaches. At the beginning, we assembled the complete precursors of human Neurexins using Ensembl and VEGA transcript databases and we used these human sequences as queries for the Blast search programs at the Zebrafish Genome Browser (http://www.ensembl.org/Danio_Rerio) and for a screening of the public EST databases available at the NCBI (http://www.ncbi.nlm.nih.gov/BLAST). In parallel, we amplified the cytoplasmic domain of zebrafish Neurexins using degenerate primers complementary to rat and cow Neurexins (Ushkaryov et al. 1992; Ushkaryov and Sudhof 1993; Ushkaryov et al. 1994; Russell and Carlson 1997). Finally, we used the obtained fragments for an extensive *in silico cloning* using the Ensembl genomic database (http://www.ensembl.org/Danio_rerio/), and we identified the positions of the members of this gene family in different Linkage Groups (LGs). We confirmed each sequence prediction by RT-PCR. On the basis of the human genes structure previously published (Rowen et al. 2002; Tabuchi and Sudhof 2002) and the Ensembl gene predictions, we assembled the α- and β-forms of zebrafish Neurexins and then we completed the partial sequences using 5′ RACE techniques.

In different regions of the zebrafish genome, we found 6 genes homologous to mammalian Neurexins that, following the Zebrafish Nomenclature Guidelines available from the Zebrafish Information Network (http://zfin.org), we designated as *nrxn1a-b*, *nrxn2a-b*, and *nrxn3a-b*. In the last releases of the Ensembl database (Zv5 and Zv6, dated, respectively, 27 May 2005 and 30 March 2006), these different paralogs are located in the following genomic positions: LG12 (*nrxn1a*), LG13 (*nrxn1b*), LG21 (*nrxn2a*), LG7 (*nrxn2b*), LG17 (*nrxn3a*), and LG20 (*nrxn3b*).

Protein-sequence alignment (fig. 1 for β-Neurexins alignment and supplementary fig. S1, Supplementary Material online, for α-forms) reveals that the zebrafish Neurexins are very similar to the human proteins. At the structural level, each protein presents the typical organization of mammalian Neurexins (fig. 1): they are transmembrane proteins with a short cytoplasmic tail, and the extracellular region differs between α- and β-forms. The α-forms contain 3 LNS(A)-EGF-LNS(B) repeats, whereas β-forms present a β-specific portion just upstream of an LNS(B) domain. The greatest differences are located in the first LNS domain and the first EGF-like repeat, as already observed in a comparison between rat and cow proteins (Missler et al. 1998). Notably, the highest level of identity is found in the C-terminal tail (in particular in the last 10 residues) that interacts with the cytoplasmic scaffold protein CASK (Hata et al. 1996). A similar pattern of conservation is found in all the species that we compared. Table 2 presents the percentage of identity and conservation among the different members of the Neurexin families. Globally, the high degree of identity and conservation at the amino acid level suggests the existence of a very strong evolutionary pressure on this gene family and a possible conservation of the functions of its members, notwithstanding the approximately 450 Myr that have passed because the human and zebrafish common ancestor were speciated into 2 separate lineages (Aparicio et al. 2002; Zdobnov et al. 2002).

The full-length nucleotide and amino acid sequences of zebrafish Neurexins were used as queries to find homologous genes in other low vertebrate and invertebrate species by performing BlastN and TBlastN searches in different genomic databases. The genomic organization of some Neurexin genes of the *Xenopus laevis*, *T. nigroviridis*, and *T. rubripes* could not be completely elucidated, because cDNA/genomic sequences were only partially available. To assess the evolutive relationships among the Neurexin family members from different species, we performed phylogenetic and synthenic analysis (fig. 2). Figure 2*A* shows a rooted NJ tree obtained aligning multiple protein sequences of different Neurexins and zebrafish homologs. The phylogenetic tree organization suggests that the *D. rerio* Neurexins that we found were produced by a recent fish lineage-specific duplication event (Taylor et al. 2003). This notion is also supported by their different genomic localization, as described above. The topology of the tree is coherent with the relationship between the different taxa and is supported by robust bootstrap values in almost all the nodes; nevertheless it is not sufficient to fully elucidate that are the orthologs of the mammalian genes between each pair of zebrafish homologs. To obtain a clear perspective on the relationship of the 6 Neurexin genes in respect to the human genes, we

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

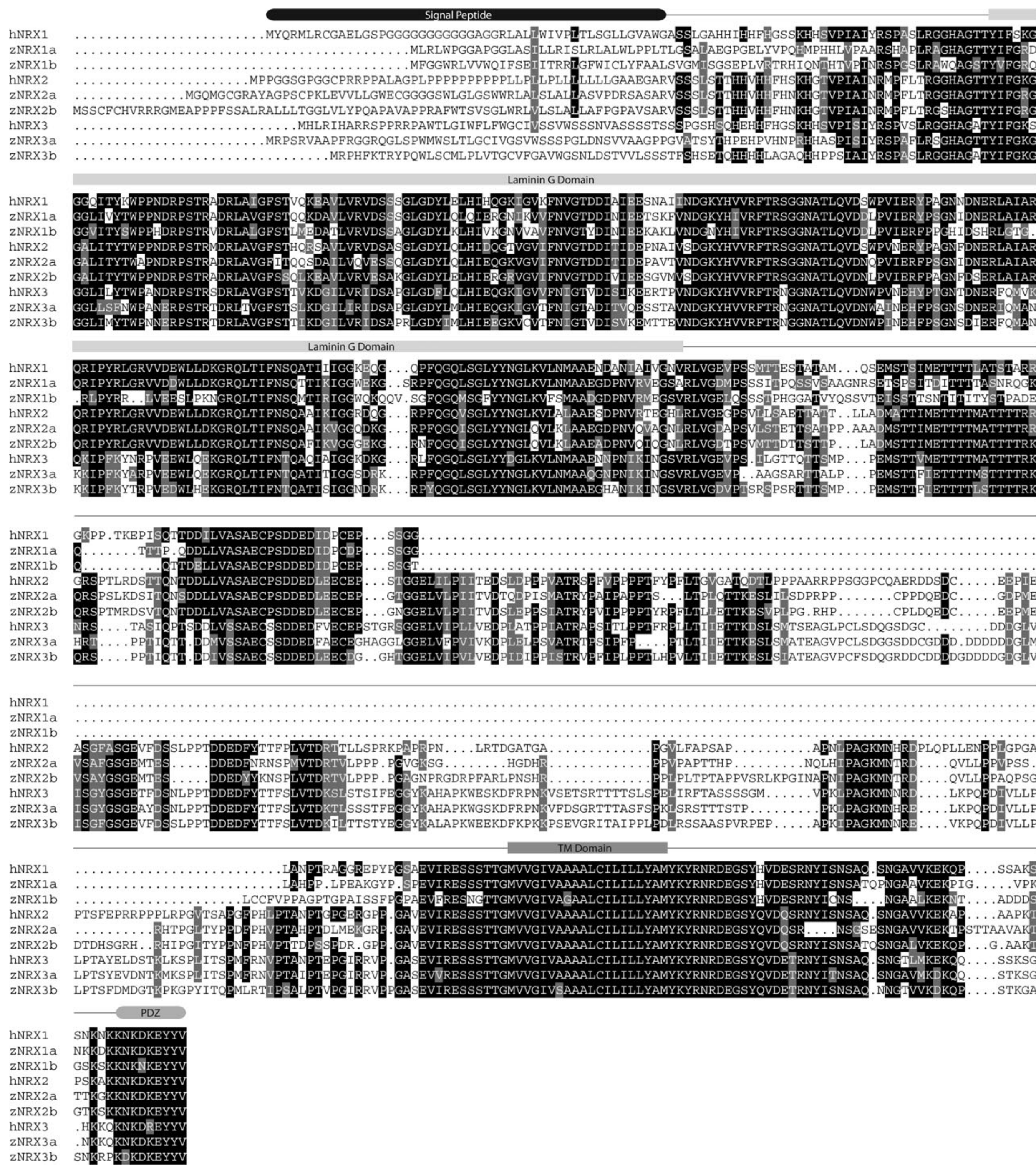

FIG. 1.—Multiple alignment of human and zebrafish β-Neurexin proteins. The amino acid sequences of human and zebrafish Neurexins β-forms were aligned using AlignX of Vector NTI Advance 10.1.1 based on the ClustalW algorithm and then edited with Boxshade. Black shading shows identical amino acids, gray indicates conservative substitutions, no shading indicates unrelated amino acids, and dots are gaps of alignment. Predicted features of human Neurexin 1β (NRX1) structure were obtained using PROSITE and are depicted above the amino acid sequences. Abbreviations: TM Domain, transmembrane domain; PDZ, PDZ interaction site; h, *Homo sapiens* and z, *Danio rerio*.

compared the neighbouring genomic regions of each zebrafish Neurexin with the mapped human genome using the comparative tools available in Ensembl. As shown in figure 2*B*, we found syntenic genes (or predicted genes) only in flanking regions of *nrxn1a*, *nrxn2a*, and *nrxn3a*. Combining data between protein-sequence alignment across different species and syntenic analysis, it became clear that teleost *nrxn1a*, *nrxn2a*, and *nrxn3a* are orthologs to the mammalian Neurexin genes.

Genomic Organization of the Zebrafish Neurexin Genes

In mammals, the 3 Neurexin genes exhibit the same genomic organization. Each gene contains 2 alternative

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 2
Percentage of Identity Conservation between Human and Zebrafish Homologs

	zNRX1a	zNRX1b	hNRX2	zNRX2a	zNRX2b	hNRX3	zNRX3a	zNRX3b
Alpha forms								
hNRX1	**79–88**	71–83	70–83	66–82	70–84	71–84	70–84	68–82
zNRX1a		75–86	68–81	66–80	69–82	68–81	67–81	66–80
zNRX1b			61–77	59–76	61–78	62–77	62–78	60–76
hNRX2				**71–84**	74–86	65–79	63–79	62–77
zNRX2a					79–90	62–79	61–78	61–76
zNRX2b						64–80	62–79	61–77
hNRX3							**82–91**	78–88
zNRX3a								81–90
Beta forms	zNRX1a	zNRX1b	hNRX2	zNRX2a	zNRX2b	hNRX3	zNRX3a	zNRX3b
hNRX1	**67–76**	57–70	68–77	65–78	65–78	63–76	59–76	59–75
zNRX1a		60–71	64–74	63–74	64–75	56–68	54–68	55–68
zNRX1b			52–66	53–67	54–70	49–64	48–64	48–65
hNRX2				**69–78**	69–79	54–69	51–67	52–67
zNRX2a					75–83	53–68	51–66	51–66
zNRX2b						52–68	50–66	50–64
hNRX3							**73–84**	70–82
zNRX3a								71–82

Numbers in bold indicate identity-conservation percentages among each couple of orthologue genes.

promoters that produce 2 major forms (α and β) and presents the same exon–intron structures (Tabuchi and Sudhof 2002). A common feature of the Neurexin is that the first LNS domain and the first EGF-like domain are coded by a unique large exon followed by several short exons that are alternatively spliced (SS1); all other LNS domains are coded by 2 or more exons, whereas all the EGF-like units are coded by single exons.

The genomic structure and the exon–intron boundaries of the zebrafish orthologs were determined by performing BlastN alignments of the different *nrxn* cDNAs against the Ensembl genomic database and then compared with the human genes. The *nrxns* present 2 main forms for each gene, 5 AS sites, the same protein-reading frame, and, except for minor exceptions, canonical (AG-GT) splice borders and the same number and size of exons. For a complete description of the genomic structure, see tables 3–5.

Because the sequencing of the zebrafish genome is not completed yet, it is difficult to calculate the exact size of the Neurexin genes. Nevertheless, *nrxn1a* and *nrxn2a* span a region of ~300 and ~680 Kb of genomic DNA, respectively. Different portions of *nrxn3a* have been found in 2 separate portions on LG17, and it is impossible to determine with enough confidence the complete size of this gene.

As the *nrxn1a* and *nrxn2a* exons are very similar in size to the human orthologs, the differences in gene sizes must depend on the intronic components, unlike the human and mouse genes that present the introns of identical size (Tabuchi and Sudhof 2002).

Sequence analysis shows that *nrxn2a* and *nrxn3a* are the most divergent members of the family. *nrxn2a* shows a partially different organization of the splice site 1 (SS1) with a very short exon 5 and an additional exon 6, found only in zebrafish (see table 4). Unlike the human gene, *nrxn3a* presents an extra exon (exon 17 in table 5) in a region that is not canonically involved in AS. Nevertheless, this exon presents some common features of alternatively spliced exons (e.g., number of base pairs divisible by 3 and shortness [Ast 2004]) and, more importantly, we detected 2 PCR products that differed for the presence or absence of exon 17. Unfortunately, given the absence of ESTs spanning this region, we cannot exclude the existence of a splicing error that is unusual, but possible (Kan et al. 2002).

The human NRXN3 gene can produce different soluble forms of the protein through the use of alternative acceptor splice sites for exon 23 (at least 4 different variants have been described). In zebrafish, a single soluble form of *nrxn3a* (and *nrxn3b*, see fig. S2 in Supplementary Material online) is encoded due to the use of a single acceptor site for exon 25.

As evidenced in tables 3–5, the analysis of intron–exon boundaries shows a very high level of conservation, especially for many of the alternatively spliced exons. This finding is of exceptional relevance considering the evolutive distance between *H. sapiens* and *D. rerio*.

Evolution of AS in Zebrafish Neurexins

As previously noted, Rowen et al. (2002) pointed out that sequences flanking some alternative exons are conserved between the 3 human Neurexin genes, in particular at SS2, SS3, and SS4. Because we found that zebrafish Neurexins present high similarity to the human genes in their amino acid sequence and genomic structure, we wondered whether they shared a common regulative mechanism of expression. With this aim, we analyzed the promoter regions and the AS regulative sequences of zebrafish orthologs. Given the lack of data concerning the human Neurexins promoters and potential transcription regulatory elements, it has not been possible to highlight a significant conservation in the genomic regions upstream the trascription start sites of the different genes. On the contrary, we identified conserved sequences in the flanking regions of different spliced sites for each zebrafish Neurexin. Intriguingly, an extensive intra- and interspecies analysis of SS2-, SS3-, and SS4-containing regions have shown a very high level of conservation (fig. 3 and supplementary fig. S3, Supplementary

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

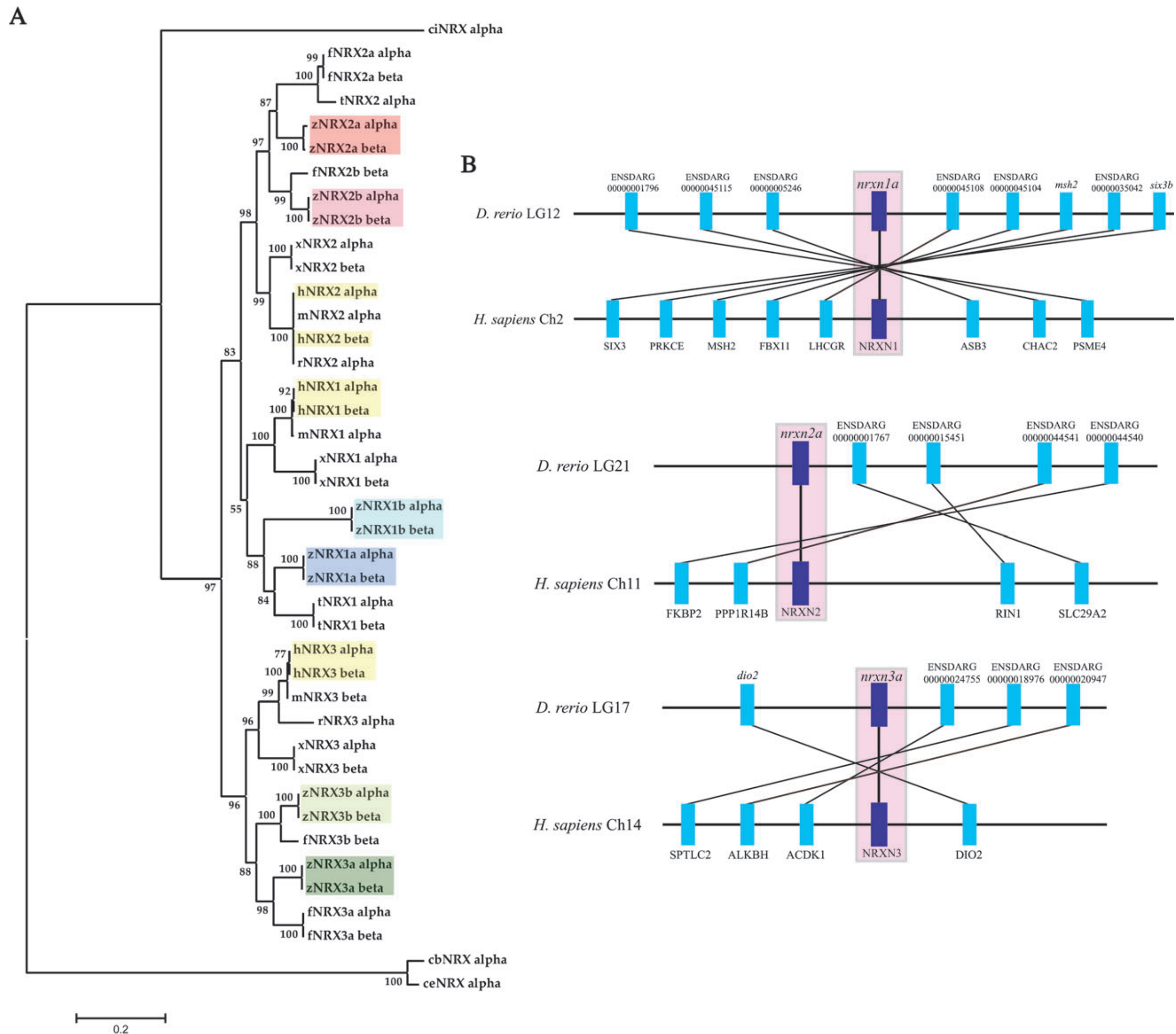

FIG. 2.—Phylogenetic relationship and syntenic analysis of Neurexins genes. (*A*) Evolutionary comparison of invertebrate and vertebrate Neurexin proteins represented in a phylogenetic-rooted tree generated using MEGA 3.1 program with a NJ method and 1,000 bootstraps. Different Neurexins are shaded in different colours. Branch lengths are measured in terms of amino acid substitutions, with the scale indicated below the tree. Numbers at nodes indicate percentage of bootstrap probabilities. Abbreviations: m, *M. musculus*; r, *R. norvegicus*; h, *Homo sapiens*; f, *Takifugu rubripes*; x, *X. tropicalis*; t, *T. nigroviridis*; z, *Danio rerio* (zebrafish); ci, *C. intestinalis*; ce, *C. elegans*; cb, *C. briggsae*. (*B*) The genomic contexts of each human Neurexin were analyzed for conservation of neighboring genes with their zebrafish orthologs. Each diagram represents the order of genes along a portion of the indicated chromosomes. Blue rectangle represents different gene that are syntenic. Identification code of Ensembl gene predictions is used when gene name is not yet available.

Material online). In particular, comparing the intronic regions that flank spliced exons at SS2 in both species (fig. 3*A* and *B*), a very high degree of sequence conservation is evident up to, respectively, 200 or 60 nucleotides from 3′ and 5′ splice site. Moreover, we were able to identify some regulatory elements (i.e., some repeats and the consensus sequence, YCAY, for the neural-specific splicing regulatory protein NOVA) (fig. 3*A*). Interestingly, these regulative elements are not present in all the zebrafish paralogs; for example, *nrxn1b* lacks the first two and the fourth repeat and all the NOVA consensus sequences, whereas both *nrxn3a* and *nrxn3b* lack the third NOVA site. This finding is very interesting because it could indicate that, after duplication, each zebrafish paralog has undergone differentiation mechanisms acting on their splicing regulative elements. As recently proposed (Su et al. 2006), AS and gene duplication may not evolve independently. In the early stage after gene duplication, duplicates may take over a certain amount of protein function diversity that was previously carried out by the AS mechanism (Su et al. 2006). This could be partially due to the accumulation of mutations at the level of intronic flanking regions of alternatively spliced exons as shown by our data. The progressive accumulation of different mutations could be the cause of AS neo or subfuctionalization phenomena.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 3
Comparison between Human Neurexin 1 and Zebrafish Neurexin 1a Gene Structure

Introns			Exons			
Danio rerio		*Homo sapiens*	*D. rerio*		*H. sapiens*	*D. rerio*
No.	Length	Length	No.	Length	Length	Flanking sequences
			1 (5′UTR)	n/a	69	n/a
1	n/a	2,786				
			2 (5′UTR + cds)	983	974	AGCAGAGCAGTC//AC,TGC,AGT,GAA,Ggtaagaggcctggattcgtc
2	8,290	101,546				
			3a* SS1	18	18	cgcctcggccgcccctccagAA,GAG,GCC,TAT,GTA,GGA,Ggtgggtacatggtttttatt
3	1,839	3,250				
			4* SS1	30	30	tgctcttgtcttcccttaagGT,CTT,GCG,CAC,C//TG,GGC,GAC,CAA,Ggtatgggtttaatttcttac
4	762	777				
			5* SS1	12	12	ccattatatcaatatcatagGG,AAA,AGA,AAA,Ggtatataatgcttctttcat
5	359,407	298,253				
			6	302	302	tttctttatgtatttgcagAG,AGA,GAA,GAA,T//AAC,CTT,CGT,CAGgtaaccgcgcagaaaggca
6	766	2,064				
			7b* SS2	45	45	ctccccccttttccacaaagCAC,TCA,GGC,ATT//CTA,CAT,TGT,TCGgtagatatcattctaattgt
7	1,207	1,021				
			8	162	162	ccctgtgctttgttgtctagGTT,ACA,ATA,TCC//TGC,TTG,AAA,GAGgtaagagatttgaacaattg
8	15,609	66,966				
			9	445	439	cttgttctttttcttaaagGTT,GTT,TAC,AAA//AT,GGA,AGA,TCA,Ggtagaatatgaatttatttc
9	18,143	13,950				
			10	384	384	ccttgtgtttgtgtgacagGC,ACC,ATT,TCC,G//CC,TGT,GAG,AGA,Ggtgagacactgagttttggc
10	24,143	6,822				
			11	204	204	cctttcttgctgtctctcagAT,GCC,ACG,ATC,CT//CT,GTC,AAT,CTA,Ggtacatctgccaataagagc
11	7,832	2,575				
			12* SS3	27	27	gtctgttctccccatctagAC,TGT,ATC,AGG,ATA,AAC,TGT,ACT,ACC,Agtaagtgactcttcccctct
12	13,769	22,007				
			13	123	123	ctgtgtgggacatgttatagGT,AAA,GGT,CCA,G//AG,CCC,TCT,GAA,Ggtacagcattatttcaagca
13	4,881	8,780				
			14	382	382	cgattattcattctggtaagGT,CAA,ATC,ACT,G//G,CTG,GTT,AAA,GGgtgagctgcttttgcttttt
14	84	1,237				
			15	191	191	cgattattcattctggtaagGT,CAA,ATC,ACT,G//G,CTG,GTT,AAA,GGgtgagctgcttttgcttttt
15	1,629	23,433				
			16	174	174	cgattattcattctggtaagGT,CAA,ATC,ACT,G//G,CTG,GTT,AAA,GGgtgagctgcttttgcttttt
16	11,604	6,736				
			17	120	120	cgattattcattctggtaagGT,CAA,ATC,ACT,G//G,CTG,GTT,AAA,GGgtgagctgcttttgcttttt
17	12,340	117,737				
			18 β (5′UTR + cds)	1,205	995	ACAGCTGCACGC//GG,GCA,GGA,CAT,Ggtgcgtagcttattcgacag
18	14,991	109,710				
			19	182	182	ctctctttctccttttgcagCG,GGA,ACC,ACA,T//CAG,CTT,CAG,ATTgtgagtcttacacaaacaaa
19	19,813	145,294				
			20	172	172	ttccctctctatctttacagGAA,AGA,GGC,AAT//GC,TAC,CCA,TCA,Ggtaagagccccaagccttta
20	7,336	36,278				
			21* SS4	90	90	tgtctctttaaactttaaagGC,AAC,ATT,GAT,A//TA,CTC,GAC,AAA,Ggtaaaaacactgattaaaac
21	1,293	1,364				
			22	317	326	gtaacgttctcccttggcagGT,CGG,CAG,CTC,A//ACA,ACT,CCA,CAGgtcagcgacaccttatcta
22	21,635	109,479				
			23b* SS5	82	88	cttctgtctctctggtacagGAC,GAT,CTT,CT//GGT,GGG,TTA,Ggttagtcatctttctttttat
23	165,405	21,452				
			24	302	308	cttctgtctctctggtacagGAC,GAT,CTT,CT//GGT,GGG,TTA,Ggttagtcatctttctttttat

Alternative spliced exons are marked with an asterisk and the number of the corresponding splice site (SS#); ''a'' or ''b'' indicates the splice donor/acceptor considered. Capital and lower case letters indicate exonic and intronic sequences, respectively. Bases marked in bold are identical in human sequences. Sizes are indicated in base pairs.

Although sequence conservation is less evident for exons at SS1 and SS5 (see below), there is no significant conservation at the borders of constitutively spliced exons (fig. 3*C* and *D*). This finding is consistent with some recent evidences, indicating that RNA sequence motifs, which are involved in the regulation of AS, are subject to strong selective pressure. The increased conservation associated with alternative exons is not due to amino acid selection pressure, but seems to reflect a purifying selection pressure on the RNA sequence itself (Xing and Lee 2006). This phenomenon has been called RNA-level selection pressure by Xing and Lee (2005).

AS and gene duplication are the 2 major sources of protein function diversity; when an alternatively spliced gene is duplicated, each duplicate copy could lose some AS isoforms due to the functional redundancy, or it could acquire new isoforms (Su et al. 2006). In this way, AS may contribute to the neo- or subfunctionalization of one or both

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 4
Comparison between Human Neurexin 2 and Zebrafish Neurexin 2a Gene Structure

Introns			Exons			
Danio rerio		*Homo sapiens*	*D. rerio*		*H. sapiens*	*D. rerio*
No.	Length	Length	No.	Length	Length	Flanking sequences
			1 (5′UTR)	158	69	GTTTCATGCAGG//GCGCACTTTCAGgtaagagcacttctgcgtgt
1	106,628	9,025				
			2 (5′UTR + cds)	1,246	974	ctttttcttttcactttcagGCATTCAAAGAA//AC,TGC,ACT,GAA,Ggtaagcacgatggtgtgcct
2	41,697	15,366				
			3* SS1	21	18	tcccaatactcctgctgcagTT,CTC,CAA,CAA,GTT,TTG,GGA,Ggtaagactgatgtttttatc
3	21,304	4,898				
			4* SS1	30	30	ttcctgtttttttccttcaagGT,CTG,GCG,CAC,C//TA,AAG,AGC,CAA,Ggtatggggcctcccccctttc
4	190	2,370				
			5* SS1	39	72	tgtcctgtccactggcccagCC,CCA,GGA,GCA,G//CT,CCT,GCA,GCA,Ggtaatgctctctttcgctct
5	25,129	n/a				
			6* SS1	15	n/a	ttctttttctttttccacagGT,GAT,CCC,AAA,TCA,Ggtaagaatccttaatcttgt
6	80,496	n/a				
			7	302	302	ttctatctcccttcaacaagGC,CAA,GTG,GAG//AAC,CTG,CGT,CAGgtaacgttcacctttgcaca
7	3,198	4,457				
			8b* SS2	45	45	tttttttcttttttttgaagCGC,GCA,GGC,GTT//CTG,CAT,TAT,ATGgtagatatcacattcaaatt
8	3,484	8,388				
			9	162	162	tatggcatttacgcaagcagGTG,ACC,ATA,TCA//TGC,CTT,AAA,GATgtaagtgcattttttaaagct
9	18,693	754				
			10	427	439	tttctctctccctgtcccagGTG,GTC,TAC,ACC//AG,GGC,AGG,AAA,Ggtcaggatattgcaagatca
10	2,478	6,110				
			11	384	387	cttcattcactttctccaagGC,TCC,ATC,TCC,G//AC,TGT,GAG,ATG,Ggtgaggaaagtagggcttaa
11	2,126	217				
			12	204	204	tcttcttctgcctctttcagAG,TCT,GCA,GTG,T//CC,GTC,AAC,CTC,Ggtaaaaccaccagactcgag
12	5,211	6,609				
			13* SS3	27	27	ctcactctctctcccaccagAT,TGT,ATC,GGG,A//AC,TGT,AAC,CTT,Agtaagtgattcttttatctg
13	16,337	1,541				
			14a* SS3	129	179	tgtgtgactttggccttctcTC,TCT,ACA,GGT,A//TG,ACC,GTG,GAA,Ggtgagtggaagacctcatac
14	2,370	398				
			15	382	382	cgtgtttggttctgccacagGC,CAG,ATG,AGC,G//G,CTT,GTA,AAA,GGgtaagtgtgagttttaatt
15	15,564	616				
			16	191	191	ggggaactgatgtattgcagG,TAC,ATC,CAT,TA//TG,GAC,CTC,AAA,Ggtatgcatattcacctatgt
16	22,486	1,540				
			17	174	174	tatttcttctgctctatagGA,GAG,CTG,TAC,A//GA,GGA,TGT,GAG,Ggtcagtatactgatgctctg
17	3,705	395				
			18	123	120	caacttgctgtttgtttcagCC,GGT,CCC,GGC,A//TC,TGC,AGT,GAC,Cgtgagtaccgctttgctctt
18	205,445	4,684				
			19 β (5′UTR + cds)	1,462	995	TGCATCTCGTGT//GA,GGA,GGC,CAC,Ggtaagcctagcgtttccttc
19	145,118	7,086				
			20	182	182	tgtttctcttcatttggcagCT,GGG,ACC,ACG,T//CAG,TTG,CAC,ATTgtgagtattaaagttcactg
20	288,475	4,697				
			21	172	172	aactgtccattcccacacagGAA,CAA,GGA,AAG//GC,TTC,CCG,TCA,Ggtaggaccacgtggagaatg
21	6,879	3,849				
			22* SS4	90	90	tctctctaaattctttgaagGG,AAC,ATT,GAT,A//TA,CTC,GAC,AAA,Ggtaataaaatgaattaatta
22	4,569	3,384				
			23	329	326	tcctctctgatccgtctcagGT,CGG,CAG,TTG,A//AGC,ATT,ACG,CAGgtaattacactgagaggcgg
23	13,367	2,380				
			24	79	79	ttttttttttcctcctgcagAAC,TCG,GAT,GAC//AG,CCG,GGC,ACT,Ggtgagaatacacacacacac
24	5,554	12,211				
			25a* SS5	788	887	tttctcctctctcctcctagGA,GGT,GAA,CTT,G//TAT,TAC,GTC,TGA

Alternative spliced exons are marked with an asterisk and the number of the corresponding splice site (SS#); ‘‘a’’ or ‘‘b’’ indicates the splice donor/acceptor considered. Capital and lower case letters indicate exonic and intronic sequences, respectively. Bases marked in bold are identical in human sequences. Sizes are indicated in base pairs.

the new genes. From this point of view, the lack and the acquisition of specific exons by *nrxn2a* and *nrxn3a* (exons 5 and 6 and exon 17, tables 4 and 5, respectively) together with the mutations at regulative regions at 3′ splice site of SS2 (and maybe other features not discovered by our analysis) could have contributed to neo- or subfunctionalization mechanisms avoiding the inactivation of 1 of the 2 new genes. Moreover, these data suggest that new functions of Neurexin genes could exist, at least, in the teleost.

One question still unanswered is the different level of sequence conservation between different splice sites. As mentioned above, there is a very high selective pressure at the intronic sequences at SS2, SS3, and SS4 in almost all the Neurexin genes (fig. 3 and supplementary fig. S3,

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 5
Comparison between Human Neurexin 3 and Zebrafish Neurexin 3a Gene Structure

Introns			Exons			
Danio rerio		*Homo sapiens*	*D. rerio*		*H. sapiens*	*D. rerio*
No.	Length	Length	No.	Length	Length	Flanking sequences
			1 (5′UTR)	178	108	GGGGGAGCGCGC//TTACGAAAATTGAGgttagtctgtttgtgttttt
1	6,295	71,682				
			2 (5′UTR + cds)	1,197	1,412	CAGCTATTGTGC//TC,TGC,AAC,GAA,Ggtaaatatatggtttatata
2	8,131	34,842				
			3* SS1	18	18	tctttgctgcaacccgccagAG,GCC,AAC,AAC,ATC,CCA,Ggtgagctcaatctcctttaa
3	3,758	19,168				
			4* SS1	30	30	ccgtttcctgttacttaaagGT,TTT,GCA,CAC,A//TG,GCG,GAT,CAA,Ggtatgccacactccttcact
4	204	2,819				
			5* SS1	12	12	ctatcatctcctttctcaagCC,AAG,GGC,AAA,Ggtacagaaccattttcttcc
5	n/a	344,428				
			6	302	302	atttatctcctcttttgtagCA,CGA,GAA,GAG,A//AAC,CTC,AGA,CAGgtaacgctaaaaatacaaaa
6	2,119	3,846				
			7b* SS2	45	45	tcttccccttttctcaaagCAT,TCA,GGT,ATT//CTA,CAT,TGT,CTGgtagatatcattcttattgt
7	3,998	1,852				
			8	162	162	ttgtcgttcctctcaatcagGTG,ACG,ATA,TCT//TGC,CTG,AAA,GAGgtaggttgtcaaactccatt
8	22,426	57,889				
			9	439	439	ctttgcttgcatccacgcagGTT,GTT,TAC,AAA//AT,GGA,AGA,TCA,Ggtcagttacacacacttaaa
9	22,229	5,100				
			10	384	384	ttttggtcttttttccagGC,ACC,ATA,TCA//CC,TGC,GAG,AGG,Ggtgagttcactgatttaca
10	26,551	88,480				
			11	204	204	cttgtgtgttttgtctacagAG,GCG,TCC,ATC,T//CG,GTC,AAC,TTA,Ggtattgtataaaacccccctc
11	4,222	6,494				
			12* SS3	27	27	acttattttcccccatctagAC,TGT,ATC,AGG,A//AC,TGT,AAC,TCC,Agtaagttaacactcaagttt
12	5,467	146,885				
			13	120	120	gtcctgaccgtcattctcagGC,AAA,GGA,CCA,G//AC,GTA,GCA,GAA,Ggtactatccttggtcctcaa
13	419	8,663				
			14	382	382	caacctgctgtggtttgtagGT,CAG,ATG,GTG,G//G,CTG,GTG,AAA,GGgtaagtacgttgaagaaaaa
14	1,661	801				
			15	191	191	ttctctttcccatcttttagG,TAT,ATA,CAT,TA//TG,GAC,CTT,AAA,Ggtacgtaaacaaatatgctt
15	1,288	772				
			16	174	174	cgctttcgtgttgtctttagGT,GAC,CTC,TTC,A//GA,GGA,TGC,GAA,Ggtacaccctctaacaactcg
16	8,013	19,675				
			17	27	n/a	ctgtttcttgcttaacaaagTT,GGT,TTC,ACC,AAA,GCA,GAT,CTA,AAA,Ggtatataaggttcaattcca
17	3,219	n/a				
			18	120	120	gctctgctttgctatcatagGT,CCT,AGC,ACT,A//AC,TGT,AAT,GAC,Cgtaagtcactctccgtctcg
18	n/a	291,410				
			19 β (5′UTR + cds)	861	681	TCTGGATGCTTA//GA,AGC,GGA,CAT,Ggtgagtttgcatggagttgt
19	n/a	186,682				
			20	182	182	ttttgtatttttcctccagCT,GGA,ACA,ACA,T//ATG,CTC,CAC,ATCgtaagtcaaatctaaatact
20	n/a	196,394				
			21	172	172	GAG,CAA,GGC,AAA//AC,TTC,CCC,TCA,G
21	n/a	28,223				
			22* SS4	90	90	ctgtcttttaaacttccaagGT,AAC,AGC,GAT,A//TT,CAG,GAG,AAA,Ggtaacttcctcctctgcttt
22	2,808	5,367				
			23	305	308	gatctttcttgcatcctcagGG,CGG,CAG,CTA,A//CCA,ACA,ATA,CAGgtaatgcacaacacacatga
23	35,810	107,174				
			24b* SS5	85	88	tttctgtctgtgtgttgtagACC,ACG,GAT,GAC//CA,GGT,GGG,CTA,Ggtcagttcacttttgcttgc
24	13,277	48,339				
			25* SS5	579	679	tccccatcccttccttacagCC,AAA,AAC,GTT,A//CAATTTGTTGATgtaagttgacaactcccaaa
25	5,261	6,821				
			26a* SS5	926	926	gtgtctgtgttcaattgtagGA,GGC,GAA,TTA,G//TAC,TAC,GTG,TGA

Alternative spliced exons are marked with an asterisk and the number of the corresponding splice site (SS#); ''a'' or ''b'' indicates the splice donor/acceptor considered. Capital and lower case letters indicate exonic and intronic sequences, respectively. Bases marked in bold are identical in human sequences. Sizes are indicated in base pairs.

Supplementary Material online). On the contrary, the exons of SS1 and SS5 present low rates of conservation among different human paralogs, but surprisingly their analysis reveals high conservation rates when they are compared with the zebrafish orthologs (data not shown). As suggested by our data (see also below), SS1 and SS5 could be involved in "minor" functions that wait to be elucidated, and as a consequence they could have been subjected to a milder selective pressure. However, it is intriguing that not all splicing sites have encountered the same evolutive destiny.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Analysis of AS during Zebrafish Development

One of the striking features of the Neurexin family is the extensive AS to which they are subject. There are several studies that describe this aspect (Ushkaryov and Sudhof 1993; Missler and Sudhof 1998; Missler et al. 1998), but all of them are focused on the adult organism or the late stages of embryonic development. Recently, Zeng et al. (2006) cloned the orthologs of the Neurexins α-forms in *X. laevis* and they partially described their AS during embryonic development. Interestingly, they found the expression of all 3 α-forms in oocytes of *X. laevis* with an oocyte-specific splicing pattern for the Neurexin 3. To date, despite the numerous efforts to understand the functional meaning of this high complexity, many questions remain unsolved. The existence of a huge number of different variants indicate that some of them can have unknown functions. In order to verify the AS pattern in vivo and discover possible new functions for these molecules outside the CNS, we closely examined their expression during zebrafish development. Using RT-PCR–based methods and specific primers designed around each splice site (except for SS5 of *nrxn2a*, see Materials and Methods), we analyzed how the pattern of AS varies from the early development (ovary and 1–2 cell stage) up to the early days of the larval phase (5 days). Figure 4 shows the results of this analysis.

In figure 5, we summarized the data and we presented them in a graphical form, showing the exon–intron organization and all the possible splice variants for each of the 5 sites of AS. Moreover, using the data published by Rowen and colleagues, we compared the different isoforms isolated in zebrafish with the human forms.

As already proved by Zeng and colleagues for *X. laevis*, each *nrxn* is expressed in the ovary and the 1–2 cell stage. These data are indirectly confirmed by an EST mapping in the 3′ UTR of *nrxn2a* (Accession. Number: DY563669) that derives from a whole ovary library and confirm the maternal origin of the transcripts.

The expression of all the transcripts follows a similar pattern. After the very early developmental stages, the expression levels decrease and, with few exceptions, they remain relatively constant until around 24 h, with the onset of an increase of expression that continues till the larval stages.

It is important to note that, for example, at the 18-somite stage (lane 8 in fig. 4*A–C*), most of the primary neurons (including, e.g., spinal primary motoneurons and Rohon-Beard neurons) start to grow axons and, beginning from ~24 hpf, the zebrafish embryo undergoes several events that induce synaptogenesis in different anatomical districts. From 72 to 120 hpf, the synaptogenesis increases progressively in all the major regions of the CNS (like the retina and the olfactory bulb) and in the myotome, as documented by several studies. This temporal window fits very well with the increased expression of both the major forms (α and β) of *nrxns*.

Surprisingly, the mechanism of AS is already finely regulated since the very early stages of development. As shown in figure 4*A*, the pattern of isoforms for SS1 of *nrxn1a* is quite different between the ovary and the 1–2 cell stage with the presence, in this stage, of specific isoforms, present also at the 64-cell stage and only during the first larval stages (72–120 hpf). The other major differences in the pattern of splicing are evident at the 50% epiboly stage, the early embryogenesis (8-somite stage, where a specific isoform is induced), the 15- to 20-somite stage, and finally from 24 to 120 hpf.

At SS2, from early stages to late somitogenesis the form without any insert predominates, with the exception of late epiboly and the 8-somite stage where the intermediate form is present too, although at a low level of expression. At 72 hpf and 5 days, the other forms are expressed at higher levels. Although at the earliest stages both forms of SS4 are present at about the same level, starting from 24 hpf onward the isoform containing the insert begins to be more represented. It has been shown that, at least in rat brain, Dystroglycan can interact with specific Neurexin 1 isoforms devoid of inserts at SS2 and SS4 (Sugita et al. 2001) and more recently, that zebrafish Dystroglycan is maternally expressed (Parsons et al. 2002). Our data could suggest a role of these isoforms in the interaction with Dystroglycan in zebrafish.

At the level of SS3, we identified both possible forms (with or without the exon 12) starting from 72 hpf; before this stage, only the larger form appears to be expressed.

Regarding SS5, the possible forms differ only by just few nucleotides that make the analysis sometime difficult; both forms are clearly present in the last 3 stages examined (maybe because both have a higher expression), whereas in the previous stages it is possible to notice the presence of at least the lower one.

In figure 4*B*, the AS analysis for *nrxn2a* gene is presented. Like *nrxn1a*, the splicing pattern at SS1 is very complicated with some different isoforms of very similar molecular weight, but, unlike *nrxn1a*, the pattern seems to be unchanged until 50% epiboly; after this stage, the expression decreases remarkably and becomes more discrete. Finally, from 72 hpf onward the expression of only-specific isoforms is evident.

Only one isoform (without the insert) is expressed at SS2 during all the embryonic stages and only during the larval stages (72 hpf and 5 days) the other variants are present at low levels.

SS3 presents a relatively complex pattern; indeed 4 different variants are possible at this splice site. The exon

→

FIG. 3.—Comparison of exonic and intronic flanking sequences between human and zebrafish Neurexins genes. Panels A and B show, respectively, the 3′ and 5′ splice sites of the alternative spliced exon at SS2. In C and D are shown the intron–exon flanking regions for the same constitutively spliced exon of different Neurexin genes (exon 8 of *nrxn1a* gene, see table 3). Sequences were aligned by ClustalW multiple sequence alignment using AlignX of Vector NTI Advance 10.1.1 and edited with Boxshade. Black shading shows identical nucleotides, gray indicates conservative substitutions, no shading indicates unrelated bases, and dots represent gaps of alignment. Note the high degree of nucleotide conservation between humans and zebrafish homologs at the alternatively spliced site (except for *nrxn1b* that has lost one of the alternative donor splice sites). On the other hand, constitutively spliced exon (C and D) do not present comparable levels of conservation between the different homologs.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

A SS2 3′ splice site

```
                                                                                Repeat
hNRX1  GGAGGCCTACTTGT......TAGTCTGCTTTTTTCTTG.TAAAATTTTTTTGCCCTTTTTCTATGTT.ACTAACATGCTT
zNRX1a GAACTCCACGGTGCAG.ATCTCGTCTGCTTCTCCTTTCCTTCTCTTCTTTTGCCCTTTTAAATGTTTACTAACCTGCTT
zNRX1b GTAGTCAGGTCAATAA....TCCACTCTATAGGACCAGTTTTTAGACATTGAGACGCCCACACTAATGCATCACGTACTG
hNRX3  GGTGGCATGGTTTTCC.A..TAGAGTCCTGGTCTCATTGTCTTCCCCCCACTCTCTTTTTCTGTGTCGGGTGCCTTTTTT
zNRX3a TGTGGCTATGATCC..TTTTTCTTTTCCACCTTTTCTT.ACTCACTCATATGCTTTGCTTGTCT.ATGGATAACATTTGT
zNRX3b AGTGGAATATCATGTCCTGCTCTTTTCTCTCTTTTCCTT....TCCTGTTTCTCAAATCTTGGGAAGAAACAGTTTTTTTT

                     Repeat        Nova Nova          Nova   Repeat
hNRX1  ........AATAACTAACATG...TCATTCATGGAATGTTTCATTCTACTAAC..C.GTTACCTGAATCAAA.....AA.
zNRX1a TTGGGGAAAGAAACTAACCTT...TCATCCATGGAATGTGTCATT.TACTAAC..CTGTTACCTCTTACTAACGTGCAAC
zNRX1b AAT...AAACACAATTACTAAA.AGCACAGAGGGAATGTGCCCTTATACTAACCAAGTGTTTCTTTAGCTTT..T..AA.
hNRX3  TTG.....AGAAACTAACACAAGATCATTCATGCAATGTGTCATTTTACTAAC..CGACTAATG..TACTAAC.....AC
zNRX3a TCAG..ATCCAAACTAACCAC...TCATG.AACCATTAACTAATT.TACTAAA....CAGAACCTAAACTAA.......
zNRX3b TTTAATTTCGAGACCAACACTAATCCATTCATGGGTCATTTTTAT.TACTAACGTA.CTTCTAACCAACTAATGTACTAA

                                                                        Branch
          Repeat                                                        Point
hNRX1  ....ATGTACTAACAT...GGTAGAGACCATCATATTTTTTAAGTAACGATCGTT.ATTCTGTTCACAGTTTTT.A.ATT
zNRX1a GCTGATGTACTAACCTTTGGCAAGAGACCATC...TTTTTTGATTGAGGAACGGT.GTTCTGTTCACAGTTTTTTA.ATT
zNRX1b ..ATATTCACATTGGTGA.TCATCATACATT..TACTTTGTGAATGACAGATGAT.ATTCAGT.CACAGTTTAT.A.ATT
hNRX3  CCTAACG.ACTCATCTTT..GTTTTTAGT......TTTTTTAATTAACAATCG....TTCTGTTCACAATTTTT.A.ATT
zNRX3a ....TCATAACTTCATTTGCGTTTATTCCA.......TTTTATTAACAACCC....TTTTGTTCA...TTGTTTA.ATT
zNRX3b CACTCTATATCTTCATCGTTATTTTCTTTCTCTTTTAATTTCATCCATTAACAACCGTTCTGTTCACTATTGTTTTGATT

          Pyrimidine Tract                       Exon
hNRX1  TCCTTTCTCCCCCCT.TCTCCA.CAAAGCACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTTCG
zNRX1a TCCTTTCTCCCCCCT.TTTCCA.CAAAGCACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTTCG
zNRX1b TGCTTTC...TTC.T.TTTCCA.CACAGCACTCAGGCGTTGGATATGCTATG....................
hNRX3  TCCTTTCTTCCCCCT.TCTCCG.CAAAGCACTCAGGCATCGGACACGCTATGGTAAACAAACTACATTGTCTG
zNRX3a TCCTTTCTTCCCCCT.TTTCTC..AAAGCATTCAGGTATTGGACACGCTATGGTAAACAAACTACATTGTCTG
zNRX3b TCCTTTCTTCCCCCCTTCTTCTTGAAAGCACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTCTG
```

B SS2 5′ splice site

```
                  Exon
hNRX1  CACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTTCGGTAGATATCA..TTCTAA.TTGTTTCTTATTTTGGGTTTGCC
zNRX1a CACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTTCGGTAGATATCA..TTCTAA.TTGTTTCTTAGTTTGGGTTTGCC
zNRX1b CACTCAGGCGTTGGATATGCTATGGTAATATCATTCAGTTTATTTCTCTTATTTT..GTGTATGTTTATAAAGTATATCTATCTATC
hNRX2  CACGCAGGGATTGGACACGCTATGGTAAACAAACTGCATTATCTGGTAGATATCA..CTCTAA.TTGTCTTACCGTTTGG.TTTGCC
zNRX2a CGCGCAGGGGTTGGACTCCCTACGGTAAACAAACTGCATTATATGGTAGATATCACATTCTAATTTGTCTAGT...TTGGTTTGGCC
zNRX2b CACTCAGGCATTGGACACGCTATGGTAAACAAACTGCATTATATGGTAGATATCA..TTCTAA.TTGTCTTCTACTTTGG.TTTGCC
hNRX3  CACTCAGGCATCGGACACGCTATGGTAAACAAACTACATTGTCTGGTAGATATCA..TTCTTA.TTGTCTCTTTTTTGGGTTTGCC
zNRX3a CATTCAGGTATTGGACACGCTATGGTAAACAAACTACATTGTCTGGTAGATATCA..TTCTTA.TTGTC.CTTTT..TTGGGTTTACC
zNRX3b CACTCAGGCATTGGACACGCTATGGTAAACAAACTACATTGTCTGGTAGATATCA..TTCTTA.TTGTCCTTT..TTTGG.TTTGCC

hNRX1  GTGTGTTTTCTT......TCTTTCTTTTAACT.GTAGACATCAT
zNRX1a GTGTGTCTATTT......TCCTTTTTTTGAATCGTAGACATCAA
zNRX1b TATCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTGTC
hNRX2  GTGGGTTTTTTAACTGTTCGATTCTCCCTCTTTTGTTTCCTCT
zNRX2a CTTTCTTTTTTTA.....TCTCTTTTTTATGCTTTTACTGT...
zNRX2b CTGGATTTCGTT...TCCTATTTCAATTCTGT.TTGCTTTACT
hNRX3  GTGTGTTGCCCCC.CTTTTCCTCTCCAACCCCCCCTTTTATCCT
zNRX3a GTGTGTTTTCCTG.GCCCTCTTTCTTCTTTTTTGGGGTT.TCAT
zNRX3b GTGTGTCCCTCTCCCTTGTTTTTTTTTCTTCTTGAGTTTCATT
```

C 3′ splice site

```
hNRX1  AAAAAAGAAAAAAGCAAAACAAAACA.AAACAAAAAAACAAAAGCAAAACAC.....AAAACAGGCATATCCCAGGATTACAGTA..TCT.AGGT
zNRX1a ACAAAAGGAAAAAG.AAATCTAAACA.AATCCAAGCATTCAAATCCTTACGAGGGGGACCAAAGCCCTGTAACAGTATTGAAGT.....T.ATAT
zNRX1b TCAGATCTGATGTGCTCTCCATTGCA.TCATCATTATGCAGCATTCATATGA.....AAGAGACTGAAGGATCAGGAGTTCAGAA..AGT.ATCC
hNRX3  ATCAAATCAAATTAGAAAGTATATCT..CATGAAAATCTTAATGAAAGTGATGCCACAAGTAGGGGAAGCCACTGGGTTATGATAG.GAT.CGTA
zNRX3a GCAGGTACACAATCAAAGACAGACCACAACCAGCAAGACGAAGCAAACCTATTCCTGGTAGAATCGCCATCACAACCTTGACGC...TCA.TGAA
zNRX3b ATTGGGGGAAAATCCAAACAAGATAA.ACCAAAACAGACAAAACAAAAACAC....GAGAGAGGTTATCGCTCAGGTTAAAAGCACTTCT.TAGA
hNRX2  AGACAAGCAGACATAAAAATGGGGGTCCCCCTCCCCTGCTGCCCTCTGGCC.....CTCCTTTGCCTAGCCGTAGCTCTTCCTAGGGAT.TTGA
zNRX2a ATTAATTTCTTTATTTTTTTAGTTGAGAAAACAAGGATTTCGGCTGTTGCAA.....AATGAAAAGATGTCCCACGTTCCCTCTTTGTTTCTTCC
zNRX2b CCACAATGACTCTTTAATCCAGTGTCTCTCTCTCCTCTCTCTGTTCCTCTGG......GGGAATCTCTCTCTGTTTCTCTCTCTCTCTCT.CTCT

                                                        Exon
hNRX1  CTGAGT.....ATAGCGTGTCCTTTCCCTGTGCT...TTGTTATCTAGGTGACAATATCAGTGGATGGG
zNRX1a GTGAGT.....ATAGCGTGTCCTTTCCCTGTGCT...TTGTTGTCTAGGTTACAATATCCGTGGATGGG
zNRX1b AGCTGC.....TCTCTCCTCTCCTCCTGTGGT...TTGTTTTGCAGGTTACGATATCTGTGGACGGA
hNRX3  AGGTCATTGTTGGGATTTTTTTTGTCCCTAT.GT...CCCTCCACCAGGTGACAATCTCTGTGGATGGC
zNRX3a TGGATTT..TTCCCTCTTCTTTTTTTTT.TGTCGTTC.CTCTCAATCAGGTGACGATATCTGTGGATGGC
zNRX3b AGGAGTTACTGATTTGTTGTTTTCTCCATATTGTCA.TCATTTAACAGGTTACTATATCGGTGGATGGT
hNRX2  TGTTGGTTGTCTCCTGTCTTTTCTCCGTCCCTCCT...GGGCTGGACAGGTGACCATCTCGGTGGACGGG
zNRX2a CCCTCTCTCTTTCTCCTTTCCGTCCGTCTATGGCATTTACGCAAGCAGGTGACCATATCAGTGGACGGG
zNRX2b CTCTCTC...TCTCTCTCTCCGTCTCTCTCTCTC..TTTCTCCCGCAGGTCACTATCTCAGTGGATGGT
```

D 5′ splice site

```
            Exon
hNRX1  TTTATGGGCTGTCTCAAAGAGGTAAATA...TCGTTGC..CATCAGAATCATTCC..T..A....ATTCATGGTC...TCATCATTTCAGATCCA
zNRX1a TTTATGGGCTGCTTGAAAGAGGTAAGAG...ATTTGAA..CA.....ATTGCTCCT.CGGAA.AAAATCAAATT....GTT.CCATCCAGAATTA
zNRX1b TTCATGGGCTGTCTGAGGGAGGTAAAGA...AACTGCCA.CACCA.AATCAGTGCAGGCTGTGTAGAGCTTAGAGTCACAATCGAATCACAATGA
hNRX3  TTCATGGGCTGCCTTAAAGAGGTAAAG......TTCACCCAATTCTATTTAATGCA..CCATGTGATTGTTTCATTTATAAACAAATGACATGTC
zNRX3a TTTATGGGTTGCCTGAAAGAGGTAGGTT...GTCAAACTCCATTGCAATTTTTCG..TCTACATGCATCATATCAGAGTGTCCTTGTCTGGACTG
zNRX3b TTCATGGGCTGCTTGAAAGAGGTAAGTT...CTGTGAA..GATTAGAATTGTTTTA.TAGAGCAAAATCAAAAG....GTTTCCACTCACTGGGA
hNRX2  TTCATGGGCTGCCTCAAGGACGTGAGTAGGGCTAGGGAGGGGCTGCAGGCTGTTCAGGAGCCCTTTGTCTCCTG.TCTTCAGCGTGACTTAGTCT
zNRX2a TTCATGGGTTGCCTTAAAGATGTAAGTG...CATTTTT...AAAGCTGTTTTGAATTTAGATCACAAAGAGGACCTTAATGTTCTTTTTTTTTGT
zNRX2b TTCATGGGCTGCTTGAAGGATGTGAGT....CTCTGCA..GA......TCAGCAC....GAACACAGTCAGCAGTGCTGTTTCTGTGCAGAACCG
```

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

14 presents 2 different 3′-splice sites (14a and 14b) that produce isoforms that alter the size of the amplicons of 9 nucleotides, so it is impossible to discern between these 2 isoforms in an agarose gel, and we consider each band as composed of these 2 different variants. Both bands are present till 50% epiboly stage (with the lower one more intense), then they decrease at very low levels and at 24 hpf both are equally present, whereas at the last 2 stages the long forms are more represented.

The SS4 presents both possible isoforms from the first stages, although in the ovary the lower one is very faint. Notably, like for *nrxn1a*, the expression increases starting at 24 hpf and the most expressed form is the longest one.

Regarding SS5, at the first stages of development, both the possible forms are visible and, starting from stage 64 cells, there is a decrease up to stage 72 and 120 hpf when 2 bands of equal intensity appear again. The decrease visible at the 64 cells, sphere, and 50% epiboly stages that is not detected for the other splice sites, could nevertheless be due to a lower efficiency of the primer pair.

For *nrxn3a* (fig. 4*C*), at SS1, SS2 bands are present since the early stage of development but only at the 24, 72, and 120 hpf the upper band is expressed.

For SS2, both inserts are present up to the 64-cell stage; at the sphere stage, a form devoid of inserts appears and from stage 8-somite onward only the short size insert is present.

For SS3 and SS4, the insert positive form is present throughout all the stages, with the higher level of expression in the early and larval stages of development. For SS3, the form without inserts is expressed only in the last 2 stages examined, whereas for SS4 these forms are detectable at the stages 50% epiboly, 8-somite, and larval stages.

As in *H. sapiens*, SS5 shows a very complex pattern of splicing. Starting from the earliest stages, we identified 2 major bands corresponding to both the transmembrane and soluble forms. Notably, the splicing pattern becomes very diversified only at the last 2 stages considered.

Finally, in figure 5 and table 6, we summarized all these data. We analyzed the different AS sites in 2 manners. First, each splice site has been considered as a distinct element and in figure 5 we presented in graphical form their global complexity. To this aim, we depicted the genomic exon–intron organization and all the hypothetical different isoforms that could be generated. Moreover, we indicated the forms isolated in our screening and those found in *H. sapiens* by Rowen and colleagues. In the last case, the authors found evidences for 51 different isoforms, that is 70.8% of the 72 total theoretical forms (see table 6); on the contrary, we detected 49 isoforms out of 75 (i.e., 65.3%). Because the 2 percentages are quite similar, the complexity of each splice site is globally maintained.

As a 2nd approach, we analyzed the combinatorial capacity of all the different splice sites in each Neurexin isoform. In table 6, we listed all the possible different isoforms that could be expressed in *D. rerio* and *H. sapiens*, assuming that each splice site is independently regulated from the others. Comparing the total number of different isoforms, it is possible to note that they are nearly identical (2,240 and 2,250 for *D. rerio* and *H. sapiens*, respectively). Nevertheless, the total number of isoforms for each gene differs between the 2 species; in particular, alpha forms of *nrxn2a* and *nrxn3a* are the most different, suggesting a greater evolutive divergence in their AS mechanism.

Potential Paternal Transfer of Specific *nrxn1a* Splicing Isoforms to the Embryo

As previously noted, one peculiar observation for all Neurexin genes has been that they are already expressed and alternatively spliced in the ovary and at the earliest stages of development, a fact suggesting a maternal origin for these transcripts. However, for SS1 of Neurexin 1 at the stage of 1–2 and 64 cells, we detected specific isoforms that were not found in the ovary (fig. 4*A*). To investigate the origin of these isoforms, we compared cDNAs from oocytes and the first stages of development (2 cells and 4–8 cells stages) with the adult testis. The band observed in adult testis and the first developmental stages (indicated by an arrow in fig. 6) was absent in unfertilized oocytes sample, suggesting a paternal origin of some Neurexin transcripts.

In all animals, there is a delay between fertilization and the activation of the zygotic genome, when the embryo relies on gene products present in the egg (maternal transcripts). Although in mammals zygotic genomes seem to be activated at very early developmental stages (2-cell stages) (Pelegri 2003), the zebrafish genome is transcriptionally inactive during the first hours of embryogenesis, and zygotic transcription starts at midblastula transition that, in zebrafish, corresponds to the 512-cell stage (Kane and Kimmel 1993; Santacruz et al. 1997; Bruce et al. 2003). Until this stage, development relies on maternal factors produced during oogenesis that are present in the egg at the time of fertilization (reviewed by Pelegri [2003]). Many studies, however, have reported RNAs in ejaculate spermatozoa and their delivery to the egg during fertilization (reviewed in Miller et al. [2005] and Miller and Ostermeier [2006]). Ostermeier et al. (2004) proved that, during fertilization, human spermatozoa deliver RNAs to the oocyte and they suggested that sperm RNA could contribute to early development; moreover, recent evidence in a mouse model also suggests that paternal RNAs can provide epigenetic marks to the developing embryo that influence the phenotype of the offspring (Rassoulzadegan et al. 2006).

We suggest that the specific band observed in adult testis and the first stages of embryo development could be due to the transfer of specific *nrxn1a* isoforms from mature spermatozoa to the embryo at fertilization. Interestingly, these specific isoforms transferred to the embryo persist until the 64 cells stage (see fig. 4*A*) and they are not degraded after their delivery, suggesting a possible functional role for these Neurexin transcripts. This could be one of the first evidences that these kinds of phenomena are not restricted just to mammals but are present in teleost too. In any case, although further studies in this direction are beyond the aim of this paper, the finding is very intriguing.

Conclusions

In this paper, we characterized the cDNA sequence, the genomic structure, and AS pattern of the complex family of Neurexins genes in the low vertebrate *D. rerio*.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

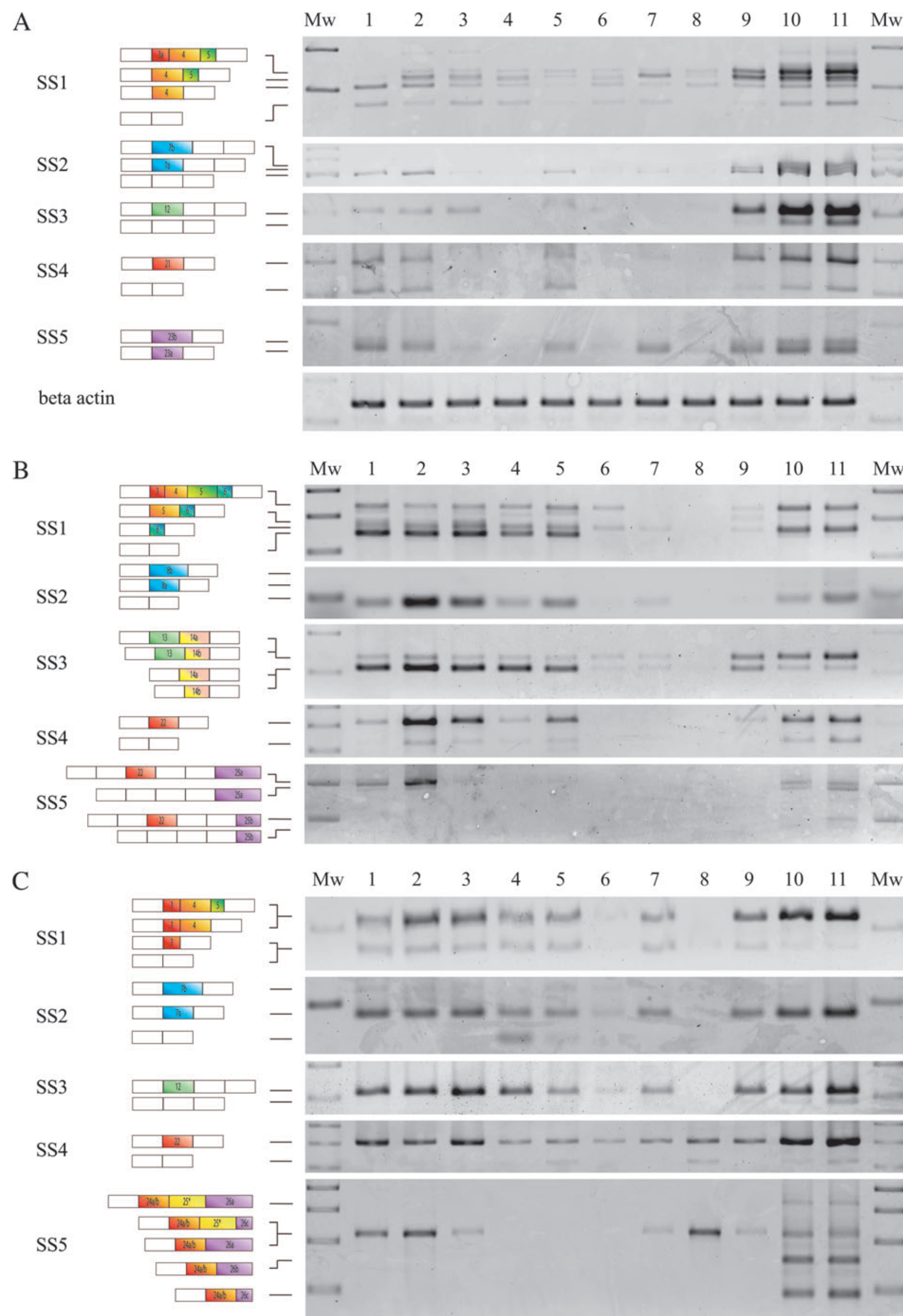

FIG. 4.—Analysis of splicing pattern for *nrxn1a* (*A*), *nrxn2a* (*B*), and *nrxn3a* (*C*) during zebrafish development. RT-PCR products were generated from total RNA extracted from adult ovary and different stages of embryonic development. Panel A shows the amplicons of *nrxn1a* and the β actin used as control, panels B and C show the splicing pattern for *nrxn2a* and *nrxn3a*, respectively. The major isoforms isolated for each splice site (from SS1 to SS5) are depicted with a schematic diagram (open boxes: constitutively spliced exons, filled boxes: alternatively spliced exons). Numbers on top indicate the different stages used: 1, adult ovary; 2, 1–2 cell stage; 3, 64-cell stage; 4, sphere stage; 5, 50% epiboly stage; 6, 80–90% epiboly; 7, 8-somite stage; 8, 15–20 somite stage; 9, 24 hpf; 10, 72 hpf; 11, 5 days; Mw, Molecular weight size marker.

The high degree of protein and gene structure conservation between zebrafish and other mammalian Neurexins suggests a very strong evolutionary pressure and moreover indicates a possible conservation of the functions of these genes in all the vertebrates.

The data presented in this paper, supported by independent data previously published (Zeng et al. 2006), suggest possible new functions (at least in low vertebrates), open new fields of research for these genes, and pose the basis for novel approaches to the functional studies of this complex protein family. Moreover, they suggest that the molecular mechanism of splicing regulation could be partially similar in zebrafish and humans. Finally, we think that they will improve the understanding of the

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

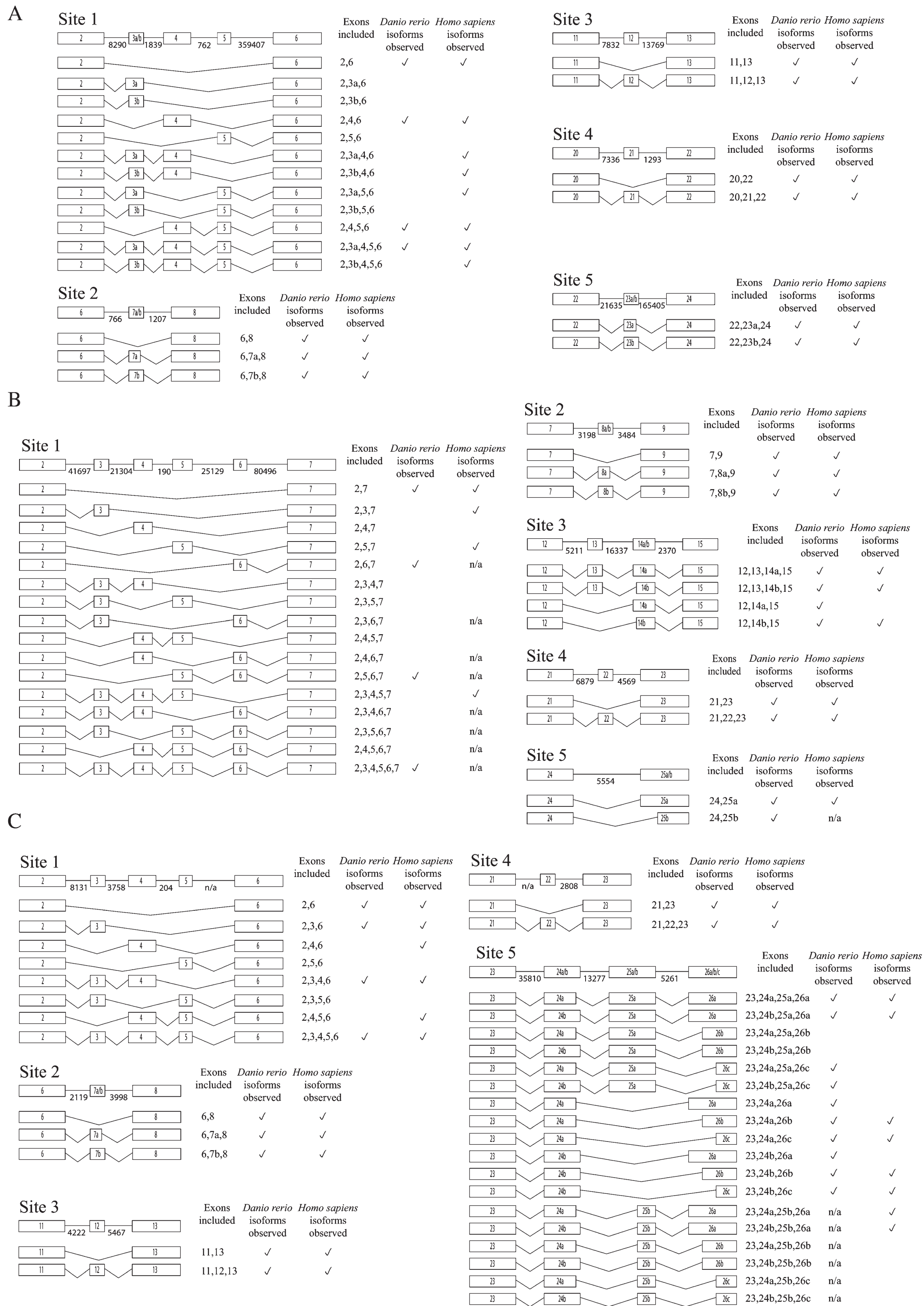

FIG. 5.—AS of *nrxn1a*, *nrxn2a*, and *nrxn3a* genes. Panels A, B, and C show the exon–intron organization of all the possible splice variants at each site of AS of *nrxn1a*, *nrxn2a*, and *nrxn3a*, respectively. On the side of each different variant, we indicated whether it has been found in *Danio rerio* (first column) and in *Homo sapiens* (2nd column). Neither exons nor introns are drawn to scale. Numbers beneath the introns (straight lines) indicate the sizes in bases.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Table 6
AS of Zebrafish Neurexins

	SS1	SS2	SS3	SS4	SS5	Total (SS1 + SS2 + SS3 + SS4 + SS5)	Total (SS1 × SS2 × SS3 × SS4 × SS5)
Neurexin 1a alpha	12 (12)	3 (3)	2 (2)	2 (2)	2 (2)	21 (21)	288 (288)
Neurexin 1a beta	—	—	—	2 (2)	2 (2)	—	4 (4)
Neurexin 2a alpha	16 (8)	3 (3)	4 (4)	2 (2)	2 (1)	27 (18)	768 (192)
Neurexin 2a beta	—	—	—	2 (2)	2 (1)	—	4 (2)
Neurexin 3a alpha	8 (8)	3 (3)	2 (2)	2 (2)	12 (18)	27 (33)	1152 (1,728)
Neurexin 3a beta	—	—	—	2 (2)	12 (18)	—	24 (36)
						75 (72)	2,240 (2,250)

The table lists the number of zebrafish and human (in parentheses) hypothetical variants for each site of AS in Neurexins.

mechanisms of splicing in low vertebrates that are only partially known.

Supplementary Material

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

Acknowledgments

This project was supported by grants from Ministero dell'Istruzione, dell'Universtà e della Ricerca (project FIRB 2001, to M.A. and F.C., PRIN 2004 to F.B and M.A), Regione Piemonte (Ricerca Applicata 2004 and Ricerca Finalizzata 2004 to M.A.) contract LSHM-CT-2003-503254 (www.EVGN.org) to F.B., and the Associazione Italiana per la Ricerca sul Cancro to F.B. We thank Giulio Pavesi, Lorenzo Sangiorgio, Diego Fontaneto, and Roberto Marotta for critically reading the manuscript. We are grateful to Prof Michele De Bortoli and Simona Cimbro for helpful discussions and Patrizia Gilardi for manuscript editing. We gratefully acknowledge the assistance of all the members of the zebrafish laboratory at the Department of Biology at the University of Milano.

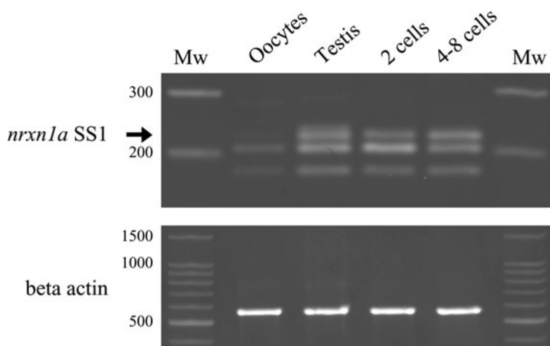

FIG. 6.—Potential paternal delivery of specific *nrxn1a* splicing isoforms to the embryo. Comparison of cDNAs from unfertilized oocytes, adult testis, and early stages of development for *nrxn1a* SS1. Analysis of β actin expression is included as a positive control and it confirms the lack of genomic contamination in RNA preparations (see Materials and Methods). Note that specific messenger RNA splicing variants (arrow) are present in the first stages of development and in adult testis; on the contrary, they are absent in the oocytes sample.

Literature Cited

Amatruda JF, Shepard JL, Stern HM, Zon LI. 2002. Zebrafish as a cancer model system. Cancer Cell. 1:229–231.

Aparicio S, Chapman J, Stupka E, et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science. 297:1301–1310.

Argenton F, Giudici S, Deflorian G, Cimbro S, Cotelli F, Beltrame M. 2004. Ectopic expression and knockdown of a zebrafish sox21 reveal its role as a transcriptional repressor in early development. Mech Dev. 121:131–142.

Ast G. 2004. How did alternative splicing evolve? Nat Rev Genet. 5:773–782.

Biederer T, Sudhof TC. 2001. CASK and protein 4.1 support F-actin nucleation on neurexins. J Biol Chem. 276:47869–47876.

Bruce AE, Howley C, Zhou Y, Vickers SL, Silver LM, King ML, Ho RK. 2003. The maternally expressed zebrafish T-box gene eomesodermin regulates organizer formation. Development. 130:5503–5517.

Butz S, Okamoto M, Sudhof TC. 1998. A tripartite protein complex with the potential to couple synaptic vesicle exocytosis to cell adhesion in brain. Cell. 94:773–782.

Chen E, Ekker SC. 2004. Zebrafish as a genomics research model. Curr Pharm Biotechnol. 5:409–413.

Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM. 2004. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. Cell. 119:1013–1026.

Hata Y, Butz S, Sudhof TC. 1996. CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neurexins. J Neurosci. 16:2488–2494.

Henry MD, Campbell KP. 1999. Dystroglycan inside and out. Curr Opin Cell Biol. 11:602–607.

Hubbard T, Andrews D, Caccamo M, et al. (52 co-authors). 2005. Ensembl 2005. Nucleic Acids Res. 33:D447–D453.

Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, Sudhof TC. 1995. Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell. 81:435–443.

Irie M, Hata Y, Takeuchi M, Ichtchenko K, Toyoda A, Hirao K, Takai Y, Rosahl TW, Sudhof TC. 1997. Binding of neuroligins to PSD-95. Science. 277:1511–1515.

Kan Z, States D, Gish W. 2002. Selecting for functional alternative splices in ESTs. Genome Res. 12:1837–1845.

Kane DA, Kimmel CB. 1993. The zebrafish midblastula transition. Development. 119:447–456.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. Dev Dyn. 203:253–310.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015

Lipscombe D. 2005. Neuronal proteins custom designed by alternative splicing. Curr Opin Neurobiol. 15:358–363.
Miller D, Ostermeier GC. 2006. Towards a better understanding of RNA carriage by ejaculate spermatozoa. Hum Reprod Update. 12:757–767.
Miller D, Ostermeier GC, Krawetz SA. 2005. The controversy, potential and roles of spermatozoal RNA. Trends Mol Med. 11:156–163.
Missler M, Fernandez-Chacon R, Sudhof TC. 1998. The making of neurexins. J Neurochem. 71:1339–1347.
Missler M, Hammer RE, Sudhof TC. 1998. Neurexophilin binding to alpha-neurexins. A single LNS domain functions as an independently folding ligand-binding unit. J Biol Chem. 273:34716–34723.
Missler M, Sudhof TC. 1998. Neurexins: three genes and 1001 products. Trends Genet. 14:20–26.
Missler M, Zhang W, Rohlmann A, Kattenstroth G, Hammer RE, Gottmann K, Sudhof TC. 2003. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature. 423:939–948.
Nguyen T, Sudhof TC. 1997. Binding properties of neuroligin 1 and neurexin 1beta reveal function as heterophilic cell adhesion molecules. J Biol Chem. 272:26032–26039.
Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. 2004. Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature. 429:154.
Parsons MJ, Campos I, Hirst EM, Stemple DL. 2002. Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos. Development. 129:3505–3512.
Patzke H, Ernsberger U. 2000. Expression of neurexin Ialpha splice variants in sympathetic neurons: selective changes during differentiation and in response to neurotrophins. Mol Cell Neurosci. 15:561–572.
Pelegri F. 2003. Maternal factors in zebrafish development. Dev Dyn. 228:535–554.
Petrenko AG, Ullrich B, Missler M, Krasnoperov V, Rosahl TW, Sudhof TC. 1996. Structure and evolution of neurexophilin. J Neurosci. 16:4360–4369.
Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature. 441:469–474.
Rowen L, Young J, Birditt B, et al. (11 co-authors). 2002. Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics. 79:587–597.
Rudenko G, Nguyen T, Chelliah Y, Sudhof TC, Deisenhofer J. 1999. The structure of the ligand-binding domain of neurexin Ibeta: regulation of LNS domain function by alternative splicing. Cell. 99:93–101.
Russell AB, Carlson SS. 1997. Neurexin is expressed on nerves, but not at nerve terminals, in the electric organ. J Neurosci. 17:4734–4743.
Santacruz H, Vriz S, Angelier N. 1997. Molecular characterization of a heat shock cognate cDNA of zebrafish, hsc70, and developmental expression of the corresponding transcripts. Dev Genet. 21:223–233.
Scheiffele P, Fan J, Choih J, Fetter R, Serafini T. 2000. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. Cell. 101:657–669.
Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. Genome Res. 16: 182–189.
Sugita S, Khvochtev M, Sudhof TC. 1999. Neurexins are functional alpha-latrotoxin receptors. Neuron. 22:489–496.
Sugita S, Saito F, Tang J, Satz J, Campbell K, Sudhof TC. 2001. A stoichiometric complex of neurexins and dystroglycan in brain. J Cell Biol. 154:435–445.
Tabuchi K, Sudhof TC. 2002. Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. Genomics. 79:849–859.
Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. 13:382–390.
Ullrich B, Ushkaryov YA, Sudhof TC. 1995. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron. 14:497–507.
Ushkaryov YA, Hata Y, Ichtchenko K, Moomaw C, Afendis S, Slaughter CA, Sudhof TC. 1994. Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences in receptor-like neuronal cell-surface proteins. J Biol Chem. 269:11987–11992.
Ushkaryov YA, Petrenko AG, Geppert M, Sudhof TC. 1992. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science. 257:50–56.
Ushkaryov YA, Sudhof TC. 1993. Neurexin III alpha: extensive alternative splicing generates membrane-bound and soluble forms. Proc Natl Acad Sci USA. 90:6410–6414.
Westerfield M. 2000. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed. Eugene: University of Oregon Press.
Xing Y, Lee C. 2005. Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci USA. 102:13526–13531.
Xing Y, Lee C. 2006. Alternative splicing and RNA selection pressure—evolutionary consequences for eukaryotic genomes. Nat Rev Genet. 7:499–509.
Zdobnov EM, von Mering C, Letunic I, et al. (36 co-authors). 2002. Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science. 298:149–159.
Zeng Z, Sharpe CR, Simons JP, Gorecki DC. 2006. The expression and alternative splicing of alpha-neurexins during Xenopus development. Int J Dev Biol. 50:39–46.

Takashi Gojobori, Associate Editor

Accepted October 10, 2006

Downloaded from http://mbe.oxfordjournals.org/ at Ernst Mayr Library of the Museum Comp Zoology, Harvard University on April 28, 2015