Gene Expression Patterns 17 (2015) 69–78

Contents lists available at ScienceDirect

Gene Expression Patterns

journal homepage: www.elsevier.com/locate/gep

Analysis of splice variants for the *C. elegans* orthologue of human neuroligin reveals a developmentally regulated transcript

Fernando Calahorro *, Lindy Holden-Dye, Vincent O'Connor

Centre for Biological Sciences, Life Sciences Building 85, University of Southampton, Southampton SO17 1BJ, UK

ARTICLE INFO

Article history:
Received 3 December 2014
Received in revised form 17 February 2015
Accepted 19 February 2015
Available online 26 February 2015

Keywords:
Neuroligin
Neurexin
Splicing
Development
Synapse
C. elegans

ABSTRACT

Neuroligins are synaptic adhesion molecules and important determinants of synaptic function. They are expressed at postsynaptic sites and involved in synaptic organization through key extracellular and intracellular protein interactions. They undergo trans-synaptic interaction with presynaptic neurexins. Distinct neuroligins use differences in their intracellular domains to selectively recruit synaptic scaffolds and this plays an important role in how they encode specialization of synaptic function. Several levels of regulation including gene expression, splicing, protein translation and processing regulate the expression of neuroligin function. We have used *in silico* and cDNA analyses to investigate the mRNA splicing of the *Caenorhabditis elegans* orthologue *nlg-1*. Transcript analysis highlights the potential for gene regulation with respect to both temporal expression and splicing. We found *nlg-1* splice variants with all the predicted exons are a minor species relative to major splice variants lacking exons 13 and 14, or 14 alone. These major alternatively spliced variants change the intracellular domain of the gene product NLG-1. Interestingly, exon 14 encodes a cassette with two distinct potential functional domains. One is a polyproline SH3 binding domain and the other has homology to a region encoding the binding site for the scaffolding protein gephyrin in mammalian neuroligins. This suggests differential splicing impacts on NLG-1 competence to recruit intracellular binding partners. This may have developmental relevance as *nlg-1* exon 14 containing transcripts are selectively expressed in L2–L3 larvae. These results highlight a developmental regulation of *C. elegans nlg-1* that could play a key role in the assembly of synaptic protein complexes during the early stages of nervous system development.

Intercellular communication between neurons is mediated by the alignment of pre and postsynaptic elements. This involves trans-synaptic interactions between cell surface molecules and selective recruitment of signalling scaffolds through the intracellular domains of such adhesion molecules. The molecular architecture generates a synaptic 'code', in which different classes of cell-type-specific contact molecules drive the maturation, stabilization and signalling of synapses. An important class of the synaptic adhesion proteins that help define the synaptic code are the neuroligins (Yamagata et al., 2003).

Five neuroligin genes exist in human (*NLGN-1*, *-2*, *-3*, *-4X*, *-4Y*) and they are expressed in different neuronal classes (Budreck and Scheiffele, 2007; Ichtchenko et al., 1996). They interact with the cell adhesion proteins, the neurexins, and other ligands at the synapse (Bourne and Marchot, 2014) producing interactions that contribute to differentiation, plasticity and specificity of synapses (Craig and Kang, 2007; Dean and Dresbach, 2006; Varoqueaux et al., 2004). Rare variants in the human genes encoding postsynaptic neuroligin-3 and 4 (*NLGN-3* and *-4*), and their presynaptic binding partner neurexin-1 (*NRXN-1*) have been linked with autism spectrum disorders (ASD) (Sudhof, 2008). The human neurexin and neuroligin genes are orthologous to the *Caenorhabditis elegans* genes *nrx-1* and *nlg-1*, respectively (Calahorro, 2014).

nlg-1 is expressed throughout the nervous system of *C. elegans* and the NLG-1 protein is located at postsynaptic sites of the body wall muscle, although it is also expressed presynaptically in some neurons (Feinberg et al., 2008). NLG-1 from *C. elegans* and the mammalian neuroligins present significant homology in terms of their amino acid sequences. Neuroligin in both human and the nematode harbour the same functional domains encompassing an extracellular cholinesterase-like domain, a type-1 transmembrane sequence and a short intracellular domain ending in a PDZ-binding motif (Calahorro and Ruiz-Rubio, 2012).

Cell type-specific gene expression programs lead to huge diversity in neuronal properties. Furthermore, alternative splicing greatly amplifies the coding capacity of the genome, providing a powerful

Abbreviations: ASD, autism spectrum disorders; SS, splice site; SL, splice leader; SH3, Src-Homology 3; EGF, epidermal growth factor; LNS, laminin/neurexin/sex-hormone-binding; nt, nucleotide; aa, amino acid.

* Corresponding author. Centre for Biological Sciences, Life Sciences Building 85, University of Southampton, Southampton SO17 1BJ, UK. Tel.: +44 23 8059 9006; fax: +44 23 8059 4459.
E-mail address: F.Calahorro@soton.ac.uk (F. Calahorro).

http://dx.doi.org/10.1016/j.gep.2015.02.002
1567-133X/

mechanism to control functional diversity. Splicing is particularly well detailed for the neurexin gene family where both cell and isoform specific splicing occur. This capacity is enhanced by activity-dependent alternative splicing which broadly acts to mediate dynamic modification of neuronal function (Treutlein et al., 2014). This generates specific splice variants modifying the signalling properties or the function of synaptic proteins (An and Grabowski, 2007; Li et al., 2007). The splicing patterns of neuroligins are less well characterized than neurexin. However, the gene organization suggests there is a splice regulation of their trans-synaptic signalling potential (Lee et al., 2010).

The reported splicing of mammalian neuroligins show they are differentially spliced at one position (splice site A, SSA) in exon 2 or 3 that encodes the beginning of the extracellular domains. Additionally, *NLGN-1* presents another splice position (splice site B, SSB) in exon 5 that encodes the central region of the acetylcholinesterase domain (Calahorro, 2014; Ichtchenko et al., 1995, 1996). Functional impact of these differences is indicated by *in vitro* observations that *NLGN-1* lacking SSB induce a rapid presynaptic differentiation of approaching neurons relative to the full length *NLGN-1*. This is due to the removal of the N-linked glycosylation site within insert B (Lee et al., 2010). Furthermore, splice variants in neuroligins have been associated with autism spectrum disorder (ASD). Specifically a *NLGN-3* transcript variant (without exon 7) that is selectively expressed in ASD cohorts is predicted to harbour a premature termination after exon 6 and results in a truncated protein (Talebizadeh et al., 2006). The truncated NLGN-3 product may have a regulatory role in the activity of neuroligins by attenuating the function of the full-length isoform, impacting on the balance between excitation and inhibition in affected neural networks. In addition, a novel *NLGN-4* isoform lacking exon 4 has been described that leads to an in-frame exclusion of 62 amino acids. However, the functional significance of removing this extracellular region of NLGN-4 is unknown (Talebizadeh et al., 2006).

Here, we describe different major splice variants in the gene encoding the only predicted *C. elegans* neuroligin, *nlg-1*. The *nlg-1* transcript variants, in contrast to the situation described for the mammalian genes, involve exons encoding the intracellular domain. In particular the major variants are predicted to encode proteins that will lack or contain a candidate polyproline SH3 binding domain and the homologous region in the mammalian neuroligins harbours the binding site for gephyrin. This is pertinent because this kind of protein scaffolding domain has been implicated in dynamic regulation of neuroligin dependent synaptic differentiation of inhibitory synapses in mammals (Soykan et al., 2014). Interestingly, we observed dynamic expression of these *nlg-1* variants during the development of *C. elegans*. This suggests alternative splicing might play a role in regulating neuroligin scaffolding during network assembly of the *C. elegans* nervous system.

1. Results

1.1. Phylogenetic analysis of neuroligins

The predicted neuroligin protein sequences of different invertebrates and mammals were used to construct a phylogenetic tree (Fig. 1A). This analysis revealed that the *C. elegans* NLG-1 is closely related to invertebrate neuroligins from *Apis mellifera*, *Drosophila melanogaster* and *Aplysia californica*. The vertebrate and mammalian neuroligins are evolutionarily more distant from the *C. elegans* neuroligin. However, the core protein including the cholinesterase-like domain and O-linked and N-linked glycosylation sites are conserved between mammals and the nematode (Calahorro, 2014). The coding sequences (CDS) and amino acid alignment of human and *C. elegans* neuroligins showed identities of 46.83–53.61% (CDS) and 23.97–25.50% (aas), respectively (Supplementary Table 1). The comparison of the exon–intron boundaries of *C. elegans* and human neuroligins revealed that the *nlg-1* transcript contains 16 exons as compared to mammalian neuroligins which have 7 (*NLGN-1*, *NLGN-2*, *NLGN-4X*) or 8 (*NLGN-3*) exons (Fig. 1B). The predicted intracellular region of NLG-1 is encompassed and encoded by exons 13–16. This is in contrast to the human neuroligins where this region is encoded by the final exon of the respective neuroligin (*i.e.* exon 7 or 8). The expression of human *NLGN-1* and/or rat *Nlgn-1* cDNAs under the *C. elegans nlg-1* promoter partially rescues abnormal sensory and locomotory behaviours in *nlg-1*-deficient mutants (Calahorro and Ruiz-Rubio, 2012; Izquierdo et al., 2013). This suggests that despite limited amino acid identity the heterologous expressed protein retains key neuroligin function conserved between nematode and mammal.

1.2. nlg-1 gene organization and splice variants

We catalogued the major splice variants of transcripts that encode NLG-1 isoforms by PCR utilizing cDNA from young adult worms (L4 + 1) as a template. We designed primers covering blocks of exons within the *nlg-1* transcript to amplify short sequences (between 0.5 and 1.2 kb), which encompassed contiguous sequences encoding the entire cDNA length. We first investigated the ill-defined 5′ end of the cDNA by amplifying the 5′ end of the gene. This made use of a sense primer for SL-1 combined with an antisense primer from exon 5 that amplified a 682 base pair fragment. In contrast performing the same PCR in which the SL-1 primer was replaced with SL-2 primer failed to amplify a product. This SL-1 amplified product was extracted from the gel and subjected to sequencing (Fig. 2Ai). Sequencing this major amplification product confirmed the extension of exon 1 by the presence of untranslated sequence encoded by a 71 nt extension from the predicted start AUG (accession: KP331805). There was no heterogeneity in the three clones that were sequenced and the lower ghost band was an unspecific product corresponding to another SL-1 trans-spliced gene (data not shown). Thus, the nature of the amplification and the sequencing suggests little evidence for alternative splicing between exon 1 and exon 5 of the SL-1 trans spliced neuroligin gene (Fig. 2Ai).

The PCR amplification for the second transcript block encoded between exon 6 and exon 11 identified a single major band 1115 bp. Sequences from this PCR identified that the amplifications were as would be predicted from the routine splicing of exons 6–11. This shows that the *nlg-1* cDNA encoding the central region of the acetylcholinesterase domain does not undergo alternative splicing (Fig. 2Aii). There was no heterogeneity in the three clones that were amplified and sequenced.

Finally, the amplification of the 3′ end of the gene with primers from exons 11 and 16 generated a complex set of products. Although a 593 bp band that defines this region of the cDNA when encoded by all predicted exons was identified, the major amplification product was at 455 bp. The sequence analysis of the distinct cDNAs amplified in this PCR revealed the variation was generated from three alternative splice variants (Fig. 2Aiii), and two minor but detectable amplifications encoded by sequences unrelated to the *nlg-1* cDNA (Fig. 2Aiii asterisked and data not shown). This region coded by exons 11–16 encompasses the transmembrane and intracellular domains in the NLG-1 protein. Of the *nlg-1* related amplifications, the dominant band corresponded to a variant in which exon 14 is alternatively skipped (accession: KP331806). Above this band we identified a cDNA that contained all of the predicted exons but this transcript contained a 5 nt sequence corresponding to an isoform in the NCBI database (NM_001270297) (accession: KP331808). In this isoform, the 5 nt sequence AACAG encodes a Glu and Gln residues while in the isoform containing the 5 nt sequence CACGT the residues encoded are Ala and Arg (Fig. 2Aiii). Finally below the major band, that lacks exon 14, we identified a variant in which both exons 13 and 14 were alternatively skipped