ELSEVIER

Gene 273 (2001) 115–122

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

The complete sequence of the human locus for NgCAM-related cell adhesion molecule reveals a novel alternative exon in chick and man and conserved genomic organization for the L1 subfamily

Kate Dry[a,1], Sue Kenwrick[a], André Rosenthal[b,2], Matthias Platzer[c,*]

[a]*Wellcome Trust Centre for Molecular Mechanisms of Disease and University of Cambridge Department of Medicine, Addenbrooke's Hospital, Cambridge CB2 2XY, UK*

[b]*Friedrich-Schiller-Universität, Schoßgasse 1, 07743 Jena, Germany*

[c]*Institut für Molekulare Biotechnologie, Abteilung Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany*

Received 20 December 2000; received in revised form 21 March 2001; accepted 20 April 2001
Received by M. D'Urso

Abstract

NrCAM is one member of the L1 subfamily of cell surface recognition molecules implicated in nervous system development and function. Here we report the complete sequence of the human *NRCAM* locus. The gene comprises 34 exons and shows extensive conservation of exon/intron structure compared to *L1*, suggesting a common evolutionary ancestor. By human–chick sequence comparison we identified exons not previously found in mammalian *NRCAM* mRNAs. One of these encodes a premature stop codon that would give rise to an isoform of NRCAM lacking ankyrin-binding capacity. The availability of the complete sequence will allow an investigation of the potential role of these splice variants, and examination of the regulatory elements controlling *NRCAM* expression as well as the relationship of *NRCAM* to disorders involving 7q.

Keywords: Neural cell adhesion molecule; Gene structure

1. Introduction

Neural cell adhesion molecules (CAMs) of the immunoglobulin (Ig) superfamily are important mediators of nervous system development with multiple functions as guidance cues and receptors for neurons and their axons (reviewed in Brümmendorf and Rathjen, 1996). The superfamily consists of proteins with Ig domains only and those with both Ig and fibronectin type III (FNIII)-like domains in their extracellular regions. Members of the Ig superfamily can be divided into subgroups based on similarities of both structure and sequence. For instance, L1, NgCAM (neuroglia cell adhesion molecule), NrCAM (NgCAM-related cell adhesion molecule), CHL1 (close homologue of L1), neuroglian and neurofascin/ABGP (ankyrin binding glycoprotein) form one group of vertebrate CAMs (Brümmendorf et al., 1998). These cell surface glycoproteins share an overall structure of six Ig domains, five FNIII domains, a single pass, transmembrane domain and a very conserved cytoplasmic tail. At the cell surface they undergo a bewildering variety of homophilic and heterophilic interactions with members of the same family as well as with other classes of CAMs and extracellular matrix proteins. Moreover, they have overlapping repertoires of extracellular ligands as well as intracellular cytoskeletal connections. Their net affect on neuronal migration or neurite/axon outgrowth may therefore in part be a function of which particular Ig superfamily CAMs and their ligands are present at the neuronal surface in combination.

Mutations in the gene for one member of this subfamily, L1, are responsible for an X-linked neurological disorder that involves abnormalities of axonal growth, testifying to the importance of this family in nervous system development (Rosenthal et al., 1992; reviewed in Fransen et al., 1997; Kenwrick et al., 2000). The genes for all other mammalian

Abbreviations: bp, base pairs; cDNA, DNA complementary to RNA; CAM, cell adhesion molecule; hnRNA, heterogeneous nuclear RNA; Ig, immunoglobulin; FNIII, fibronectin type III; kb, kilobases; mRNA, messenger RNA; PCR, polymerase chain reaction; UTR, untranslated region

* Corresponding author. Tel.: +49-3641-656241; fax: +49-3641-656255.

E-mail address: mplatzer@imb-jena.de (M. Platzer).

[1] Present address: Proteom Limited, Babraham Hall, Babraham CB2 4AT, UK.

[2] Present address: metaGen Gesellschaft für Genomforschung mbH, Ihnestrasse 63, 14195 Berlin, Germany.

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00493-0

members of this subgroup of CAMs, CHL1, NrCAM and ABGP map to different autosomal locations and have not to date been associated with a Mendelian disorder. The gene for human NrCAM (*NRCAM*) lies in 7q in a region that is deleted in a variety of tumors (Lane et al., 1996; Wang et al., 1998). Unlike L1, *NrCAM* is subject to extensive alternative splicing and has the potential to produce a large number of isoforms (Grumet et al., 1991; Kayyem et al., 1992; Lane et al., 1996; Grumet, 1997). These isoforms may have tissue- or cell-specific functions as *NRCAM* is also expressed at high levels at specific sites outside the nervous system (Wang et al., 1998). Furthermore, exon utilization is related to the tissue of expression. In order to facilitate examination of *NRCAM* in disorders that map to 7q and investigate the basis for tissue-specific gene expression and alternative splicing, we have determined the complete structure of the *NRCAM* locus. Comparison of genomic with cDNA sequences has revealed a conserved splicing event that would result in a truncated version of the protein. This would have the potential to act as a negative regulator of intracellular NRCAM function. The presence of an exon that has the capacity to direct intracellular trafficking of L1-like CAMs to the neuronal growth cone is also found within the *NRCAM* gene, although this exon is rarely included in neural *NRCAM* mRNA. The organization of the gene is compared to that of close relative L1 and similarities and differences are discussed.

2. Materials and methods

2.1. Library screening and clone characterization

Probes were PCR-amplified from human brain cDNA using primers flanking Ig domains 3 and 4 (primers 19 and 20; Table 1) or FNIII domains 2 and 3 (primers 21 and 23). The RPC111 human male BAC library (Lawrence Livermore National Laboratory) and ICRF chromosome 7-specific cosmid libraries were screened by Steve Scherer and Jack Huizenga (Hospital for Sick Children, Toronto, Canada). Five BAC (140C21, 267017, 57N23 and 54P11) and five cosmid (87H8, 214M2, 148C10, 214G1 and 86D6) positive clones were pulled out. Both STS analysis and restriction enzyme fingerprinting verified clone overlap. For the former, five different primer pairs were used to amplify sections of the *NRCAM* gene (Table 1). These included products from both the 5′ and 3′ ends of the gene. No *NRCAM* intragenic deletions were observed.

For fingerprinting, clones were digested with either *Bam*HI or *Hind*III. Fragments (0.5–25 kb) were resolved on a CHEF DRII pulsed-field gel electrophoresis apparatus (1% agarose with switch times of 1–6 s for 11 h at 6 V) and visualized by ethidium bromide staining. Substantial overlap was observed for all clones.

2.2. Genomic sequencing and analysis

BACs were sequenced by the ‘shotgun method’ as described previously (Platzer et al., 1997). Sequence reactions were electrophoresed on ABI 377 sequencers. Homology searches in public domain databases were performed using BLAST, version 1.4 (Altschul et al., 1990) and FASTA, version 2.0 (Pearson and Lipman, 1998). G + C content and distribution were calculated using GCG, version 9.0 (Genetics Computer Group, Inc.). Criteria for identification of CpG islands were G + C > 50%, ratio of CpG observed/expected (o/e) > 0.6, and length > 200 bp CpG island finder (Gardiner-Garden and Frommer, 1987). Genome-wide repeats were identified by RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker).

Table 1
NRCAM primer sequences

Primer	5′ to 3′	Location[a]	Location
7	CTG TCT GAA GGA GTT CGA CTT C	3716–3695	Exon 34
12	tga gga tat ttt cct gta ggt c		Intron 29
13	CTC CAT TCT TCT TTG Cct atg g	3314–intron	Exon 30/intron 29
19	CAT GGA ATT CCA GCA GAA GCA ACC TAT TTC TG	670–691	Exon 9
20	ACA CCT CGA GGA TTC GTG GTA GCT CAG CC	1356–1337	Exon 15
21	AAT GGA ATT CGA ATT TAA TGG GCC AGG C	2305–2322	Exon 23
23	CAT CCT CGA GAG AGT CCA GTG TTG GAT TC	2886–2868	Exon 26
35	CCG CTC GAG AAG TTC TTA TAA ACT GTC CTG	947–927	Exon 12
36	CCG CTC GAG ACC TCG GCT AAG GTA CTG TTC	2576–2556	Exon 24
53	AAA GTT CCC CGC ATG AAA AT		5′ UTR
54	CCT GCT GAG ACT CAC ACA CTG		5′ UTR
55	TTA AGC TCA AAG CCA ATA TTC C		3′ UTR
56	TGT TCA TAG TAT GAG AGG GCT G		3′ UTR
67	CAG GGC TGG TTC ATT GGT CTG	3484–3504	Exon 31
68	GAA TAC AGg ctg ttt tca ttt g	3655–intron	Exon 32/33a
72	TTT CCT TCA GCC GGC TCT TTC	3848–3828	Exon 34
73	GGA CAT TTG GAG AAT ACA GGT CT	4073–4095	Exon 32/33

[a] Locations of coding sequence of primers relative to the sequence of AJ001057.

Aligning mRNA sequence to genomic sequence using SIM4 (Florea et al., 1998) identified exon boundaries. Furthermore, the program GCG DIVERGE (Li et al., 1985) was used for estimation of evolutionary distances.

2.3. Analysis of alternative splicing of ASCYT1 and ASCYT2

Poly A^+ brain RNA from a 14-week-old human fetus was reverse-transcribed using super RT reverse transcriptase (HT Biotechnology Ltd.) as described in Wang et al. (1998). Primers 67 and 72 were used to amplify from exon 31 through to exon 34 with Amplitaq Gold (Roche) in a 50 μl reaction and an annealing temperature of 55°C. Product (8 μl) was visualized on agarose gels. For nested PCR reactions with primers 68 and 7 or 73 and 72, 2 μl of the primary 67–72 reaction was used as a template for a new reaction carried out with annealing at 61°C. Sequencing was carried out using the ABI Prism® cycle sequencing kit and end primers.

3. Results and discussion

3.1. The human NRCAM locus consists of 34 exons spanning over 316 kb

Human male BAC and chromosome 7 specific cosmid libraries were screened with cDNA probes representing Ig domains 3–4 and FNIII domains 2–3 of NRCAM. Five BAC and five cosmid clones were identified. *Eco*RI digestion followed by Southern hybridization indicated that two out of five cosmids (214G1 and 214M2) and five out of five BACs contained both regions of the gene. The remaining three cosmid clones contained a sequence encoding only the Ig domains. To determine the genomic content of each clone in more detail, PCR analysis was performed using a range of primer pairs spanning the *NRCAM* cDNA sequence. Three human cDNA sequences have been described (Accession number U55285, 4134 bp, Lane et al., 1996; AB002341, 6218 bp, Nagase et al., 1997; AJ001057, 3900 bp, Wang et al., 1998) that differ in the amounts of 5′ and 3′ untranslated sequence reported. Two potential initiating methionines are present, 15 bp apart. PCR analysis showed that four out of five BACs (not clone 140C21) appeared to contain the complete coding sequence, 3′ untranslated region (UTR) and the 5′ end of the gene as described by Lane (U55285) but not the additional upstream 5′ UTR described by Nagase (AJ0010557). Two of the cosmid clones contained the same 5′ end of *NRCAM* (87H8 and 86D2) but none extended beyond FNIII domain V. Two of the BAC clones were therefore selected for sequencing (BACs 267O17 and 256L2).

The two BAC clones were sequenced generating a contig of 227,054 bp (Accession number AF172277), with 102 kb of this contig (position 46,609 to 149,112) covered by the

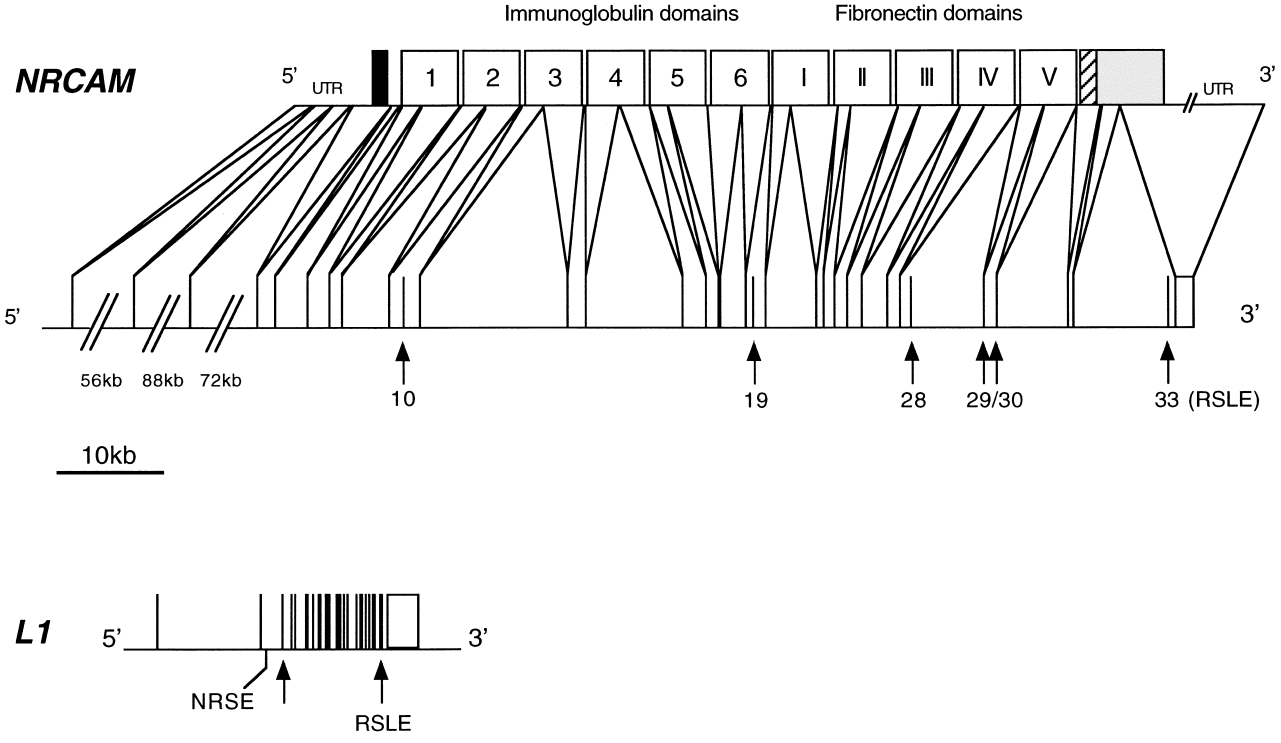

Fig. 1. The exon/intron organization of human *NRCAM* compared to that of *L1*. *NRCAM* exons are shown as vertical lines and mapped onto a cartoon of Ig and FNIII domains (white boxes). Exons encoding sequence N-terminal to Ig domains, the transmembrane region and cytoplasmic domain are mapped onto black, hatched and gray boxes, respectively. Exons belonging to 5′ and 3′ UTRs are mapped onto the line. Exons subject to alternative splicing are shown with an arrow.

overlap of both BACs. Sequence analysis revealed 31 *NRCAM* exons spanning 99 kb of genomic DNA including the entire coding region and 3′ end of the gene (Fig. 1) The exons range in size from 12 to 2499 bp in length separated by introns ranging in size from 90 bp to more than 14 kb (Table 2). The initiator methionine was found within a 212 bp exon preceded by a consensus splice acceptor sequence and containing 106 untranslated bases. This suggested that additional upstream exons existed although none could be found in the 46 kb upstream of this exon in BAC clone BAC267O17.

To determine if an additional 5′ *NRCAM* sequence was available in existing databases these were screened with the longer *NRCAM* cDNA sequence described by Nagase et al. (1997). This revealed two BAC clones from 7q31 submitted by the Genome Sequencing Center of Washington University School of Medicine (Accession numbers AC005058 and AC007567) extending our sequence contig for 287,290 bp. This region was found to contain the additional 5′ untranslated sequence described in the longest cDNA entry. Interestingly, the remaining 5′ UTR sequence was distributed over three additional exons separated by very large introns from each other and the first coding exon (57, 88 and 73 kb). Thus, *NRCAM* contains 34 exons distributed over 316,572 bp. The first exon is located within a pronounced CpG island, with 70.9% GC and 205 CpG dinucleotides. Moreover, a second CpG island lies 15 kb 5′ to the first coding exon (846 bp, 70.2% GC, 66 CpGs, o/e = 0.74). This element may not be functional, as it is part of an LTR-related genome-wide repeat. In addition, database searches revealed a third WashU BAC clone (Accession number AC005683) derived from the *NRCAM* locus. Although not annotated as containing the *NRCAM* gene this clone was found to contain exons 14–34 in agreement with our data.

3.2. NRCAM and L1 have a partially conserved genomic structure

Our laboratories have previously described the complete gene sequence for human L1 (Coutelle et al., 1998). The

Table 2
NRCAM and *L1* exons are of comparable size

Domain	*NRCAM* exon number	*NRCAM* exon size (bp)[a]	L1 exons (bp)[a]	Intron phase[b]	*NRCAM* intron (bp)
	1	84	125	–	56,649
	2	158		–	88,150
	3	67		–	72,524
	4	106/**106**	108/**76**	**1**	2189
	5	**18**	**15**	**1**	3064
Ig 1	6	**106**	**106**	**2**	2060
Ig 1	7	**197**	**203**	**1**	1172
Ig 2	8	**123**	**123**	**1**	4652
Ig 2	9	**171**	**171**	**1**	506
	10, **AE19**	**57**		1	1808
Ig 3	11	**112**	**112**	**2**	14,119
Ig 3	12	**185**	**185**	**1**	1759
Ig 4	13	**132**	**132**	**1**	9427
Ig 4	14	**144**	**144**	**1**	2083
Ig 5	15	**112**	**112**	**2**	1341
Ig 5	16	**167**	**167**	**1**	90
Ig 6	17	**148**	**157**	**2**	2170
Ig 6	18	**125**	**125**	**1**	445
	19, **AE10**	**30**		1	1507
FNIII 1	20	**102**	**111**	**1**	11,106
FNIII 1	21	**198**	**198**	**1**	105
FNIII 2	22	**71**	**71**	**0**	1328
FNIII 2	23	**226**	**223**	**1**	757
FNIII 3	24	**116**	**116**	**0**	1397
FNIII 3	25	**205**	**202**	**1**	2110
FNIII 4	26	**123**	**123**	**1**	1417
FNIII 4	27	**177**	**174**	**1**	1072
	28, **AE12**	**36**		1	6919
FNIII 5	29, **AE93**	**126**	**120**	**1**	1203
FNIII 5	30, **AE93**	**153**	**156**	**1**	6716
	31	**132**	**135**	**1**	819
	32	**79**	**73**	**2**	8757; 8710
	33, **ASCYT2**; 33a, **ASCYT1**	**12**; **34**/25	**12**	**2**; –	547
	34	**238**/2261; 2499	**232**/1145	–	

[a] Bold, coding.
[b] Bold, conserved between *NRCAM* and *L1CAM*.

current study offers the potential for direct comparison of gene organization for these two members of the L1 subfamily. NrCAM and L1 have six Ig domains and five fibronectin (FNIII) domains in their extracellular domains. Each of these domains is coded for by two exons, a feature of genomic organization that has also been described for other Ig superfamily CAMs including F11 (Plagge and Brümmendorf, 1997). Comparison of the protein domain structure with the exon/intron structure of the genes reveals that there are two types of intron. For both *L1* and *NRCAM* the introns that lie between the pairs of exons that encode domains are of phase 1, i.e. are inserted after the first nucleotide of the interrupted codon. By contrast, the introns that split these pairs of exons can be of any phase. This was also found to be true for *F11* and has led to the suggestion that these genes evolved by duplication of whole-domain rather than half-domain ancestry of Ig-like domains. For *L1* and *NRCAM* the phase of 'intradomain' introns is absolutely conserved.

Comparison of the genomic structure of these genes illustrates that the *NRCAM* gene is much larger than *L1*. The 28 coding exons of *L1* are contained within only 15 kb with an additional non-coding exon 10 kb upstream of the start codon. However, all of the exons of *NRCAM* that code for structural domains are almost identical in size to those of *L1* (Table 2). *NRCAM* has three additional small exons (10, 19 and 28) but the size difference between the genes is primarily accounted for by much larger introns across the whole locus. The conservation of exon size and intron phase observed for these two loci indicates that they arose from a common ancestral gene. In order to date the duplication event, the coding sequences (CDS) of human *NRCAM* (Accession number AF172277) and *L1CAM* (Accession number U52112) were compared by the program DIVERGE using mouse *L1cam* (Accession number 133093) as a reference. Assuming a divergence time of primates and rodents of about 80 Myr B.P. estimates of the duplication time ranging from 330 Myr (synonymous sites) to 840 Myr (non-synonymous sites) were observed. Whether subsequent divergence in gene structure primarily involved loss of exons and intronic material from a large gene similar to *NRCAM* or the addition of DNA to a more *L1*-like gene cannot be ascertained. Resolution of the structure of other L1 family members in different species may help answer this question. Interestingly, a similar relationship exists between the genes for two other Ig superfamily CAMs, axonin-1 and F11. These chick proteins share a domain structure and GPI linkage to the plasma membrane. Although these genes have conserved intronic positions relative to coding sequences, the sizes of the introns are invariably larger for *F11* so that the locus is over twice the size of *axonin-1* (Kozlov et al., 1995).

3.3. Alternative splicing of NRCAM hnRNA involves whole exons and a cryptic donor site in intron 32

Determination of the genomic structure of *NRCAM* has allowed a number of questions concerning alternative splicing to be resolved. *NRCAM* hnRNA is subject to extensive alternative processing. Six alternatively-spliced regions have been described in chick: AE19 encodes a 19 amino acid section between Ig domains 2 and 3; AE10 encodes a ten amino acid section between Ig domain 6 and FNIII domain 1; AE12 encodes a 12 amino acid section between FNIII domains 4 and 5; AE93 encodes a 93 amino acid section corresponding to the whole of FNIII domain 5; ASCYT1 and ASCYT2 (AE4) affect the cytoplasmic domain (Grumet et al., 1991; Kayyem et al., 1992; Grumet, 1997).

Four of these alternatively-spliced regions, AE19, AE10, AE12 and AE93, have also been described for human *NRCAM* (Lane et al., 1996; Wang et al., 1998). The four alternatively-spliced regions described in human and chick are coded for by complete exons: AE19 (exon 10), AE10 (exon 19), AE12 (exon 28) and AE93 (exons 29 and 30). With the exception of AE93, which affects a whole domain, all of these variations in splicing alter the length of peptides between folded Ig or FNIII domains. These exons may therefore play a role in the tertiary structure of the protein or its flexibility. Interestingly, in a related insect protein hemolin, the four Ig domains comprising the extracellular part of the molecule can fold into a horseshoe structure due to the presence of a seven residue spacer between Ig domains 2 and 3 (Su et al., 1998). Thus, the presence or absence of interdomain amino acids may radically affect the presentation of the sections of mature protein. Alternatively, the presence or absence of these regions may directly alter the ligand binding potential of the mature protein by contributing to binding sites.

The remaining two regions contributing to splicing variation in chick, ASCYT1 and ASCYT2 (AE4), have previously only been observed there (Kayyem et al., 1992) but not in mammals. Out of five cDNA clones isolated from a chick cerebellum cDNA library, one clone contained only ASCYT2 and another contained both ASCYT1 and ASCYT2. Inclusion of ASCYT1 in mRNA produced a frame shift that resulted in a stop codon within the cytoplasmic region of chick NrCAM. As only the cDNA sequence was available it was impossible to determine whether this represented a true variant or incomplete processing of chick mRNA.

We aligned the chick ASCYT1 and ASCYT2 sequences with the human *NRCAM* genomic sequence. ASCYT1 matched bases 144,838–144,884 (Accession number AF172277) with 92.9% identity over 47 bp. There is an acceptor site at the 5′ boundary but no donor site at the 3′ end of this conserved sequence. The ASCYT2 sequence is contiguous with ASCYT1 in the genomic sequence and shows 100% identity over 12 bp (bases 144,885–144,896) immediately 3′ to the ASCYT1 match. Both acceptor and donor sites are present in the human genomic sequence flanking ASCYT2 and it has therefore been designated exon 33. Inclusion of ASCYT1 in human *NRCAM* mRNA

would require use of a cryptic acceptor site upstream of the exon 33 acceptor site within intron 32 (Fig. 2). We have therefore designated ASCYT1 as exon 33a.

The conservation of exons 33a and 33 over the long evolutionary distance between chick and man represents striking evidence for their functional importance. In order to screen for their presence in *NRCAM* cDNA, RT-PCR was carried out on RNA from human fetal brain (14 weeks gestation). Using PCR primers flanking this region (primer 67 from exon 31 and primer 72 from exon 34) a single band was obtained (Fig. 2, top, lane 2). Sequence analysis revealed that this 365 bp product represented a single transcript lacking both exons 33a and 33. The same result was obtained using adult brain cDNA. Thus, although exon 33 is flanked by consensus splice signals, it was not included at high copy number in human brain mRNA samples.

However, when a sense primer (primer 68) spanning the last eight bases of exon 32 and the first 14 bases of ASCYT1 was used with an antisense primer from exon 34 (primer 7) a faint 121 bp product was obtained (Fig. 2, second panel, lane 2). This was more apparent in a nested PCR reaction, and was found to contain both 33a and 33 sequences (lane 6). Thus, some *NRCAM* transcripts in fetal brain do result from splicing exon 32 to the cryptic acceptor site upstream of exon 33. To test the possibility that some *NRCAM* mRNA species contain exon 32 directly spliced to exon 33, i.e. contain exon 33 without 33a, a primer that bridged exons 32 and 33 (primer 73) was used in conjunction with one in exon 34 (primer 72). The expected band of 217 bp was obtained only in a nested reaction (Fig. 2, third panel, lane 6). Therefore, all three splicing variations depicted in Fig. 2 are found in human brain mRNA.

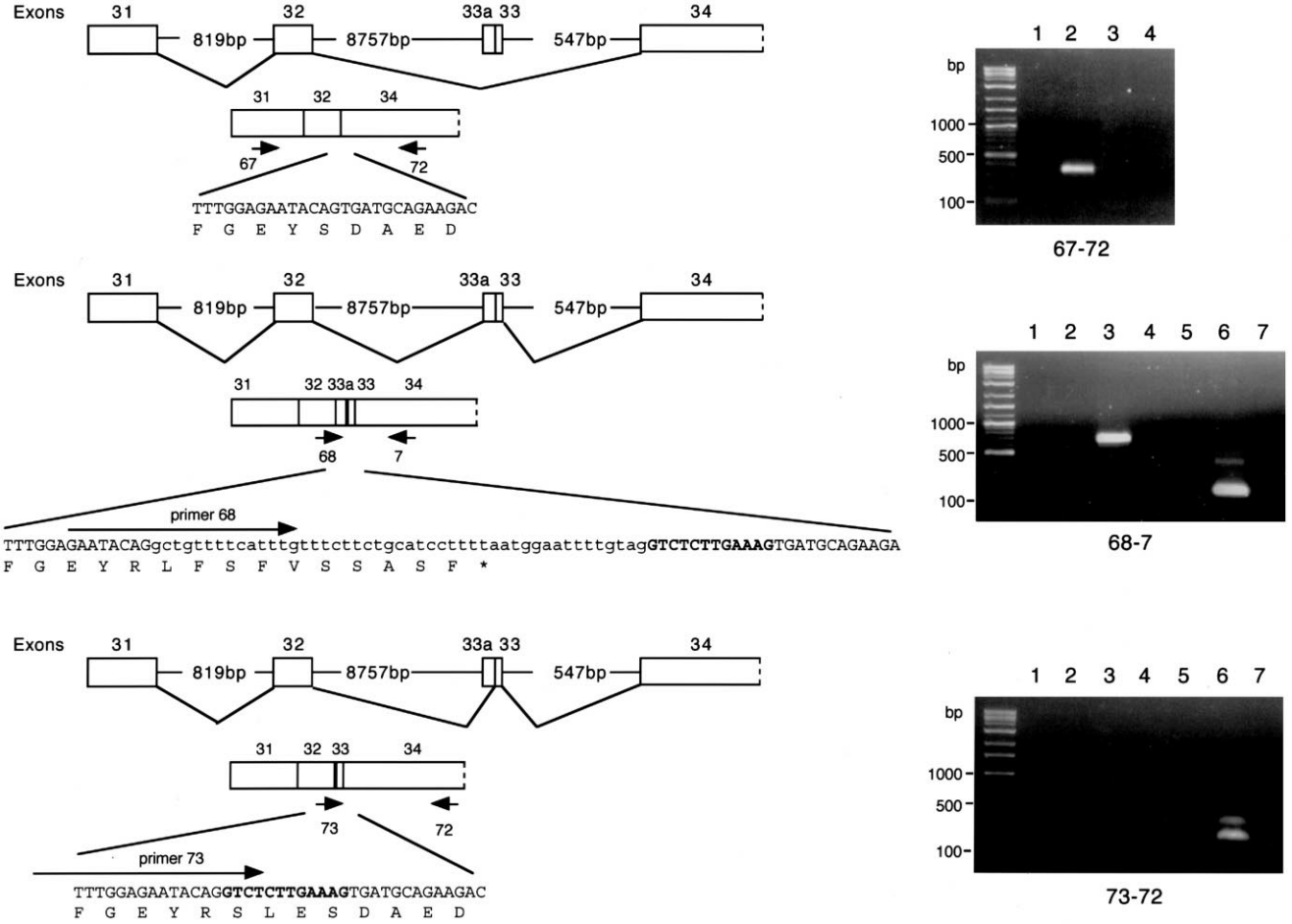

Fig. 2. Potential alternative splicing events at the 3′ end of the human *NRCAM* gene. Three possible splicing events and the sequences they would generate are shown alongside the results of RT-PCR experiments on 14-week-old fetal brain cDNA. The positions of primers used for RT-PCR and sequencing are shown by arrows. Exons and introns are in upper and lower case letters, respectively. The exon 33 sequence is depicted in bold type. (Top) RT-PCR amplification of fetal brain cDNA between primers 67 and 72 produced one product. Sequence analysis showed that this contained exon 32 spliced directly to exon 34. Lanes 1 and 2, minus RT and plus RT cDNA template; lane 3, genomic DNA as template; lane 4, minus template control. (Middle) PCR using primers 68 and 7 on the following templates: lanes 1 and 2, minus and plus RT fetal brain cDNA; lane 3, genomic DNA; lane 4, minus template control; lanes 5–7, nested reactions using aliquots of reactions in lanes 1, 2 and 4 of 67–72 products (top). The nested product was sequenced and found to contain both exons 33a and 33. (Bottom) PCR using primers 73 and 72 on the following templates: lanes 1 and 2, minus and plus RT fetal brain cDNA; lane 3, genomic DNA; lane 4, minus template control; lanes 5–7, nested reactions using aliquots of reactions in lanes 1, 2 and 4 of 67–72 products (top). The smaller band corresponds to cDNA containing exons 32 spliced to 33. The larger product did not yield a gene-specific sequence.

Exon 33 is 100% identical to ASCYT2 in chick and exon 27 in *L1*. In *L1* this exon is also alternatively spliced, being always present in neuronal tissues but absent from most non-neuronal tissues (Reid and Hemperly, 1992; Takeda et al., 1996). Kamiguchi and Lemmon (1998) demonstrated that this exon codes for part of a signal motif, YRLSE, that is essential for correct targeting of L1 to the extending growth cone of neurons. The identification of this same exon in *NRCAM* suggests that this molecule has the potential to be targeted in the same pathway. However, the only description of this exon in *NRCAM* cDNAs was in a single chick cDNA clone (Kayyem et al., 1992). It would appear from our PCR analysis that the major species of *NRCAM* mRNA in human fetal brain lacks this exon. Therefore, at least at this stage of development, the majority of the protein may not be directed to the developing growth cone but to the cell body. This is interesting in the light of the cellular distribution of NrCAM compared to L1 in neural tissue. Whereas L1 is expressed primarily by neurons NrCAM is also expressed on glial cells, where a motif regulating axonal transport may not be functional.

Similarly, we have demonstrated that ASCYT1 (exon 33a) is also present in a subpopulation of mRNA species. Utilization of the cryptic splice donor sequence in intron 32 to incorporate exon 33a would result in a frame shift leading to a termination codon within the highly conserved cytoplasmic domain. This would result in a truncated protein with the last 82 amino acids substituted with ten novel residues. Such a protein would not have the ankyrin-binding region described by Davis and Bennett (1994) that lies over 50 residues downstream of the exon 33-encoded sequence. NrCAM is associated with both ankyrin and neurofascin, another member of the L1 subfamily, at the nodes of Ranvier where it is thought to play a role in the formation or maintenance of the structure of the nodes (Davis et al., 1996). A novel, shorter version of NrCAM may be present in cells, or compartments of cells where ankyrin binding is not required. A truncated version of the protein may also have a role in negatively regulating the activities of the full-length protein. Whatever the role of this truncated isoform it is highly conserved between chick and man.

3.4. Confirmation of the localization of NRCAM close to D7S666 in 7q31

NRCAM was localized to 7q by PCR screening of the Genebridge 4 radiation hybrid panel (Wang et al., 1998). This placed *NRCAM* closest to D7S666 and between anchor markers WI-5853 and D7S658. Confirmation of this map position was sought by screening the genomic data for mapped STSs. SWSS3079, WI-6227 and WI-6368 were found within the 3′ UTR of *NRCAM* (bases 113,742–113,979, 114,122–114,386 and 115,493–115,640, respectively). SWSS2888 (95,942–96,349) lies within intron 27 and WIAF-4217 lies within BAC clone 267017 upstream of *NRCAM* exon 1. Both WI-6227 and WI-6368 have been placed on the WICGR radiation hybrid map and have been linked to YAC contig WC7.6, a contig that also houses marker D7S666. This confirms the genetic localization of *NRCAM* and physically links the gene to the nearest known polymorphic marker (D7S666), although this marker is now described as belonging to contigs in 7q31 rather than 7q21. *NRCAM* lies within the following context of known genes: *MDG1* (microvascular endothelial differentiation gene 1); *RPL7* (ribosomal protein L7); *IPLA2R* (similar to calcium-independent phospholipase A2); *NRCAM*; and *LAMB1R* (laminin beta 1-related).

4. Conclusions

In summary, we report the complete sequence of the human gene for NRCAM. The genomic sequence reveals remarkable conservation of exon/intron arrangement compared to other members of the L1 subfamily of CAMs. Furthermore, the gene contains two conserved 3′ exons previously not found in mammalian *NRCAM* mRNAs. Recruitment of these exons would alter either the trafficking or binding capacity of NRCAM protein. The availability of the complete sequence will also allow examination of the regulatory elements controlling *NRCAM* expression as well as the relationship of *NRCAM* to disorders involving 7q. For example, allelic loss of markers spanning the 7q22 region has been described in uterine leiomyomas, suggesting the presence of a neighboring tumor suppressor gene (Ishwad et al., 1997). In particular, loss of *D7S666*, the closest genetic marker to *NRCAM*, is frequently deleted.

Acknowledgements

We thank Nicole Steinlein and Hella Ludewig for expert technical assistance. This work was supported by grant BEO 0311108/0 from the German BMBF.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Brümmendorf, T., Rathjen, F.G., 1996. Structure/function relationships of axon-associated adhesion receptors of the immunoglobulin superfamily. Curr. Opin. Neurobiol. 6, 584–593.

Brümmendorf, T., Kenwrick, S., Rathjen, F.G., 1998. Neural cell recognition molecule L1: from cell biology to human hereditary brain malformations. Curr. Opin. Neurobiol. 8, 87–97.

Coutelle, O., Nyakatura, G., Taudien, S., Elgar, G., Brenner, S., Platzer, M., Drescher, B., Jouet, M., Kenwrick, S., Rosenthal, A., 1998. The neural cell adhesion molecule L1: genomic organisation and differential splicing is conserved between man and the pufferfish Fugu. Gene 208, 7–15.

Davis, J.Q., Bennett, V., 1994. Ankyrin binding activity shared by the neurofascin/L1/NrCAM family of nervous system cell adhesion molecules. J. Biol. Chem. 269, 27163–27166.

Davis, J.Q., Lambert, S., Bennett, V., 1996. Molecular composition of the node of Ranvier: identification of ankyrin-binding cell adhesion molecules neurofascin (mucin + /third FNIII domain-) and NrCAM at nodal segments. J. Cell Biol. 135, 1355–1367.

Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., Miller, W., 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res. 8, 967–974.

Fransen, E., Van Camp, G., Vits, L., Willems, P.J., 1997. L1-associated diseases: clinical geneticists divide, molecular geneticists unite. Hum. Mol. Genet. 6, 1625–1632.

Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282.

Grumet, M., 1997. Nr-CAM: a cell adhesion molecule with ligand and receptor functions. Cell Tissue Res. 290, 423–428.

Grumet, M., Mauro, V., Burgoon, M.P., Edelman, G.M., Cunningham, B.A., 1991. Structure of a new nervous system glycoprotein, Nr-CAM, and its relationship to subgroups of neural cell adhesion molecules. J. Cell Biol. 113, 1399–1412.

Ishwad, C.S., Ferrell, R.E., Hanley, K., Davare, J., Meloni, A.M., Sandberg, A.A., Surti, U., 1997. Two discrete regions of deletion at 7q in uterine leiomyomas. Genes Chromosomes Cancer 19, 156–160.

Kamiguchi, H., Lemmon, V., 1998. A neuronal form of the cell adhesion molecule L1 contains a tyrosine-based signal required for sorting to the axonal growth cone. J. Neurosci. 18, 3749–3756.

Kayyem, J.F., Roman, J.M., de la Rosa, E.J., Schwarz, U., Dreyer, W.J., 1992. Bravo/Nr-CAM is closely related to the cell adhesion molecules L1 and Ng-CAM and has a similar heterodimer structure. J. Cell Biol. 118, 1259–1270.

Kenwrick, S., Watkins, A., Angelis, E.D., 2000. Neural cell recognition molecule L1: relating biological complexity to human disease mutations. Hum. Mol. Genet. 9, 879–886.

Kozlov, S.V., Giger, R.J., Hasler, T., Korvatska, E., Schorderet, D.F., Sonderegger, P., 1995. The human TAX1 gene encoding the axon-associated cell adhesion molecule TAG-1/axonin-1: genomic structure and basic promoter. Genomics 30, 141–148.

Lane, R.P., Chen, X.N., Yamakawa, K., Vielmetter, J., Korenberg, J.R., Dreyer, W.J., 1996. Characterization of a highly conserved human homolog to the chicken neural cell surface protein Bravo/Nr-CAM that maps to chromosome band 7q31. Genomics 35, 456–465.

Li, W.H., Wu, C.I., Luo, C.C., 1985. A new method for estimating synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2, 150–174.

Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., Ohara, O., 1997. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res. 4, 141–150.

Pearson, W.R., Lipman, D.J., 1998. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

Plagge, A., Brümmendorf, T., 1997. The gene of the neural cell recognition molecule F11: conserved exon-intron arrangement in genes of neural members of the immunoglobulin superfamily. Gene 192, 215–225.

Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y., Rosenthal, A., 1997. Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. Genome Res. 7, 592–605.

Reid, R.A., Hemperly, J.J., 1992. Variants of human L1 cell adhesion molecule arise through alternate splicing of RNA. J. Mol. Neurosci. 3, 127–135.

Rosenthal, A., Jouet, M., Kenwrick, S., 1992. Aberrant splicing of neural cell adhesion molecule L1 mRNA in a family with X-linked hydrocephalus (published erratum appears in Nat. Genet., 1993, 3(3), 273). Nat. Genet. 2, 107–112.

Su, X.D., Gastinel, L.N., Vaughn, D.E., Faye, I., Poon, P., Bjorkman, P.J., 1998. Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. Science 281, 991–995.

Takeda, Y., Asou, H., Murakami, Y., Miura, M., Kobayashi, M., Uyemura, K., 1996. A nonneuronal isoform of cell adhesion molecule L1: tissue-specific expression and functional analysis. J. Neurochem. 66, 2338–2349.

Wang, B., Williams, H., Du, J.S., Terrett, J., Kenwrick, S., 1998. Alternative splicing of human NrCAM in neural and nonneural tissues. Mol. Cell. Neurosci. 10, 287–295.