ELSEVIER

Gene 208 (1998) 7–15

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

The neural cell adhesion molecule L1: genomic organisation and differential splicing is conserved between man and the pufferfish *Fugu*[1,2]

Oliver Coutelle [a,c], Gerald Nyakatura [a], Stefan Taudien [a], Greg Elgar [c], Sydney Brenner [c], Matthias Platzer [a], Bernd Drescher [a], Monique Jouet [c], Sue Kenwrick [c], André Rosenthal [a,b,*]

[a] *Institute of Molecular Biotechnology, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena, Germany*
[b] *Friedrich Schiller University Jena, Faculty of Biology and Pharmacy, 07743 Jena, Germany*
[c] *Department of Medicine, University of Cambridge Clinical School, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK*

Received 13 August 1997; accepted 14 October 1997; Received by K. Gardiner

Abstract

The human gene for the neural cell adhesion molecule L1 is located on Xq28 between the ALD and MeCP2 loci. Mutations in the L1 gene are associated with four related neurological disorders, X-linked hydrocephalus, spastic paraplegia (SPG1), MASA syndrome, and X-linked corpus callosum agenesis. The clinical relevance of L1 has led us to sequence the L1 gene in human and to investigate its conservation in the vertebrate model genome of the pufferfish, *Fugu rubripes* (*Fugu*), a species with a compact genome of around 40 Mb. For this purpose we have sequenced a human and a *Fugu* cosmid clone containing the corresponding L1 genes. For comparison, we have also amplified and sequenced the complete *Fugu* L1 cDNA. We find that the genomic structure of L1 is conserved. The human and *Fugu* L1 gene both have 28 exons of nearly identical size. Differential splicing of exons 2 and 27 is conserved over 430 million years, the evolutionary time span between the teleost *Fugu* and the human L1 gene. In contrast to previously published *Fugu* genes, many introns are larger in the *Fugu* L1 gene, making it slightly larger in size despite the compact nature of the *Fugu* genome. Homology at the amino acid and the nucleotide level with 40% and 51%, respectively, is lower than that of any previously reported *Fugu* gene. At the level of protein structure, both human and *Fugu* L1 molecules are composed of six immunoglobulin (Ig)-like domains and five fibronectin (Fn) type III domains, followed by a transmembrane domain and a short cytoplasmic domain. Only the transmembrane and the cytoplasmic domains are significantly conserved in *Fugu*, supporting their proposed function in intracellular signalling and interaction with cytoskeletal elements in the process of neurite outgrowth and fascicle formation. Our results show that the cytoplasmic domain can be further subdivided into a conserved and a variable region, which may correspond to different functions. Most pathological missense mutations in human L1 affect conserved residues. Fifteen out of 22 reported missense mutations alter amino acids that are identical in both species. © 1998 Elsevier Science B.V.

Keywords: Comparative genomics; Gene organisation; Differential splicing; X chromosome

1. Introduction

We and others have previously identified the gene for the human neural cell adhesion molecule L1 (Hlavin and Lemmon, 1991; Rosenthal et al., 1991). Fluorescence in situ hybridisation (FISH) analysis with the mouse L1 cDNA showed that the L1 gene mapped to the distal region of the long arm of the human X chromosome (Djabali et al., 1990). We are presently sequencing a 3 Mb region in Xq28, a chromosomal region that is particularly gene dense and harbours many disease genes. Genomic sequencing of a 400 kb

* Corresponding author. Tel.: +49 3641 656240; Fax: +49 3641 656255; e-mail: arosenth@imb-jena.de

[1] O.C. and G.N. share first authorship.

[2] EMBL Accession Nos. Z29373 and Z71926.

Abbreviations: aa, amino acid(s); ACC, corpus callosum agenesis; ALD, adrenoleukodystrophy; bp, base pair(s); CAM, cell adhesion molecule; cDNA, DNA complementary to RNA; EMBL, European Molecular Biology Laboratory; FGF, fibroblast growth factor; FISH, fluorescence in situ hybridisation; Fn, fibronectin; *Fugu*, *Fugu rubripes*; G6PD, glucose-6-phosphate dehydrogenase; HD, Huntington's disease; Ig, immunoglobulin; kb, kilobase(s); MASA, mental retardation, aphasia, shuffling gait, adducted thumbs; Mb, megabase(s); MeCP, methyl CpG binding protein; no(s), number(s); PCR, polymerase chain reaction; RT, reverse transcriptase; SPG, spastic paraplegia; UTR, untranslated region.

0378-1119/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0378-1119(97)00614-8

cosmid contig performed in our laboratory showed that the L1 gene is located between the ALD and the MeCP2 locus (GenBank Accession Nos U52111 and U52112). We also showed that mutations in L1 are responsible for three clinically related neurological disorders: X-linked hydrocephalus (Rosenthal et al., 1992; Jouet et al., 1993), spastic paraplegia (SPG1) and MASA syndrome (McKusick, 1993; Jouet et al., 1994) (McKusick: Nos. 30700, 31290, and 303350). Other groups confirmed our findings (Van Camp et al., 1993; Vits et al., 1994) and demonstrated that X-linked corpus callosum agenesis (ACC) is also caused by mutations in the L1 gene (for a recent review, see Fransen et al., 1995).

The function of L1 in the pathology of these disorders is still unclear, but L1 has been implicated in neuronal cell adhesion, neurite outgrowth, axon fasciculation and the development of neuromuscular junctions (Lindner et al., 1983; Stallcup and Beasley, 1985; Landmesser et al., 1988; Grumet, 1991; Sonderegger and Rathjen, 1992; Barami et al., 1994). L1 is highly conserved amongst mammalian species with amino acid identity of 80–95% in the extracellular domains, and complete conservation of the cytoplasmic domain (Moos et al., 1988; Miura et al., 1991). The high degree of similarity makes it difficult to identify conserved amino acid domains that are of functional importance. We therefore investigated if sequence comparison with a more distantly related species could prove more informative.

It has been shown that the Japanese puffer fish *Fugu rubripes (Fugu)* provides a useful model for comparative genome analysis (Brenner et al., 1993). The *Fugu* genome is approximately 7- to 8-fold smaller than that of mammals. Since Tetraodontoid fish share the body plan of higher vertebrates, the total number of *Fugu* genes is expected to be similar, however, the genomic structure of most genes should be more compact. A significant reduction in intron size and intergenic distances in *Fugu* has been shown by a shotgun sequencing survey of *Fugu* DNA (Brenner et al., 1993) and by analysing several *Fugu* genes (Trower et al., 1996; Venkatesh et al., 1996) including homologues for the Huntington's disease (HD) gene (Baxendale et al., 1995), *G6PD* gene (Mason et al., 1995) and P55 gene (Elgar et al., 1995). The clinical relevance of the region around L1 has prompted us to investigate the corresponding region in *Fugu*. Here we report a sequence comparison between the human and the *Fugu* loci for the neural cell adhesion molecule L1.

2. Materials and methods

2.1. Resources and sequencing strategy

2.1.1. Human L1 cosmid

A genomic cosmid clone (B5) containing the human L1 gene was completely sequenced. Cosmid DNA was prepared, sonicated, end-repaired and size-selected. A 0.8–1.4 kb fraction was then subcloned into the *Sma*I site of M13mp18. A total of 500 recombinant M13 phages were sequenced using cycle sequencing and Taq dye terminator chemistry (Applied Biosystems Division of Perkin Elmer). In order to obtain anchor clones for ordering contigs after sequence assembly as well as for 'primer walking', a 6–8 kb fraction was cloned into the *Sma*I site of the pUC18 vector. Forty plasmid clones were sequenced in both directions using universal forward and reverse primers and Taq dye terminators. Raw data were collected on ABI 373A automated gel readers.

2.1.2. Human L1 cDNA

Parts of the human L1 cDNA were cloned (Rosenthal et al., 1991) and sequenced to determine the exon/intron structure of the L1 gene.

2.1.3. Fugu *L1 cosmid*

A cosmid library was constructed in lawrist 4 vector. A total of 38016 clones were picked and spotted onto nylon membranes, such that the entire library could be screened on two filters. The *Fugu* cosmid library was screened by Southern hybridisation after attempts to PCR-amplify regions of L1 from genomic *Fugu* DNA using degenerate oligonucleotides based on mammalian sequencing alignments remained unsuccessful. Southern hybridization was carried out at 50°C according to the original protocol described by Church and Gilbert (see Mason et al., 1995) using a human L1 cDNA clone containing approximately 1.5 kb of the 3'- end of gene as a probe. Four hybridisation positive cosmids were picked. Fingerprinting analysis identified two overlapping cosmids (32L19 and 30O21) which were analyzed further by Southern blotting before cosmid 30O21 was selected for sequencing. Cosmid DNA was fragmented by sonication and subcloned for 'shotgun sequencing' into pUC18 plasmids. Plasmid DNA was prepared for sequencing using QIAwell96Ultra plasmid Kits. Sequencing was performed with Dye primer chemistry using Autoread Kits (Pharmacia) and raw data were collected on a Pharmacia A.L.F. gel reader. Plasmid shotgun clones were sequenced on both strands using universal primers before gaps remaining between the forward and reverse reads were closed by sequencing with custom-made walking primers.

2.1.4. Fugu *L1 cDNA*

Fugu brain RNA was subjected to RT–PCR according to standard methods. L1-specific cDNA species were amplified from a dilution of the RT–PCR product using nested primer sets. In the first round of PCR amplification, two overlapping cDNA fragments were produced. Fragment 1, beginning 130 bp upstream of the putative translation start site extending to the end of exon 14, and fragment 2, extending from the beginning of exon

14 to 400 bp into the non-coding region of the L1 gene. Nested primers were then designed to amplify the complete L1 coding region in 14 overlapping fragments suitable for direct sequencing. All fragments were sequenced from both ends and the sequences used to determine the exact exon/intron boundaries.

2.2. Database searches and computer analysis

Raw data were processed and assembled into contigs with the Staden XBAP software package (Staden, 1987; Dear and Staden, 1991). Sequence motif and structural homology searches were carried out with the Staden package, FASTA/TFASTA, Blastn using the following databases: EMBL38, SWISS-PRROT28. The human sequence was deposited in the EMBL database (Accession No. Z29373). The *Fugu* sequence for L1 was submitted to EMBL database (Accession No. Z71926). Splice sites and exon/intron boundaries were predicted by the XGRAIL2 package (Xu et al., 1994) on the basis of base content and confirmed by alignment of the genomic sequence with the corresponding RT–PCR fragments.

3. Results

3.1. Genomic structure of L1 in human and Fugu

We have sequenced a human and a *Fugu* cosmid clone containing the gene for the neural cell adhesion molecule L1. In addition, we have sequenced both human and *Fugu* L1 cDNAs to determine whether the genomic organisation of the L1 gene is conserved between these two species.

The overall size of the *Fugu* L1 gene from the putative translation start site to the termination codon is 13 988 bp, which compares with 12 942 bp for the human L1 gene. L1 has 28 exons in both human and *Fugu* (Fig. 1), and a comparison of exon and intron sizes is shown in Table 1. The intron phase is completely conserved. Exon sizes are almost identical in both species, with the exception of the first and last exon. Substantial size differences between the *Fugu* and human L1 genes were observed for the corresponding introns. Surprisingly, despite the small genome size, some *Fugu* introns are larger than their human counterparts. Only 10 of the 27 (37%) *Fugu* introns in L1 are smaller than 120 bp, a figure well below the 75% previously reported (Brenner et al., 1993; Elgar et al., 1996).

The usage of splice consensus sites is similar in the *Fugu* and the human L1 gene. CAG splice sites are the most common, with 25 sites (92%) in man and 21 (78%) in *Fugu*, followed by TAG with two sites (8%) in human and 4 (15%) in *Fugu*. In 2 (7%) of the *Fugu* introns AAG acceptor sites were present, none were found in the human gene.

We have compared 5 kb of genomic sequence upstream of exon 1 in the human and *Fugu* L1 gene using a local similarity search programme (Huang, 1991) and did not find significant homology in the 5′-promoter regions between the two species. Pairwise analysis of the L1 introns did not reveal any conserved sequence elements between the two species.

The *Fugu* L1 homologue has a G,C content of 46% compared with 60% for the human L1 locus (Fig. 1). In the human L1 gene, there are three Alu repeats in the 5′ UTR (positions 1–144, 146–343, 365–657), one Alu repeat in intron 18 (position 11 050–11 337), and two MER42c repeats in intron 18 (positions

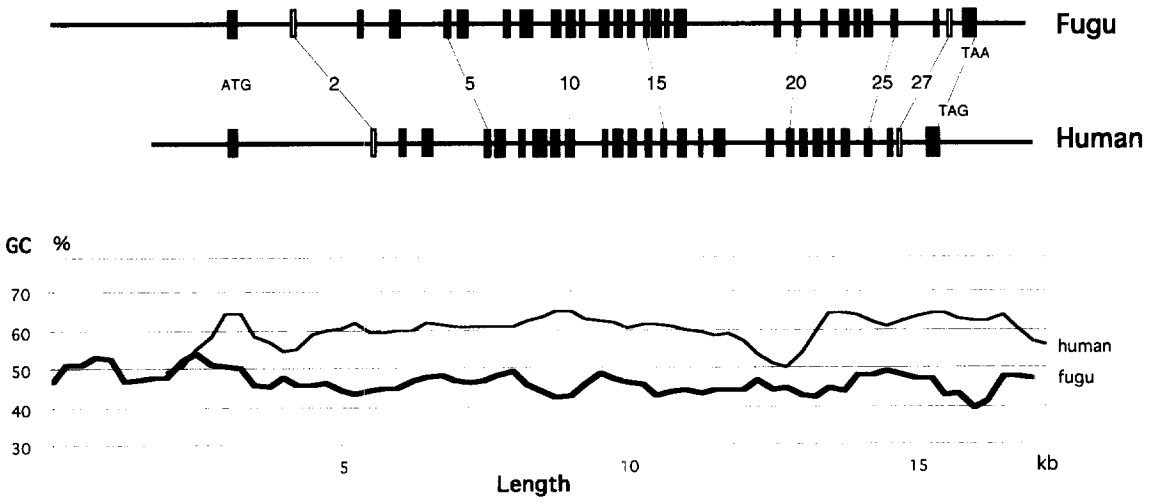


Fig. 1. Genomic map of the *Fugu* and human L1 gene and distribution of G,C content. The G,C distribution was obtained with a moving window of 1 kb, step 0.25 kb using the database entries Z71926 and Z29373.

Table 1
Exon/intron structure of the *Fugu* and human L1 gene

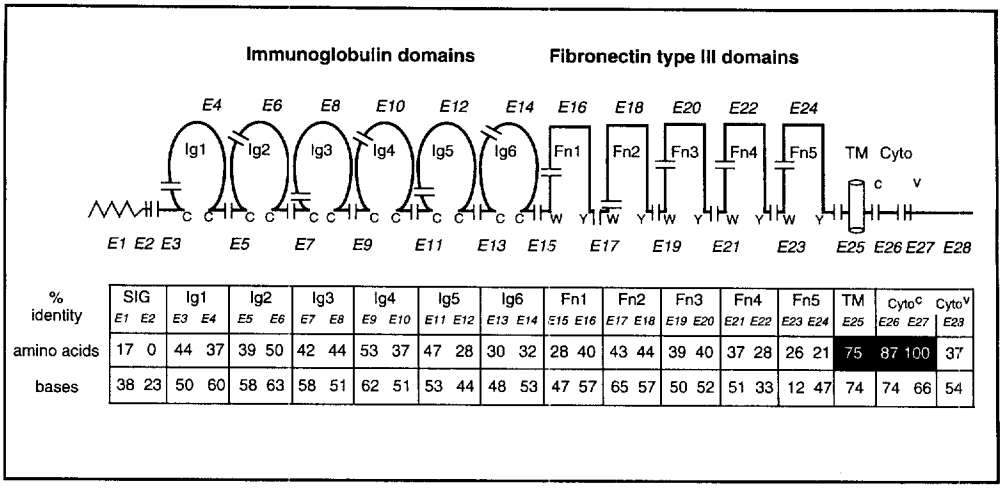
No.	Exon			Intron			Intron-phase
	Fugu	Human	Δ	*Fugu*	Human	Δ	
1	124	76	48	1085	2518	−1433	1
2	15	0	1176	530	646	1	
3	103	106	−3	439	237	202	2
4	203	203	0	865	974	−109	1
5	123	123	0	122	96	26	1
6	171	171	0	699	289	410	1
7	115	112	3	216	148	68	2
8	185	185	0	407	121	286	1
9	132	132	0	100	139	−39	1
10	147	144	3	82	565	−483	1
11	112	112	0	296	113	183	2
12	161	167	−6	86	102	−16	1
13	160	157	3	98	180	−82	2
14	125	125	0	84	87	−3	1
15	102	111	−9	113	246	−133	1
16	198	198	0	98	231	−133	1
17	71	71	0	101	182	−81	0
18	226	223	3	1642	829	813	1
19	116	116	0	275	203	72	0
20	205	202	3	283	88	195	1
21	123	123	0	231	93	138	1
22	174	174	0	86	116	−30	1
23	123	120	3	86	107	−21	1
24	150	156	−6	362	303	59	1
25	132	135	−3	643	333	310	1
26	73	73	0	171	97	74	2
27	12	12	0	308	473	−165	2
28	253	232	21				
Sum	3834	3774		1 0154	9400		

10 805–10 888, 11 340–11 455). The *Fugu* L1 locus contains several microsatellite repeats: pentanucleotide CCTCT in the 5′ UTR (position 1074–1276), tetranucleotide CTCA in intron 18 (position 12 187–12 276), and two large dinucleotides CA in intron 3 and intron 18 (position 5882–5924; and 13 202–13 279).

3.2. *Homology between human and* Fugu *L1 cDNA*

The overall identity between the *Fugu* and human cDNA is 51% at the nucleotide level and 40% at the amino acid level. Fig. 2 shows the sequence identity on the amino acid and nucleotide level between exons for both species. Complete conservation of cysteine and tryptophan residues involved in the immunoglobulin fold and conservation of tryptophan and tyrosine (or phenylalanine) residues of the fibronectin repeating unit allowed us to identify the structural domains characteristic of L1. The *Fugu* L1 homologue is composed of six immunoglobulin (Ig-like) and five fibronectin (Fn) type III domains. This arrangement is unique to all L1 molecules characterised from vertebrate organisms including human, mouse, and rat and distinguish it from other molecules e.g. NCAM of the immunoglobulin superfamily (Fig. 2). The Ig-specific cysteine residues and the Fn-specific tryptophan and tyrosine residues are highlighted. Fig. 2 also shows that except for the transmembrane domain, all immunoglobulin (Ig)-type 2 and fibronectin (Fn)-type III domains are formed by two exons. While the Ig domains 1, 3, and 5 are composed of one very short 5′ exon followed by a longer 3′ exon, the Ig domains 2, 4, and 6 are composed of two exons of almost identical size.

Between human and *Fugu* L1, only the transmembrane and the cytoplasmic domains show a high degree of cross-species homology (Fig. 2). At the amino acid level, the transmembrane domain is 75% conserved between the two species. The cytoplasmic domain appears to be subdivided into a conserved region of 85–100% identity containing exons 26 and 27 and a variable region of 37% identity containing exon 28. Amongst vertebrate species (human, rodent, fish), the transmembrane and the following 44 amino acid residues (Fig. 3, position 17–69) of the cytoplasmic domain are highly conserved. In contrast, in *Drosophila* neuroglia the cytoplasmic domain shows little homology with mammalian L1 molecules and no alignment can be made. Remarkably, in part of the variable region (Fig. 3,



% identity	SIG E1 E2	Ig1 E3 E4	Ig2 E5 E6	Ig3 E7 E8	Ig4 E9 E10	Ig5 E11 E12	Ig6 E13 E14	Fn1 S15 E16	Fn2 E17 E18	Fn3 E19 E20	Fn4 E21 E22	Fn5 E23 E24	TM E25	Cyto^{c} E26 E27	Cyto^{v} E28
amino acids	17 0	44 37	39 50	42 44	53 37	47 28	30 32	28 40	43 44	39 40	37 28	26 21	75	87 100	37
bases	38 23	50 60	58 63	58 51	62 51	53 44	48 53	47 57	65 57	50 52	51 33	12 47	74	74 66	54

E = Exon
SIG = Signal peptide
TM = Transmembrane region
Cyto = Cytoplasmic domain
c = conserved region of cytoplasmic domain
v = variable region of cytoplasmic domain

Fig. 2. Domain structure of the the *Fugu* and human L1 molecule and their homology on the amino acid and nucleotide level. E=exons, Sig=signal peptide, TM=transmembrane region, Cyto=cytoplasmic domain, con=constant region of cytoplasmic domain, var=variable region of cytoplasmic domain.

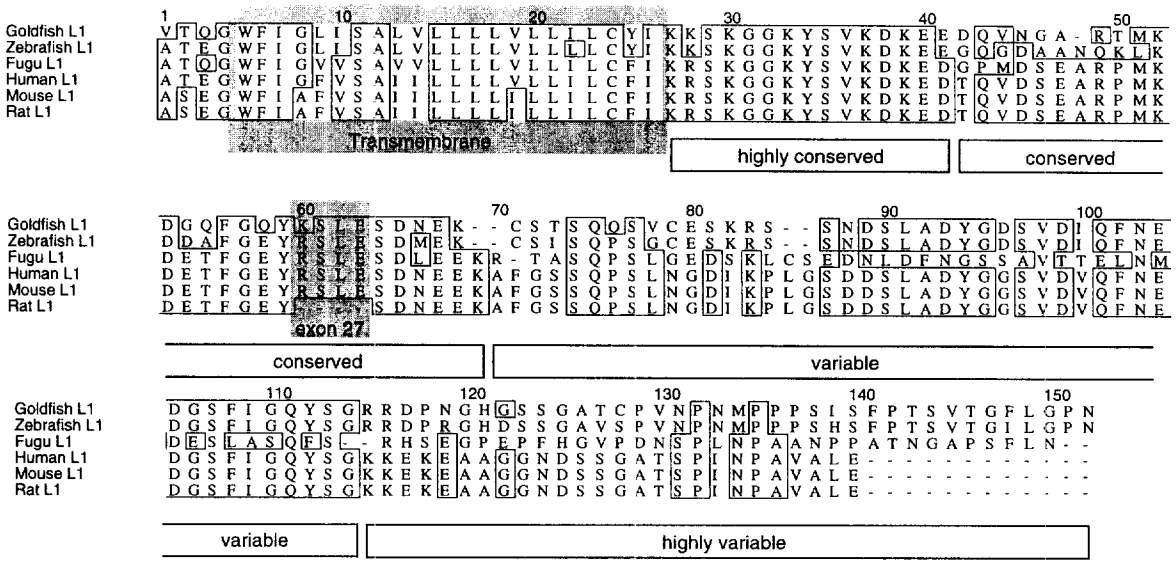


Fig. 3. Alignment of transmembrane and cytoplasmic domains of L1 molecules from goldfish (U55211), zebrafish (X89204), *Fugu*, human, mouse (X12875) and rat (X59149). The transmembrane domain is shaded grey. The highly conserved part of the cytoplasmic domain starts with the amino acids KKSK.

position 87–114) the *Fugu* L1 molecule differs significantly from all other vertebrate L1 proteins including zebrafish and goldfish.

L1 is a transmembrane glycoprotein. Nearly twice as many potential *N*-glycosylation sites were predicted for the human protein (22 sites) compared with the *Fugu* homologue (12 sites) when the amino acid sequence was examined using the PROSITE database. Only three of the glycosylation sites are conserved between the two species and are located at amino acid positions 505, 825, 1022 (human) and 520, 839 and 1035 (*Fugu*).

3.3. *Alternative splicing of exons 2 and 27 is conserved in* Fugu *and human*

Tissue-specific differential splicing leading to skipping of mini exons 2 and 27 in the resulting L1 isoform has been observed in human and shown to be conserved in

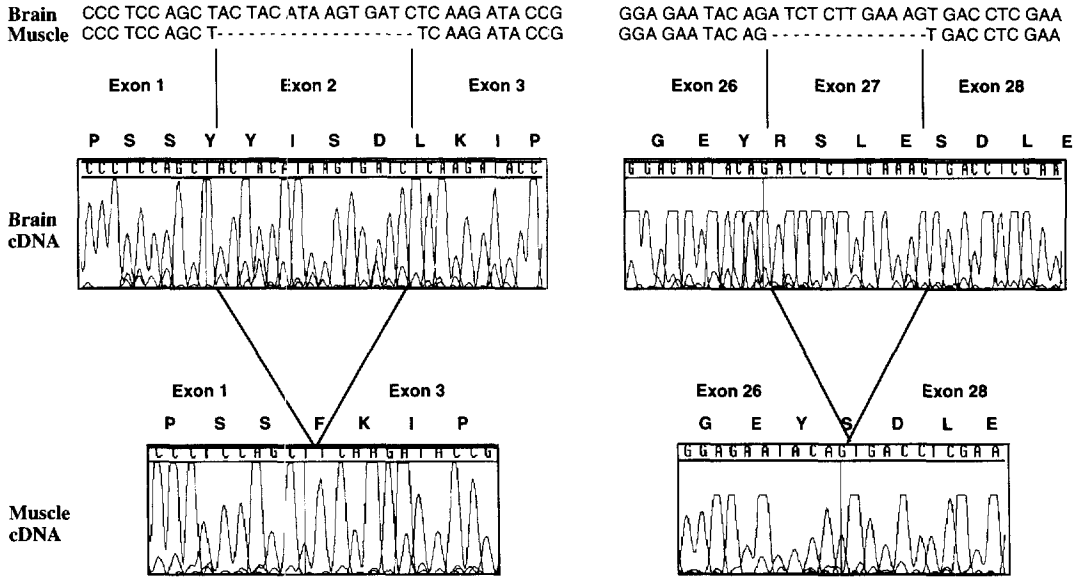


Fig. 4. Differential splicing of the mini-exons 2 and 27 of the *Fugu* L1 homologue. Due to splicing of exons 2 in total muscle tissue, the first tyrosine is changed into phenylalanine and five amino acids (YISDL) are lost. Splicing of exon 27 in *Fugu* muscle leads to loss of the four amino acids RSLE.

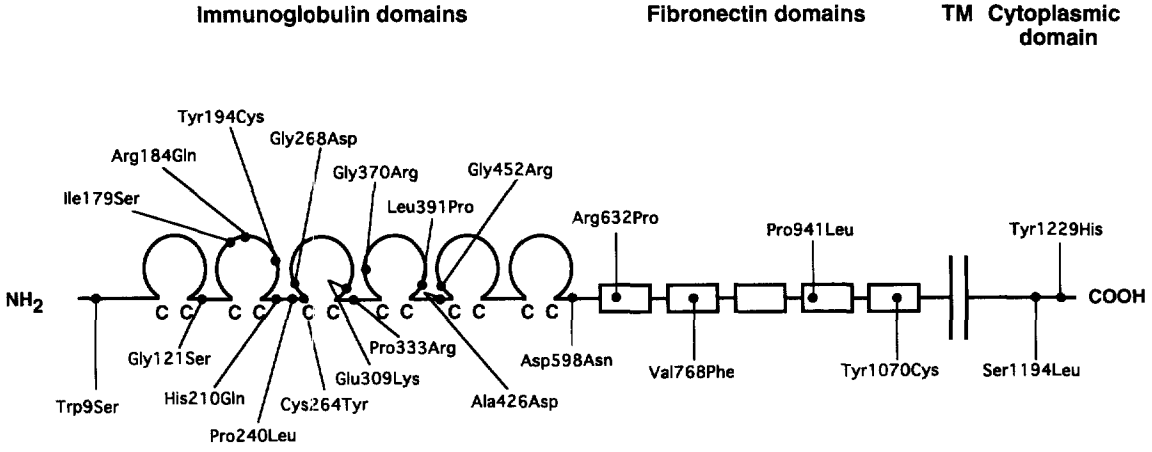


Fig. 5. Missense mutations and their location in the human L1 molecule.

mouse (Jouet et al., 1995). Both exons are present in brain mRNA, but spliced from the leukocytes transcripts and seven human tissues like heart, kidney, liver, lung, skeletal muscle, pancreas, and placenta (Taudien et al., unpublished). To examine if the differential splicing pattern is evolutionary conserved, we sequenced PCR fragments generated from *Fugu* total muscle and brain cDNA. We find that the alternative splicing is conserved in *Fugu*. While in muscle transcripts exons 2 and 27 are absent, both exons are retained in the brain variant in the L1 mRNA (Fig. 4).

An interesting observation is that at the amino acid level, exon 2 with a length of 15 bp is not conserved between man and *Fugu*, whereas exon 27 which is part of the cytoplasmic domain and has a length of 12 bp has remained completely conserved over 430 million years of evolution (Fig. 2).

3.4. Pathological mutations in L1 affect conserved residues

To date, 52 mutations in L1 have been described, in patients with either X-linked hydrocephalus, MASA syndrome, SPG1, or X-linked corpus callosum agenesis (Rosenthal et al., 1992; Jouet et al., 1993; Jouet et al., 1994), (Van Camp et al., 1993; Vits et al., 1994; Fransen et al., 1995; http://dna-lab-www.uia.ac.be/dnaLab/l1.html). The 22 missense mutations identified to date are summarised in Fig. 5. Mutations are distributed across most of the domains and each is unique to a single family. Comparison of human with *Fugu* L1 reveals that many of these mutations affect residues conserved between the two species. Out of 22 reported missense mutations, 15 affect or cause substitutions of amino acids that are identical in both species (Table 2).

Table 2
Amino acid comparison at sites of missense mutations in man, mouse, rat, chicken, *Fugu*, goldfish, zebrafish and *Drosophila*

Human	Mouse	Rat	Chicken	*Fugu*	Goldfish	Zebrafish	*Drosophila*
Trp9Ser	+	Leu	Leu	Leu	n.d.	n.d.	Leu
Gly121Ser	+	+	+	+	+	+	+
Ile179Ser	+	+	+	+	+	+	+
Arg184Gln	+	+	+	+	+	+	Asn
Tyr194Cys	+	+	+	+	+	+	Lys
His194Cys	Asn	Asn	+	Asn	Asn	Asn	Ser
Pro240Leu	+	+	+	+	+	Ala	His
Cys264Tyr	+	+	+	+	+	+	+
Gly268Asp	+	+	+	+	+	+	+
Glu309Lys	+	+	+	+	+	+	Thr
Pro333Arg	+	+	+	+	+	+	+
Gly370Arg	+	+	+	+	+	+	+
Leu391Pro	+	+	+	+	+	+	Ile
Ala426Asp	+	+	Leu	Pro	Pro	+	Pro
Gly452Arg	+	+	+	+	+	+	+
Asp598Asn	+	+	+	+	+	+	+
Arg632Pro	His	His	+	Thr	Thr	Thr	Glu
Val768Phe	+	+	+	Thr	Ala	+	Ile
Pro941Leu	+	+	Thr	+	+	+	+
Tyr1070Cys	+	+	Ser	Ser	Ser	Ser	Ser
Ser1194Leu	+	+	Gly	+	+	+	Asn
Tyr1229His	+	+	+	Phe	+	+	+

Amino acids identical with the wild-type human gene are indicated by '+', alterations in the aa are given as a three-letter aa-code.

Five mutations (Trp9Ser, His210Gln, Ala426Asp, Arg632Pro, Tyr1070Cys) affect residues that are conserved in other species, while two mutations (Val768Phe, Tye1229His) are not conserved.

While the overall homology between *Fugu* and human L1 at the amino acid level is only 40%, we find that 75% of amino acid changes associated with clinical pathology in patients are conserved, suggesting evolutionary constraints on these functionally important residues of the protein (Bateman et al., 1996).

3.5. *The RGD motif*

The Ig-like domain VI harbours two RGD-sequences at amino acid residues 553–555 and 562–564, which can serve as recognition sequences for integrin-type receptors. Computer-assisted modelling of the L1 molecule showed both RGD sequences localized at the molecular surface (Drescher et al., 1996) are an obvious prerequisite for interaction with integrins. In contrast, *Fugu* L1 does not have the RGD site, indicating that this region may be functionally less important.

4. Discussion

Comparison of the human and the *Fugu* L1 gene has shown that the genomic organisation of both genes is very similar. As a reflection of the identical number of exons with very similar sizes, the overall structural organisation of the protein is identical in both human and *Fugu*. L1 is a transmembrane glycoprotein of the immunoglobulin (Ig) superfamily of cell adhesion molecules (CAMs) (Grumet, 1991; Sonderegger and Rathjen, 1992). The molecular weight is approximately 200 kDa, with six Ig-like type C2 domains, followed by five fibronectin type III domains, a single-pass transmembrane region and a cytoplasmic domain.

The *Fugu* L1 gene identified by us shows the same genomic organisation and identical intron phases as the human L1 gene. In addition, the deduced protein of this *Fugu* gene exhibits the same domain architecture as the human neural cell adhesion molecule L1. Since this domain architecture is unique to L1 molecules among the Ig superfamily and among vertebrate species, we are convinced that the *Fugu* gene described is a true homologue of the human gene for the neural cell adhesion molecule L1.

Despite the apparent conservation of structural features between human and *Fugu*, conservation at the amino acid and nucleotide level is generally low (40% and 51%, respectively). The high degree of sequence conservation (80–95%) between the mammalian L1 homologues does not permit identification of functionally important regions on the basis of evolutionary conservation. However, alignment with the more distant vertebrate *Fugu* gene shows that conservation at both the nucleotide and the amino acid level is essentially confined to the transmembrane and the adjacent part of the cytoplasmic domain with 80% and 98% amino acid homology, respectively. The cytoplasmic domain of the L1 molecule has been implicated in two functions during

neuronal development: (i) interaction with cytoskeletal elements; and (ii) intracellular signalling. Detailed comparison of the L1 cytoplasmic amino acid sequence between human and *Fugu* showed that the cytoplasmic domain can be further subdivided into a conserved and a variable region with a length of 44 and 80 amino acids, respectively. Fig. 3 shows that in the conserved region *Fugu* is more similar to mammals than to other fish species. In contrast, within the variable region *Fugu* L1 differs remarkbably from all other vertebrate L1 molecules, including those of other fish species. The amino acid sequence Arg–Ser–Leu–Glu (RSLE) present in exon 27 encoding part of the cytoplasmic domain resembles a motif required for casein-II-dependent phosphorylation of L1 (Sadoul et al., 1989; Wong et al., 1995). This phosphorylation event may have an effect on the ability of L1 to interact with the FGF receptor, which is thought to elicit a second messenger cascade that affects neurite outgrowth (Williams et al., 1994). The cytoplasmic domain of L1 and related molecules has also been found to interact with ankyrin molecules that associate with the spectrin-based cytoskeleton, thus binding it close to the plasma membrane. Studies by Davis and Bennett (1994) indicate that the domain primarily responsible for binding lies in the C-terminal 'variable' region of the cytoplasmic domain that is much less conserved in *Fugu*. This may indicate that *Fugu* L1 has binding capabilities that are different from the mammalian protein, or alternatively, *Fugu* ankyrin may be correspondingly altered to compensate for these differences.

Our findings suggest that the glycosylation levels of the human and *Fugu* protein may be significantly different, as the number of potential *N*-glycosylation sites in *Fugu* is only 50% compared with the number of sites found for the human L1 protein. Interestingly, only three of the 22 glycosylation sites in human are conserved in *Fugu*, and they may represent sites of functional significance.

In contrast to previously described *Fugu* genes, the L1 gene is not reduced in size relative to the human homologue, despite the fact that the *Fugu* genome is nearly 8-times smaller than mammalian genomes. Since teleosts are the most distant vertebrate precursors of mammals (Powers, 1991), this may reflect evolutionary constraints on functionally important elements located within the introns. However, pairwise alignment of all L1 introns between man and *Fugu* did not identify conserved intronic elements. Despite the poor sequence conservation of the homologues, we find tissue-specific differentially spliced isoforms in both human and *Fugu* to be conserved over 430 million years of separation between mammals and teleosts. In both human and *Fugu*, the mini-exons 2 and 27 are spliced from the muscle L1 mRNA, whereas they are retained in brain. The human exons 2 and 27 are also spliced out in most tissues (unpublished data) including B cells (Jouet et al., 1995). The fact that splicing occurs is not only indicative of the fact that we have identified a true L1 homologue, but also suggests a crucial albeit unknown role of the regions covered by exons 2 and 27. The functional significance of the two isoforms is at present not understood. The RSLE motif of exon 27 might be required for casein-II-dependent phosphorylation (Sadoul et al., 1989; Wong et al., 1995) thus affecting the ability of L1 to interact with the FGF receptor, which is thought to elicit a second messenger cascade that affects neurite outgrowth (Wong et al., 1995). It is interesting to note that the RSLE motif is present in all vertebrate L1 molecules including *Fugu* but absent in *Drosophila* neuroglian, an invertebrate cell adhesion molecule with similar protein architecture to L1. In contrast to the RSLE motif of exon 27, the amino acid sequence of exon 2 (*Fugu*: YISDL; mouse: YKGHH; human: YEGHH) involved in tissue-specific splicing is not conserved. However, the exon size was maintained over 430 million years of evolution, suggesting that this region may have a distinct function in the nervous system.

References

Barami, K., Kirschenbaum, B., Lemmon, V., Goldman, S.A., 1994. N-cadherin and Ng-CAM/8D9 are involved serially in the migration of newly generated neurons into the adult songbird brain. Neuron 13, 567–582.

Bateman, A., Jouet, M., MacFarlane, J., Du, J.S., Kenwrick, S., Chothia, C., 1996. Outline structure of the human L1 cell adhesion molecule and the sites where mutations cause neurological disorders. EMBO J. 15, 6050–6059.

Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S., 1995. Comparative sequence analysis of the human and pufferfish Huntington's disease genes. Nat. Genet. 10, 67–76.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., Aparicio, S., 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. Nature 366, 265–268.

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.

Dear, S., Staden, R., 1991. A sequence assembly and editing program for efficient management of large projects. Nucl. Acids Res. 19, 3907–3911.

Djabali, M., Mattei, M.G., Nguyen, C., Roux, D., Demengeot, J., Denizot, F., Moos, M., Schachner, M., Goridis, C., Jordan, B.R., 1990. The gene encoding L1, a neural adhesion molecule of the immunoglobulin family, is located on the X chromosome in mouse and man. Genomics 7, 587–593.

Drescher, B., Spiess, E., Schachner, M., Probstmeier, R., 1996. Structural analysis of the murine cell adhesion molecule L1 by electron microscopy and computer-assisted modelling. Eur. J. Neurosci. 8, 2467–2478.

Elgar, G., Rattray, F., Greystrong, J., Brenner, S., 1995. Genomic structure and nucleotide sequence of the p55 gene of the puffer fish *Fugu rubripes*. Genomics 27, 442–446.

Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B.,

Brenner, S., 1996. Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). TIGS 12, 145–150.

Fransen, E., Lemmon, V., Van Camp, G., Vits, L., Coucke, P., Willems, P.J., 1995. CRASH syndrome: clinical spectrum of corpus callosum hypoplasia, retardation, adducted thumbs, spastic paraparesis and hydrocephalus due to mutations in one single gene, L1. Eur. J. Hum. Genet. 3, 273–284.

Grumet, M., 1991. Cell adhesion molecules and their subgroups in the nervous system. Curr. Opin. Neurobiol. 1, 370–376.

Hlavin, M.L., Lemmon, V., 1991. Molecular structure and functional testing of human L1CAM: an interspecies comparison. Genomics 11, 416–423.

Huang, X., 1991. A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics 12, 337–357.

Jouet, M., Rosenthal, A., Armstrong, G., MacFarlane, J., Stevenson, R., Paterson, J., Metzenberg, A., Ionasescu, V., Temple, K., Kenwrick, S., 1994. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. Nat. Genet. 7, 402–407.

Jouet, M., Rosenthal, A., Kenwrick, S., 1995. Exon 2 of the gene for neural cell adhesion molecule L1 is alternatively spliced in B cells. Mol. Brain Res. 30, 378–380.

Jouet, M., Rosenthal, A., MacFarlane, J., Kenwrick, S., Donnai, D., 1993. A missense mutation confirms the L1 defect in X-linked hydrocephalus. Nat. Genet. 4, 331

Landmesser, L., Dahm, L., Schultz, K., Rutishauser, U., 1988. Distinct roles for adhesion molecules during innervation of embryonic chick muscle. Dev. Biol. 130, 645–670.

Lindner, J., Rathjen, F.G., Schachner, M., 1983. L1 mono- and polyclonal antibodies modify cell migration in early postnatal mouse cerebellum. Nature 305, 427–430.

Mason, P.J., Stevens, D.J., Luzzatto, L., Brenner, S., Aparicio, S., 1995. Genomic structure and sequence of the *Fugu rubripes* glucose-6-phosphate dehydrogenase gene (G6PD). Genomics 26, 587–591.

McKusick, V., 1993. Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes. John Hopkins University Press, Baltimore.

Miura, M., Kobayashi, M., Asou, H., Uyemura, K., 1991. Molecular cloning of cDNA encoding the rat neural cell adhesion molecule L1. Two L1 isoforms in the cytoplasmic region are produced by differential splicing. FEBS Lett. 289, 91–95.

Moos, M., Tacke, R., Scherer, H., Teplow, D., Fruh, K., Schachner, M., 1988. Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. Nature 334, 701–703.

Powers, D., 1991. Evolutionary genetics of fish. Adv. Genet. 29, 119–223.

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

Rosenthal, A., MacKinnon, R.N., Jones, D.S., 1991. PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1). Nucl. Acids Res. 19, 5395–5401.

Sadoul, R., Kirchhoff, F., Schachner, M., 1989. A protein kinase activity is associated with and specifically phosphorylates the neural cell adhesion molecule L1. J. Neurochem. 53, 1471–1478.

Sonderegger, P., Rathjen, F.G., 1992. Regulation of axonal growth in the vertebrate nervous system by interactions between glycoproteins belonging to two subgroups of the immunoglobulin superfamily. J. Cell. Biol. 119, 1387–1394.

Staden, R., 1987. Computer handling of DNA sequencing projects. In: Bishop, M., Rawlings, J. (Eds), Nucleic Acids and Sequence Analysis—A Practical Approach. IRL Press, Oxford, pp. 173–217.

Stallcup, W.B., Beasley, L., 1985. Involvement of the nerve growth factor-inducible large external glycoprotein (NILE) in neurite fasciculation in primary cultures of rat brain. Proc. Natl. Acad. Sci. USA 82, 1276–1280.

Trower, M.K., Orton, S.M., Purvis, I.J., Sanseau, P., Riley, J., Christodoulou, C., Burt, D., See, C.G., Elgar, G., Sherrington, R., Rogaev, E.I., St George-Hyslop, P., Brenner, S., Dykes, C.W., 1996. Conservation of synteny between the genome of the pufferfish (*Fugu rubripes*) and the region on human chromosome 14 (14q24.3) associated with familial Alzheimer disease (AD3 locus). Proc. Natl. Acad. Sci. USA 93, 1366–1369.

Van Camp, G., Vits, L., Coucke, P., Lyonnet, S., Schrander-Stumpel, C., Darby, J., Holden, J., Munnich, A., Willems, P.J., 1993. A duplication in the L1CAM gene associated with X-linked hydrocephalus. Nat. Genet. 4, 421–425.

Venkatesh, B., Tay, B.H., Elgar, G., Brenner, S., 1996. Isolation, characterization and evolution of nine pufferfish (*Fugu rubripes*) actin genes. J. Mol. Biol. 259, 655–665.

Vits, L., Van Camp, G., Coucke, P., Wilson, G., Schrader-Stumpel, C., Schwartz, C., Willems, P., 1994. MASA syndrome is due to mutations in the L1CAM gene. Nat. Genet. 7, 408–413.

Williams, E.J., Furness, J., Walsh, F.S., Doherty, P., 1994. Activation of the FGF receptor underlies neurite outgrowth stimulated by L1 N-CAM, and N-cadherin. Neuron 13, 583–594.

Wong, E.V., Kenwrick, S., Willems, P., Lemmon, V., 1995. Mutations in the cell adhesion molecule L1 cause mental retardation. TINS 18, 168–172.

Xu, Y., Mural, R., Shah, M., Uberbacher, E., 1994. Recognizing exons in genomic sequence using GRAIL II. Genet. Eng. NY 16, 241–253.