Int. J. Dev. Biol. 46: 385-391 (2002)

Conserved modularity and potential for alternate splicing in mouse and human *Slit* genes

MELISSA LITTLE*, BREE RUMBALLE[1], KYLIE GEORGAS, TOSHIYA YAMADA and ROHAN D. TEASDALE

Institute for Molecular Bioscience and Centre for Functional and Applied Genomics, University of Queensland, St. Lucia, Australia and [1]Department of Biochemistry, University of Queensland, St. Lucia, Australia

ABSTRACT The vertebrate *Slit* gene family currently consists of three members; *Slit1*, *Slit2* and *Slit3*. Each gene encodes a protein containing multiple epidermal growth factor and leucine rich repeat motifs, which are likely to have importance in cell-cell interactions. In this study, we sought to fully define and characterise the vertebrate *Slit* gene family. Using long distance PCR coupled with *in silico* mapping, we determined the genomic structure of all three *Slit* genes in mouse and man. Analysis of EST and genomic databases revealed no evidence of further *Slit* family members in either organism. All three *Slit* genes were encoded by 36 *(Slit3)* or *37 (Slit1* and *Slit2)* exons covering at least 143 kb or 183 kb of mouse or human genomic DNA respectively. Two additional potential leucine-rich repeat encoding exons were identified within intron 12 of *Slit2*. These could be inserted in frame, suggesting that alternate splicing may occur in *Slit2*. A search for STS sequences within human *Slit3* anchored this gene to D5S2075 at the 5' end (exon 4) and SGC32449 within the 3' UTR, suggesting that *Slit3* may cover greater than 693 kb. The genomic structure of all *Slit* genes demonstrated considerable modularity in the placement of exon-intron boundaries such that individual leucine-rich repeat motifs were encoded by individual 72 bp exons. This further implies the potential generation of multiple Slit protein isoforms varying in their number of repeat units. cDNA library screening and EST database searching verified that such alternate splicing does occur.

KEY WORDS: *leucine-rich repeat, epidermal growth factor repeat, in silico mapping, Slit genes*

Introduction

The *Drosophila slit* gene encodes a large extracellular protein comprising seven epidermal growth factor motifs (EGF1-7), four N-terminal regions of leucine-rich repeat regions (LRR1-4) and a C-terminal cysteine-rich knot. Each of these motifs has been shown to play a role in protein-protein interactions (Rothberg *et al.,* 1990, 1992). Three members of a vertebrate *Slit* gene family have been isolated in human and mouse (Itoh *et al.,* 1998; Holmes *et al.,* 1998; Nakayama *et al.,* 1998). The encoded vertebrate proteins contain the same groups and arrangement of structural motifs as the fly, although they contain two additional EGF motifs. *Drosophila slit* is expressed in the midline glial cells of the developing fly neural tube and acts as a midline chemorepellant thereby facilitating the proper development of axonal scaffolds and commissural events (Rothberg *et al.,* 1990; Kidd *et al.,* 1999). This role appears to have been conserved in vertebrates (Brose *et al.,* 1999; Li *et al.,* 1999). All three of the *Slit* genes discovered to date are expressed in the developing central nervous system in the floor plate, considered to be the vertebrate equivalent of the fly midline glial cells (Itoh *et al.,* 1998; Holmes *et al.,* 1998; Nakayama *et al.,* 1998; Yuan *et al.,* 1999). Vertebrate Slit2 protein has been shown to act as a chemorepellant, as does slit in *Drosophila* (Brose *et al.,* 1999). As in fly, the vertebrate Slit proteins interact with the product of the vertebrate orthologs of the fly roundabout (*robo*) gene to elicit this chemorepulsive cue (Brose *et al.,* 1999, Li *et al.,* 1999). Together with defects in neural patterning,

Abbreviations used in this paper: AMCN, Arthrogryposis Multiplex Congenita, Neurogenic Type; BAC, bacterial artifical chromosome; bp, base pair; cDNA, complementary deoxynucleic acid; cR, centiRad; dbEST, EST database; EGF, epidermal growth factor; ESM, electronic supplementary material; EST, expressed sequence tag; htgs, high throughput genomic sequences; kb, kilobase pairs; LRR, leucine rich repeat; LRRCT, C-terminal conserved region adjacent to leucine rich repeats; LRRNT, N-terminal conserved region adjacent to leucine rich repeats; MCR, minimal critical region; NCBI, National Centre for Biotechnology Information; nr, non-redundant; PAC, P1-derived artificial chromosome; STS, sequence tagged site; TIGR, The Institute for Genomic Research; UTR, untranslated region.

***Address correspondence to:** Dr. M. Little. Institute for Molecular Bioscience, University of Queensland, St. Lucia, 4072, Australia. Fax: +61-7-3365-4388. e-mail: M.Little@imb.uq.edu.au

Electronic Supplementary Material for this paper, entitled *"Intron-exons boundary sequences for each of the three vertebrate* Slit *genes"* is available at the following addresses: http://www.ijdb.ehu.es/abstract.0204/esmX.htm or http://www.imb.uq.edu.au/groups/little/SLIT/

0214-6282/2002/$25.00

Printed in Spain
www.ijdb.ehu.es

Drosophila slit mutants also display defects in mesoderm migration and muscle attachment (Rothberg *et al.,* 1990, 1992). In support of additional non-neuronal functions for vertebrate Slits, we and others have shown distinct expression patterns for the vertebrate *Slit* genes in non-neuronal tissues, including the limbs, skeletal muscle and developing urogenital tract (Itoh *et al.,* 1998; Holmes *et al.,* 1998; Piper *et al.,* 2000). Functional evidence for non-neuronal Slit2 activities have also been demonstrated and include regulation of prechordal mesoderm migration in zebrafish (Yeo *et al.,* 2001), and inhibition of leucocyte chemotaxis in mammals (Wu *et al.,* 2001).

Slit proteins are large, highly modular proteins containing both LRR and EGF repeats. In many other protein families in which multiple extracellular protein-protein interacting motifs are present, alternate splicing results in the generation of protein isoforms with distinct activities upon ligand binding. This has been exemplified by the neural cell adhesion molecules of the immunoglobulin superfamily (Walsh and Doherty, 2001). The aim of this study was to fully characterise the genomic structure of each of the *Slit* genes in human and mouse using long distance PCR and *in silico* mapping, search for additional members of the vertebrate *Slit* gene family and search for evidence of alternate splicing of *Slit* genes. Analysis of the exon-intron boundaries in both mouse and human has revealed that all three vertebrate *Slit* genes are highly conserved and highly modular, indicating considerable potential for alternate splicing resulting in the generation of multiple protein isoforms. *In silico* analysis has provided support for this possibility by revealing the existence of *Slit* transcripts which vary in their motif complement.

Results

In Silico *Genomic Mapping of Human* Slit *Genes*

Initial characterisation of human Slit3 was commenced using long distance reverse transcription-polymerase chain reaction (RT-PCR) on either genomic DNA or Slit3-containing P1-artificial chromosome (PAC) sequences. This indicated that this gene was likely to have many exons with very large introns (data unpublished). Hence, as the human genome sequence became available, an *in silico* approach was adopted for all three known human *Slit* genes. The *in silico* approach involved identifying a number of overlapping genomic sequences for each *Slit* gene (see Electronic Supplementary Material 1) using the sequences described by Itoh *et al.* (1998) for each gene. These were NM_003061 for *Slit1*, NM_004787 for *Slit2* and NM_003062 for *Slit3*. Each unordered contig was broken down into its derivative fragments for recompilation in order according to the exons detected within each fragment.

Slit3 (NM_003062) (Itoh *et al.,* 1998) was found to be encoded by a total of 36 exons encompassing at least 255 kb (see ESM. 1; Fig. 1; Tables 1,2). While the sequence of exon 5 and exon 18 were not encoded by the genomic sequences within the public genome databases they were defined based on the flanking exons. The largest gap was between exons 4 and 5, suggesting an intron of >64kb. While *in silico* mapping predicted the presence of all intron-exon boundaries, the nature of the unordered contigs was such that not all intron lengths were predicted. In the case of introns 16 and 18, the intron length was actually clarified using long distance-PCR (LD-PCR) as 821bp and 3.2 kb respectively. In addition, this confirmed the precise nature of the exon 18 sequence that was not contained in the *in silico* map.

In silico mapping was subsequently performed for *Slit1* (NM_003061) and *Slit2* (NM_004787) (Itoh *et al.,* 1998) as for *Slit3*. This process predicts total genomic lengths of at least 183 kb and 275 kb for *Slit1* and *Slit2* respectively, with each gene consisting of 37 exons. There is remarkable uniformity of exon-intron boundary placement and intron length between all three *Slit* genes (Figs. 1,2; Tables 1,2). The single additional exon present in Slit1 and Slit2 is exon 15 (24bp), which encodes in frame 8 amino acids, these being inserted between two adjacent LRR repeats (see ESM. 1 and Fig. 1). We were unable to detect any sequences equivalent to exon 15 in the *Slit3* genomic contigs. As in *Slit3*, both *Slit1* and *Slit2* genes contain a large gap, 92 kb and >123 kb respectively, within the LRR1-encoding region (between exons 4 and 5) indicating the existence of a very large intronic region (Fig. 1, Table 1). Curiously, within intron

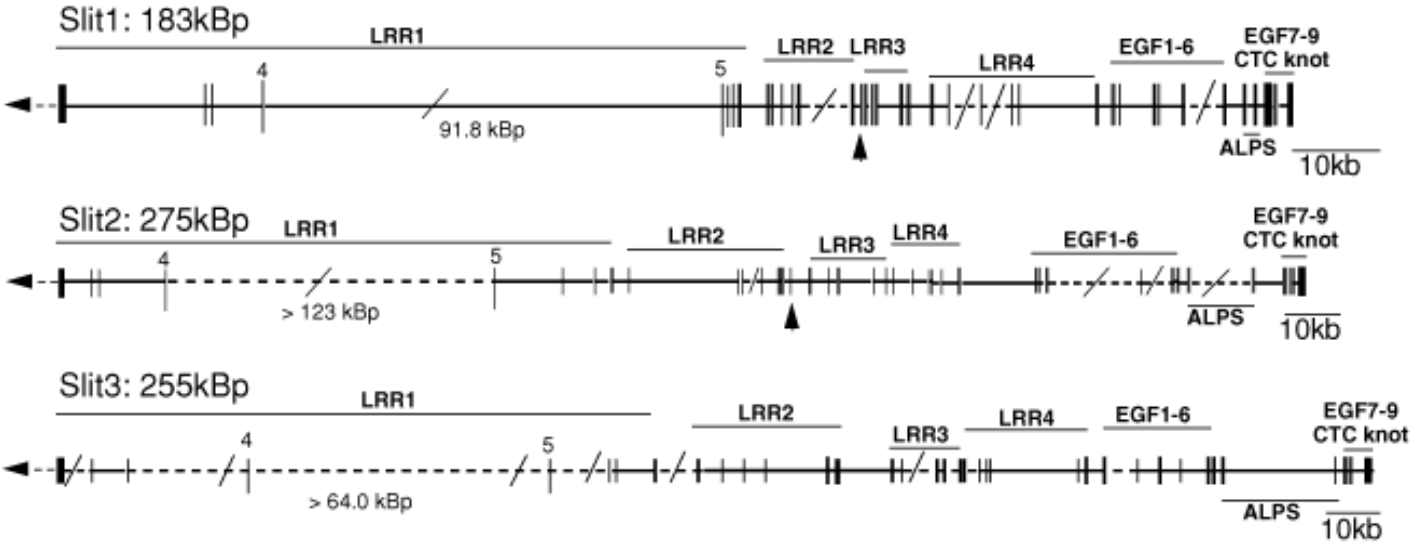

Fig. 1. Genomic maps of the three vertebrate *Slit* genes. *These ideograms indicate the position and size of all introns and exons. Dashed lines refer to introns of unknown length. The minimum total length of each gene is indicated as is the protein motifs encoded by different regions of the gene. Arrows indicate the position of exon 15 which is absent in* Slit3. *A diagonal line indicates sizes of intron may not be accurately to scale.*

TABLE 1

A SUMMARY OF THE GENE STRUCTURE FOR THE THREE VERTEBRATE SLIT GENES

SLIT1								SLIT2								SLIT3					
	Exon length	mRNA position	Intron length	AL512424	AL442123	AC024551	AC356126		Exon length	mRNA position	Intron length	AC012142	AL136447	AC013310	AC356126	AC021118		Exon length	mRNA position	Intron length	AC008479	AC011365
Exon 1	429	(0-429)	20607					Exon 1	383	(0-383)	2678						Exon 1	460	(0-460)	>6094		
Exon 2	72	(430-501)	1366					Exon 2	72	(384-455)	1125						Exon 2	72	(461-532)	6609		
Exon 3	72	(502-573)	5492					Exon 3	72	(456-527)	10871						Exon 3	72	(533-604)	>9332		
Exon 4	72	(574-645)	91849					Exon 4	72	(528-599)	>123167						Exon 4	72	(605-676)	>64012		
Exon 5	72	(646-717)	1125					Exon 5	72	(600-671)	12565						Exon 5	72	(677-748)	>15919		
Exon 6	72	(718-789)	566					Exon 6	72	(672-743)	5413						Exon 6	72	(749-820)	>1482		
Exon 7	72	(790-861)	550					Exon 7	72	(744-815)	2548						Exon 7	72	(821-892)	5797		
Exon 8	164	(862-1025)	2668					Exon 8	164	(816-979)	2779						Exon 8	164	(893-1056)	>7660		
Exon 9	148	(1026-1173)	482					Exon 9	139	(980-1118)	18734						Exon 9	142	(1057-1198)	4375		
Exon 10	72	(1174-1245)	556					Exon 10	72	(1119-1190)	71						Exon 10	72	(1199-1270)	5882		
Exon 11	72	(1246-1317)	2176					Exon 11	72	(1191-1262)	>3846						Exon 11	72	(1271-1342)	3605		
Exon 12	72	(1318-1389)	746					Exon 12	72	(1263-1334)	4309						Exon 12	72	(1343-1414)	11540		
Exon 13	144	(1390-1533)	>7267					Exon 13	144	(1335-1478)	111						Exon 13	144	(1415-1558)	1289		
Exon 14	164	(1534-1697)	967					Exon 14	164	(1479-1642)	972						Exon 14	164	(1559-1722)	10100		
Exon 15	24	(1698-1721)	139					Exon 15	24	(1643-1666)	3779											
Exon 16	145	(1722-1866)	625					Exon 16	151	(1667-1817)	2891						Exon 15	151	(1723-1873)	1601		
Exon 17	75	(1867-1941)	194					Exon 17	75	(1818-1892)	1509						Exon 16	75	(1874-1948)	>6113		
Exon 18	144	(1942-2085)	3141					Exon 18	144	(1893-2036)	5732						Exon 17	144	(1949-2092)	>821		
Exon 19	144	(2086-2229)	303					Exon 19	144	(2037-2180)	1869						Exon 18	144	(2093-2236)	3199		
Exon 20	167	(2230-2396)	2781					Exon 20	167	(2181-2347)	875						Exon 19	164	(2237-2400)	1036		
Exon 21	133	(2397-2529)	2222					Exon 21	133	(2348-2480)	3399						Exon 20	133	(2401-2533)	>5985		
Exon 22	69	(2530-2598)	>1229					Exon 22	69	(2481-2549)	2389						Exon 21	69	(2534-2602)	1410		
Exon 23	72	(2599-2670)	2790					Exon 23	72	(2550-2621)	569						Exon 22	72	(2603-2674)	604		
Exon 24	72	(2671-2742)	778					Exon 24	72	(2622-2693)	1702						Exon 23	72	(2675-2746)	9904		
Exon 25	72	(2743-2814)	9344					Exon 25	72	(2694-2765)	2979						Exon 24	72	(2747-2818)	1223		
Exon 26	164	(2815-2978)	2143					Exon 26	164	(2766-2929)	13298						Exon 25	164	(2819-2982)	2793		
Exon 27	125	(2979-3103)	300					Exon 27	125	(2930-3054)	132						Exon 26	125	(2983-3107)	>5905		
Exon 28	98	(3104-3201)	4428					Exon 28	98	(3055-3152)	1250						Exon 27	98	(3108-3205)	4149		
Exon 29	140	(3202-3341)	185					Exon 29	140	(3153-3292)	>15408						Exon 28	140	(3206-3345)	3594		
Exon 30	94	(3342-3435)	2611					Exon 30	94	(3293-3386)	>6950						Exon 29	94	(3346-3439)	5497		
Exon 31	138	(3436-3573)	>3403					Exon 31	138	(3387-3524)	581						Exon 30	138	(3440-3577)	1050		
Exon 32	238	(3574-3811)	2380					Exon 32	241	(3525-3765)	1611						Exon 31	241	(3578-3818)	1589		
Exon 33	131	(3812-3942)	471					Exon 33	131	(3766-3896)	>8481						Exon 32	131	(3819-3949)	10660		
Exon 34	155	(3943-4097)	1076					Exon 34	155	(3897-4051)	>4483						Exon 33	155	(3950-4104)	1696		
Exon 35	289	(4098-4386)	335					Exon 35	289	(4052-4341)	241						Exon 34	286	(4105-4390)	1213		
Exon 36	212	(4387-4598)	1019					Exon 36	212	(4342-4553)	1119						Exon 35	209	(4391-4599)	3302		
Exon 37	495	(4599-5094)						Exon 37	382	(4554-4935)							Exon 36	415	(4600-5015)			
		Total length +cDNA	183403							Total length +cDNA	275371								Total length +cDNA	254686		

Intron placement, exon length and the genomic contigs (shaded) used for the in silico mapping are indicated. The total length of each gene includes the length of the open reading frame (ORF).

12 of Slit2, there are two additional exons very similar to exons 13 and 14. We have termed these 'exon13like' and 'exon14like'. If these sequences were included in a Slit2 mRNA, they would encode an additional two LRR repeats within LRR2 in frame (ESM. 1).

The *Slit1* positive genomic sequences (AL442123, AL356126 and AL442123) are annotated as being localised to Chromosome 10 which is consistent with the published chromosomal localisation for *Slit1* (Nakayama *et al.,* 1998). Likewise, *Slit2* positive genomic sequences (AC046194, AC012142 and AC02118) annotated to be located on Chromosome 4 correlates to the published chromosomal localisation of *Slit2*, which characterised a localisation of human 4p15.2 (Georgas *et al.,* 1999). It was of note that the genomic sequences AL024551 (*Slit1* 5' exons) and AL136447 (*Slit2* 5' exons) appear to have incorrect annotation of their chromosomal localisation. For example, AL024551 is annotated as been localised on chromosome 4 but large segments (>10 kb) are greater than 99% identical to the sequence AL512424 which is annotated to be located on Chromosome 10 suggesting that AL024551 has been incorrectly assigned.

Comparison with Gene Structure Predictions from the Draft Human Genome

Next we compared the assembled SLIT genes with those generated as part of the analysis of the human genome project. We identified the genes annotated as SLIT1 (ENSG00000052758); SLIT2 (ENSG00000067745); SLIT3 (ENSG00000094749) in Ensembl v 1.0 (http://www.ensembl.org/ June 2001). As anticipated, comparison of our SLIT gene contigs to those within Ensembl revealed numerous discrepancies. Our assembled genes contained more sequence detail, particularly within the larger introns. Surprisingly, given that the cDNA for each SLIT is known, each of the Ensembl predicted transcripts contained inaccuracies that resulted in the disruption of the protein open-reading frame. The following discrepancies were identified within the predicted exons. *SLIT1*:- 1) Incorrect splice donor sequence for exon 29; 2) Incorrect splice acceptor sequence for exon 30. *SLIT2*:- 1) Incorrect splice donor sequence for exon 14; 2) Incorrect splice donor sequence for exon 33; 3) Exclusion of exon 34; 4) Inclusion of unrelated sequence as exon 33; 5) Fragmented exon 35 into two separated exons. SLIT3:- 1) Fragmented exon 1 into three separated exons with gaps; 2) Inclusion of unrelated sequence as exon 4; 3) Incorrect splice acceptor sequence for exon 2; 4) Incorrect splice acceptor sequence for exon 6; 5) Incorrect splice acceptor sequence for exon 19. This would suggest that gene structure prediction algorithms are flawed without parallel analysis of known cDNA and protein information and that focussed *in silico* analyses such as ours continue to more accurately predict gene structure.

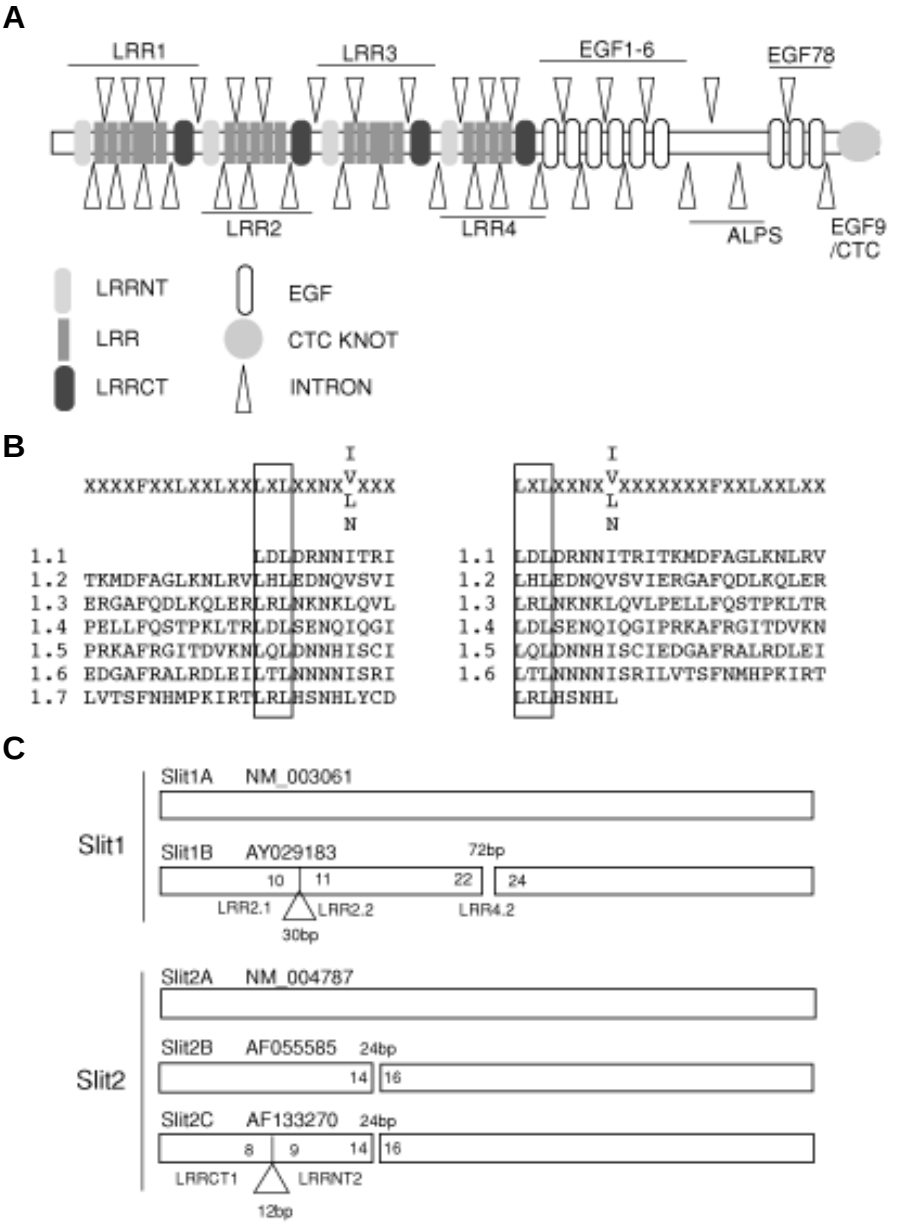

Fig. 2. Modularity within human *Slit* genes. (A) *Generic ideogram of a Slit protein, based upon Slit3, demonstrating the relationship between each structural motif and placement of introns.* **(B)** *Redefinition of the LRR motif based upon intron placement in* Slit *genes. The LRR motifs from within the first group of LRRs in human Slit3 is shown. The left alignment is according to that previously used by Taguchi* et al. *(1996). The right alignment reflects a redefinition of the LRR based upon the observation of uniform intron placement at the LXL of each repeat.* **(C)** *Novel* Slit1 *and* Slit2 *gene alternately spliced transcripts determined via cDNA library screening or analysis of EST databases. Triangles refer to insertions and gaps refer to deletions with respect to the A transcript. Size of inserts and deletions is indicated in bp with the flanking exons marked within the ideogram and the motifs they encode indicated below.*

Analysis of the Human Genome for Additional Slit Protein Family Members

Bioinformatics was also used to investigate the possibility that there are further undescribed Slit family members. Initially, the public human genome database was analysed for sequences with significant relationship to the three human Slit proteins. This failed to detect any additional predicted gene or genomic sequence related to the currently defined human Slit proteins other than those already known. In addition, analysis of the Celera Genomics and public human expressed sequence tag (EST) databases failed to detect any further mouse or human family members. Therefore, within the currently available human and mouse DNA sequences, the Slit family appears to only include the three previously described human members.

In Silico *Genomic Mapping of Murine* Slit *Genes*

To determine if the genomic structure of the murine SLIT genes showed the same conserved modularity as the human genes, these genes were identified in the assembled mouse genome generated by Celera Genomics (www.celera.com). The genomic lengths of these genes, based on the Celera assembly, is predicted to be at least 143kb, 330 kb and 530 kb for murine *Slit1, Slit2* and *Slit3* respectively. Each gene consisted of the same number of exons as the human counterpart and each contained a large intron between exons 4 and 5. With the exception of exon 32 from murine Slit1, which contained an additional nine nucleotides, each exon in the three murine SLIT genes was identical in length and position relative to the coding sequence as the human Slit genes. While the region is incomplete, analysis of the intronic region between exon 12 and exon 13 of murine Slit2 genomic sequences failed to detect the two putative exons present in the human gene.

Modularity of Genomic Structure within the LRR-Encoding Regions

The leucine rich repeat regions of the three human *Slit* genes contain a large number of very small exons. In most instances, individual leucine rich repeats of 24 amino acids were encoded by individual exons (72bp) (ESM. 1, Fig. 2A, Table 1). This highly modular structure would allow individual repeat motifs to be alternately spliced in or out within altering the frame. For example, each of the four blocks of leucine-rich repeats could be inserted or removed without altering the rest of the amino acid sequence. Indeed, in most instances the individual leucine-rich motifs within each block of repeats have the potential to be removed or inserted without altering frame. This implies that these genes may be highly alternately spliced. LRRs are structurally composed of a beta-alpha unit (Kobe and Deisenhofer, 1994) forming an elongated non-globular structure classically involved in protein-protein interactions. The Pfam website would identify the protein sequence XLXXLXLXXN Xn I Xn FXXL as an LRR. However, Taguchi *et al.* (1996), when comparing comparing novel LRR

proteins with those previously isolated, aligned the leucine-rich repeat motif as XXXX(F/L)XXLXXLXXLXXLXXNX(IVLN)XX(I/M) (Fig. 2B). Here we demonstrate that LRR-encoding exons always started on the second position base of the codon encoding the first leucine of the LXL (as underlined). We have therefore redefined the LRRs as a repeat unit of LXLXXNXXIX$_7$FXXLXXLXXLXX (see Fig. 2B). In contrast to the LRRs, there is approximately 1 exon encoding each EGF repeat. However, these exons vary in length such that loss of one such exon would usually shift the frame resulting in premature termination of the protein.

Evidence for Alternate Splicing resulting in Multiple Slit Protein Isoforms

Given the provocative intron placement in the 5' end of these genes and the existence of additional putative exons within intron 12 of *Slit2*, we searched for evidence that alternate splicing does occur. This was approached by screening a human fetal brain cDNA library using a combination of human *Slit* gene cDNA probes. cDNA library screening resulted in the recloning of multiple partial transcripts, with most sequences been identical to the original either *Slit1*, *Slit2* or *Slit3* mRNAs. However, one clone, which we term *Slit1B*, aligned with Slit1 except for the deletion of exon 23 encoding LRR4.2 (Fig. 2C). Additionally, *Slit1B* contained a 30bp insertion between *Slit1* exons 10 and 11 which is present within the genomic sequence of *Slit1* (NM_003061). This inserted 10 amino acids in frame between LRR2.1 and LRR2.2, according to our revised description of an LRR. This insertion was generated by the use of an alternate splice acceptor site within the intron preceding exon 11. Normally, the intron/exon junction at the 5' end of exon 11 is ccaccttgcag/AGACCTG, demonstrating a characteristic splice acceptor site of ^{6}Pyncag/N (Lewin, 2000). In this clone cDNA, an alternate site of agggattcag/ACG was employed insert-

TABLE 2

DEFINITION OF SLIT PROTEIN STRUCTURE

Pfam Domain	SLIT1			SLIT2			SLIT3		
	Open-Reading Frame			Open-Reading Frame			Open-Reading Frame		
LRRNT 1	Exon 1	66	(1-65.67)	Exon 1	60	(0-59.67)	Exon 1	66	(0-65.67)
LRR 1.1	Exon 2	24	(66-89.67)	Exon 2	24	(60-83.67)	Exon 2	24	(66-89.67)
LRR 1.2	Exon 3	24	(90-113.67)	Exon 3	24	(84-107.67)	Exon 3	24	(90-113.67)
LRR 1.3	Exon 4	24	(114-137.67)	Exon 4	24	(108-131.67)	Exon 4	24	(114-137.67)
LRR 1.4	Exon 5	24	(138-161.67)	Exon 5	24	(132-155.67)	Exon 5	24	(138-161.67)
LRR 1.5	Exon 6	24	(162-185.67)	Exon 6	24	(156-179.67)	Exon 6	24	(162-185.67)
LRR 1.6	Exon 7	24	(186-209.67)	Exon 7	24	(180-203.67)	Exon 7	24	(186-209.67)
LRRCT 1	Exon 8	54.67	(210-264.33)	Exon 8	54.67	(204-258.33)	Exon 8	54.67	(210-264.33)
LRRNT 2	Exon 9	49.33	(264.67-313.67)	Exon 9	46.33	(258.67-304.67)	Exon 9	47.33	(264.67-311.67)
LRR 2.1	Exon 10	24	(314-337.67)	Exon 10	24	(305-328.67)	Exon 10	24	(312-335.67)
LRR 2.2	Exon 11	24	(338-361.67)	Exon 11	24	(329-352.67)	Exon 11	24	(336-359.67)
LRR 2.3	Exon 12	24	(362-385.67)	Exon 12	24	(353-376.67)	Exon 12	24	(360-383.67)
LRR 2.4/2.5	Exon 13	48	(386-433.67)	Exon 13	48	(377-424.67)	Exon 13	48	(384-431.67)
LRRCT 2	Exon 14	54.67	(434-488.33)	Exon 14	54.67	(425-479.33)	Exon 14	54.67	(432-486.33)
	Exon 15	8	(488.67-496.33)	Exon 15	8	(479.67-487.33)			
LRRNT 3	Exon 16	48.33	(496.67-544.67	Exon 16	50.33	(487.67-537.67)	Exon 15	50.33	(486.67-536.67)
LRR 3.1	Exon 17	25	(545-569.67	Exon 17	25	(538-562.67)	Exon 16	25	(537-561.67)
LRR 3.2/3.3	Exon 18	48	(570-617.67)	Exon 18	48	(563-610.67)	Exon 17	48	(562-609.67)
LRR 3.4/3.5	Exon 19	48	(618-665.67)	Exon 19	48	(611-658.67)	Exon 18	48	(610-657.67)
LRRCT 3	Exon 20	55.67	(666-721.33)	Exon 20	55.67	(659-714.33)	Exon 19	54.67	(658-712.33)
LRRNT 4	Exon 21	44.33	(721.67-765.67)	Exon 21	44.33	(714.67-758.67)	Exon 20	44.33	(712.67-756.67)
LRR 4.1	Exon 22	23	(766-788.67)	Exon 22	23	(759-781.67)	Exon 21	23	(757-779.67)
LRR 4.2	Exon 23	24	(789-812.67)	Exon 23	24	(782-805.67)	Exon 22	24	(780-803.67)
LRR 4.3	Exon 24	24	(813-836.67)	Exon 24	24	(806-829.67)	Exon 23	24	(804-827.67)
LRR 4.4	Exon 25	24	(837-860.67)	Exon 25	24	(830-853.67)	Exon 24	24	(828-851.67)
LRRCT 4	Exon 26	54.67	(861-915.33)	Exon 26	54.67	(854-908.33)	Exon 25	54.67	(852-906.33)
EGF 1	Exon 27	42	(915.67-957.33)	Exon 27	41.67	(908.67-950)	Exon 26	41.67	(906.67-948)
EGF 2	Exon 28	32.33	(957.67-989.67)	Exon 28	32.67	(950.33-982.67)	Exon 27	32.67	(948.33-980.67)
EGF 3	Exon 29	46.67	(990-1036.33)	Exon 29	46.67	(983-1029.33)	Exon 28	46.67	(981-1027.33)
EGF 4	Exon 30	31.33	(1036.67-1067.67)	Exon 30	31.33	(1029.67-1060.67)	Exon 29	31.33	(1027.67-1058.67)
EGF 5	Exon 31	46	(1068-1113.67)	Exon 31	46	(1061-1106.67)	Exon 30	46	(1059-1104.67)
EGF 6	Exon 32	79.33	(1114-1193)	Exon 32	80.33	(1107-1187)	Exon 31	80.33	(1105-1185)
Laminin_G	Exon 33	43.67	(1193.33-1236.67)	Exon 33	43.67	(1187.33-1230.67)	Exon 32	43.67	(1185.33-1228.67)
	Exon 34	51.67	(1237-1288.33)	Exon 34	51.67	(1231-1282.33)	Exon 33	51.67	(1229-1280.33)
EGF 7	Exon 35	96.33	(1288.67-1384.67)	Exon 35	96.67	(1282.67-1379)	Exon 34	95.33	(1280.67-1375.67)
EGF 8/9	Exon 36	70.67	(1385-1455.33)	Exon 36	70.67	(1379.33-1449.67)	Exon 35	69.67	(1376-1445.33)
CTC-KNOT	Exon 37	78.33	(1455.67-1534)	Exon 37	84	(1450-1577)	Exon 36	78.33	(1445.67-1524)

The relationship between each exon and the amino acids and structural motifs encoded for each of the three vertebrate *Slit* genes is tabulated. LRR, leucine-rich repeat; LRRNT, N-terminal conserved region from within a leucine rich repeat domain; LRRCT, C-terminal conserved region from within a leucine rich repeat domain; laminin_G, agrin-laminin-parlecan-slit domain; EGF, epidermal growth factor repeat; CTC-KNOT, cysteine rich knot motif. LRR domains were detected using the criteria defined in this manuscipt (see text for details). Other structural motifs were as predicted by Pfam (http://pfam.wustl.edu).

ing the amino acids RPLSFCSPCR within the predicted protein. The function of this insertion is unknown. Slit1B has been submitted to Genbank (AY029183).

Nucleotide database searching was also employed in an attempt to identify additional alternatively spliced forms of the three Slit proteins. Searches were carried out on both the nr nucleotide database at NCBI and various EST databases located including dbEST, Unigene and TIGR Human and Mouse gene indexes. The majority of sequences identified matched with the original mRNAs. However, evidence for alternate splicing was found. The full-length cDNA for *Slit2* characterised by Holmes et.al (1998) (AF055585) varies from that described by Itoh *et al.* (1998) (NM_004787) in that it does not contain the small exon 15 that encodes for eight amino acids of unknown function located between the second LRRCT and third LRRNT domains. Therefore the sequences at this position not only varies between the Slit proteins but can vary within the individual Slits. We redefine these alternative transcripts as *Slit2A* (NM_004787) and *Slit2B* (AF055585) (Fig. 2C). An additional alternative full-length cDNA for *Slit2*, which we term *Slit2C*, was described by Brose *et al.* (1999) (AF133270). *Slit2C*, like Slit2B, is missing exon 15, but also contains an additional 12bp exon located within the intron between exon 8 and exon 9 of *Slit2A* (Fig. 2C). This results in an insertion of the four amino acids between the first LRRCT domain and the second LRRNT. Analysis of all human Slit3 ESTs did not detect any alternatively spliced forms that varied from NM_003062. However, the partial *Slit3* cDNA AF075240, representing the fully sequenced EST 21651 from human brain (Holmes *et al.,* 1998), contains a 300bp deletion encompassing part of exons 31 and 32. This results in the in frame deletion of 100 amino acids, including one EGF repeat (EGF 6), corresponding to residues 1117-1217 of *Slit3* (NM_003062). There are no apparent cryptic slice donor or acceptor sites to explain this event. Hence, this is unlikely to be an alternative splice variant, but may represent a recombination event either in the genomic context or during the plasmid replication in the bacterium, between the sequence tggtc in each exon. Additional cDNAs containing variant nucleotide sequences that do not maintain frame were also detected for the three *Slit* genes. These included cDNAs that contained intact introns, presumably representing pre-mRNA. Several variant cDNAs were also detected that could not be readily explained by alternative splicing events. These included deletions within individual exons and the duplication of partial exons within the one cDNA. Further characterisation of these cDNAs would be required to confirm they occur *in vivo*. No additional alternative splice variants were detected in the mouse nucleotide databases.

Discussion

In this study we have described the genomic structure of all three human *Slit* genes, demonstrated that the conserved modularity can result in alternate splicing, have used this to redefine the LRR motif and have found no evidence for the existence of additional human *Slit* gene family members. This study emphasizes the power of *in silico* mapping, even for genes of this size, and demonstrates that careful analysis of gene families such as this with knowledge about the predicted protein can generate more reliable information than that generated in exon prediction databases such as Ensembl.

Chromosomal Orientation and Implication for AMCN

Nakayama *et al.* (1998) have demonstrated that *Slit1* maps to human chromosome 10q24. We have previously reported the mapping of *Slit2* to human chromosome 4p15.2, distal to the *Msx1* gene (Georgas *et al.,* 1999). *Slit3* has been localized to human chromosome 5q35.5, distal to the *Msx2* gene (Nakayama *et al.,* 1998), implying an evolutionary relationship between chromosome 4p and 5q. The localisation of *Slit3* is also close to the minimal critical region for Arthrogryposis Multiplex Congenita, Neurogenic Type (AMCN), a mild motor neuron disorder characterised by rigidity of the joints with the basic defect being one of neuronal degeneration in the anterior horn (Shohat *et al.,* 1997). During the characterisation of the human *Slit3* gene, a search for sequence tagged sites (STS) anchored *Slit3* to D5S2075 at the 5' end (exon 4) and SGC32449 at the 3' end (UTR), suggesting that *Slit3* may cover greater than 693kBp. This also indicates that the *Slit3* gene lies in a telomeric to centromeric direction (5' to 3'). The minimal critical region for AMCN has only been roughly defined as lying between D5S1456 and D5S498 (Shohat *et al.,* 1997), within which D5S2075 and SGC32449 lie. However, this MCR is estimated to cover 24cR or around 7.1Mbp (as indicated by the MIT chromosome 5 map). Hence, further analysis will be required to define whether there is an involvement of *Slit3* in this condition.

Materials and Methods

In Silico *Mapping*

Using full length *Slit* gene cDNA sequences (SLIT1 (NM_003061); SLIT2 (NM_004787); SLIT3 (NM_003062)), BlastN analyses were performed searching either the unfinished High Throughput Genomic Sequences (htgs) or the completed genomic sequences (McPherson *et al.,* 2001) using the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/BLAST/). Each human genomic BAC clone identified to encode for an exon from one of the SLIT genes was further analysed. Initially the unordered genomic sequences were fragmented into their constituent segments. Contiguous sequences from all the *Slit* positive BAC clones were then assembled using Sequencher™ 3.0 (Gene Codes Corporation, Ann Arbor, MI) and manually verified. Exact positioning of exon/intron boundaries was verified manually using the consensus splice donor and acceptor sequences (Lewin, 2000). Identification of the murine genomic sequences for each full length murine *Slit* gene cDNA sequences (SLIT1 (AF144627); SLIT2 (AF144628); SLIT3 (AF144629) was performed using BlastN analyses on the Mouse Genome database using the Celera Discovery System™ [#].

Screening for New Slit *Genes and Alternate Splice Forms*

To search for evidence of additional human or murine *Slit* genes, the corresponding genome sequences and EST databases were searched using TBlastN with the full-length Slit protein. To search for evidence for alternate splicing events, the nr (http://www.ncbi.nlm.nih.gov/BLAST/; GenBank+EMBL+DDBJ+PDB sequences), dbEST (http://www.ncbi.nlm.nih.gov/BLAST/), TIGR gene index (http://www.tigr.org/tdb/hgi/) and UniGene (http://www.ncbi.nlm.nih.gov/UniGene/) were analysed for sequences that varied from the original *Slit* cDNA sequence used above. In addition, a Clontech human fetal brain cDNA library was screened with a partial human *Slit2* cDNA clone at low stringency (60°C) in

an attempt to simultaneously isolate multiple *Slit* family members. All isolated cDNAs and in *silico* identified alternative splice events were compared to the parental *Slit* cDNA sequences (see above) using the Blast2sequences tool. The presence of genomic (intronic) sequences was detected by comparison to the assembled genomic contigs generated above.

Chromosomal Orientation

To determine the orientation of *Slit3* on human chromosome 5, each fragment from within the *Slit3*-containing genomic BACs was compared to the dbSTS using BlastN. The resulting STS sequences from each end of Slit3 were searched for using Entrez / Gene Maps / Homo sapiens (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez) and simultaneously viewing the STS, cytogenetic, radiation-hybrid and morbid maps.

Acknowledgements

This work was funded by the National Health and Medical Research Council of Australia (Grant ID. 990605). ML is a Sylvia and Charles Viertel Senior Research Fellow. The Centre for Functional and Applied Genomics is a Special Research Centre of the Australian Research Council. In memory of our colleague TY, 1960-2001.

Electronic Supplementary Material for this paper, entitled *"Intron-exons boundary sequences for each of the three vertebrate* Slit *genes"* is available at the following address: http://www.ijdb.ehu.es/abstract.0204/esmX.htm or http://www.imb.uq.edu.au/groups/little/SLIT/

This data was generated through use of the Celera Discovery System and Celera's associated databases.

References

BROSE, K., BLAND, K. S., WANG, K.H., ARNOTT, D., HENZEL, W., GOODMAN, C.S., TESSIER-LAVIGNE, M. and KIDD, T. (1999) Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell* 96,795-806

GEORGAS, K., BURRIDGE, L., SMITH, K., HOLMES, G.P., CHENEVIX-TRENCH, G., IOANNOU, P.A. and LITTLE, M.H. (1999) Assignment of the human slit homologue *SLIT2* to human chromosome band 4p15.2. *Cytogenet. Cell Genet.* 86: 246-7.

HOLMES, G.P., NEGUS, K., BURRIDGE, L., RAMAN, S., ALGAR, E., YAMADA, T. and LITTLE, M.H. (1998) Distinct but overlapping expression patterns of two vertebrate *slit* homologs implies functional roles in CNS development and organogenesis. *Mech. Dev.* 79: 57-72.

ITOH, A., MIYABAYASHI, T., OHNO, M. and SAKANO, S. (1998) Cloning and expressions of three mammalian homologues of *Drosophila slit* suggest possible roles for Slit in the formation and maintenance of the nervous system. *Mol Brain Res.* 20: 175-86.

KIDD, T., BLAND, K.S. and GOODMAN, C.S. (1999) Slit is the midline repellent for the robo receptor in *Drosophila*. Cell 96: 785-94.

KOBE, B. and DEISENHOFER, J. (1994) The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* 19: 415-421.

LEWIN, B. (2000) Genes VII. Oxford University Press.

LI, H.S., CHEN, J.H., WU, W., FAGALY, T., ZHOU, L., YUAN, W., DUPUIS, S., JIANG, Z.H., NASH, W., GICK, C., ORNITZ, D.M., WU, J.Y. and RAO, Y. (1999) Vertebrate slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons. *Cell* 96: 807-18.

MCPHERSON, J.D., MARRA, M., HILLIER, L., WATERSTON, R.H., CHINWALLA, A., WALLIS, J., et al. (2001) *Nature.* 409: 934-41.

NAKAYAMA, M., NAKAJIMA, D., NAGASE, T., NOMURA, N., SEKI, N., and OHARA, O. (1998) Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening. Genomics 51: 27-34.

PIPER, M., GEORGAS, K., YAMADA, T. and LITTLE, M. (2000) Expression of the vertebrate *Slit* Gene family and their putative receptors, the *Robo* genes, in the developing murine kidney. *Mech Dev.* 94: 213-217.

ROTHBERG, J.M., JACOBS, J.R., GOODMAN, C.S. and ARTAVANIS-TSAKONAS, S. (1990) slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes. Dev.* 4: 2169-87.

ROTHBERG, J.M. and ARTAVANIS-TSAKONAS, S. (1992) Characterisation of a conserved carboxy-terminal sequence in secreted proteins and a motif implicated in extracellular protein interactions. *J. Mol. Biol.* 227: 367-370.

SHOHAT, M., LOTAN, R., MAGAL, N., SHOHAT, T., FISCHEL-GHODSIAN, N., ROTTER J.L. and JABER L. (1997) A gene for arthrogryposis multiplex congenita neuropathic type is linked to D5S394 on chromosome 5qter. *Am J Hum Genet* 61: 1139-43.

TAGUCHI, A., WANAKA, A., MORI, T., MATSUMOTO, K., IMAI, Y., TAGAKI, T and TOHYAMA, M. (1996) Molecular cloning of novel leucine-rich repeat proteins and their expression in the developing mouse nervous system. *Mol. Brain Res.* 35: 31-40.

WALSH, F.S. and DOHERTY, P. (2001) Neural cell adhesion molecules: role in axon growth and guidance. *Ann. Rev. Cell Dev. Biol.* 13: 425-456.

WU, J.Y., FENG, L., PARK, H.T., HAVLIOGLU, N., WEN, L., TANG, H., BACON, K.B., JIANG, Z.H., ZHANG, X.C. and RAO, Y. (2001) The neuronal repellent Slit inhibits leukocyte chemotaxis induced by chemotactic factors. *Nature* 410: 948-52.

YEO, S.Y., LITTLE, M.H., YAMADA, T., MIYASHITA, T., HALLORAN, M.C., KUWADA, J.Y., HUH, T.L. and OKAMOTO, H. (2001) Overexpression of a slit homologue impairs convergent extension of the mesoderm and causes cyclopia in embryonic zebrafish. *Dev Biol.* 230: 1-17.

YUAN, W., ZHOU, L., CHEN, J.H., WU, J.Y., RAO, Y. and ORNITZ, D.M. (1999) The mouse SLIT family: secreted ligands for ROBO expressed in patterns that suggest a role in morphogenesis and axon guidance. *Dev. Biol.* 21: 290-306.