Hum Genet (2000) 106 : 227–235
Digital Object Identifier (DOI) 10.1007/s004399900217

ORIGINAL INVESTIGATION

**Ida Biunno · Loris Bernard · Paul Dear
Monica Cattaneo · Sara Volorio · Laura Zannini
Alan Bankier · Massimo Zollo**

SEL1L, the human homolog of *C. elegans sel-1*: refined physical mapping, gene structure and identification of polymorphic markers

Received: 14 September 1999 / Accepted: 22 November 1999 / Published online: 26 January 2000

Abstract We have cloned the human full-length cDNA *SEL1L*, which is highly similar to the *C. elegans sel-1* gene, an important negative regulator of the "notch" pathway which acts as a key regulator of the cellular proliferation and specification processes in both vertebrates and invertebrates. The *SEL1L* gene maps to 14q24.3–31 and here we report its fine localization by HAPPY mapping, which determines its molecular distance to microsatellite markers isolated in the region. We have found two new polymorphic (CA)n microsatellites located in the gene, and have identified the exon-intron boundaries. The gene is composed of 21 exons spanning 70 kb of genomic DNA. Human SEL1L protein exhibits a high degree of similarity compared to the mouse and nematode homologs.

Introduction

LIN-12/Notch-like proteins (Adam et al. 1998; Artavanis-Tsakonas et al. 1995; Crowe et al. 1998; Fleming 1998; Gray et al. 1999; Heitzler and Simpson 1991) act as receptors involved in intercellular signalling and the determination of cell fate during development (Adam et al. 1998; Lewis 1998). In addition, they mediate inductive or lateral cell-cell interactions (Greenwald 1998). To date, negative regulators of lin-12 activity have been described in *C. elegans*. The best characterized are sel-10 and sel-1. Sel-10 (Hubbard et al. 1997; Sundaram and Greenwald 1993) is a member of the F-box/WD40 repeat family of proteins, which has been shown to be involved in ubiquitin dependent proteolysis of beta-catenin and IkappaB proteins (Hatakeyama et al. 1999; Kitagawa et al. 1999). Sel-1 protein acts as a suppressor of lin-12 and glp-1 activities in *C. elegans* (Grant and Greenwald 1996; Grant and Greenwald 1997; Sundaram and Greenwald 1993). Interestingly, sel-1 protein (Grant and Greenwald 1997) contains a 38 amino acid motif in its hydrophobic carboxy terminal region, which it shares with the *Saccharomyces cerevisiae* Hrd3 protein (Hampton et al. 1996).

Hrd3 is essential in sterol synthesis and is one of several HRD-encoded proteins required for the degradative activity of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral membrane protein acting in the yeast endoplasmic reticulum.

We have previously reported the isolation of a partial cDNA from the human gene *SEL1L* (Biunno et al. 1997) whose nucleotide and amino acid sequences are homologous to the most carboxy terminal region of sel-1 protein.

SEL1L is ubiquitously expressed in human foetal tissues, but it exhibits high mRNA levels only in adult pancreas (Biunno et al. 1997) and in islets of Langerhans (insulin-producing β cells) (Donoviel et al. 1998); mRNA levels in other adult tissues are low or undetectable. Donoviel et al. (1998) and the present report show that the murine homolog, *m-SEL-1L*, is particularly expressed in the neural tube and in dorsal root ganglia, in the floor plate and in the acini of the pancreas in mouse embryos at stage E10.5–E11.5.

SEL1L resides on human chromosome 14q24.3-q31, a region linked to an insulin-dependent diabetes mellitus locus, *IDDM 11* (OMIM 601208; Field et al. 1996) centred around the microsatellite marker D14S67. The IDDM 11 disorder appears to act primarily in diabetic families in which there is less evidence of susceptibility to the major genetic determinant HLA. However, a genome-wide scan

The first two authors contributed equally to the work.

I. Biunno · M. Cattaneo · L. Zannini
Istituto Tecnologie Biomediche Avanzate del CNR,
Via F.lli Cervi 93-20090 Segrate, Milano, Italy

L. Bernard · S. Volorio · M. Zollo (✉)
Telethon Institute of Genetics and Medicine (TIGEM),
San Raffaele Biomedical Science Park,
Via Olgettina 58, Milano, Italy
e-mail: zollo@tigem.it,
Tel.: +39-02-21560208, Fax: +39-02-21560220

P. Dear · A. Bankier
MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

Electronic supplementary material to this paper can be obtained electronically by using the Springer Link server located at http://dx.doi.org/10.1007/s004399900217

analysis of affected sibpair UK and USA families questioned the previous findings (Concannon et al. 1998; Mein et al. 1998). It is interesting to note that IDDM pathologies are characterised by a destruction of the insulin-producing β cells where SEL1L is mostly expressed.

Several correlations between mutations in Notch-like genes and pathways to human diseases have been described. In fact, three human disorders including a neoplasia (a T cell acute lymphoblastic leukaemia/lymphoma), a late onset neurological disease (CADASIL), and a developmental disorder (the Alagille syndrome), are associated with mutations in, the *Notch1*, *Notch3* (Larsson et al. 1994; Lindsell et al. 1995) and *Jagged1* (Li et al. 1997) genes, respectively, highlighting the broad spectrum of Notch activity in humans (Joutel and Tournier-Lasserve 1998).

In view of the pancreas-specific pattern of expression and its location in a chromosomal region harbouring a putative diabetes type I mellitus disease locus (*IDDM 11*), we have undertaken the task of refining the map position of *SEL1L* with respect to microsatellite markers isolated in the region. Furthermore, we have isolated the entire full-length cDNA and identified its genomic structure. In addition, we have isolated two new polymorphic (CA)n microsatellites, within *SEL1L*, useful for new linkage analysis studies. This report adds new information to help determine if *SEL1L* is a candidate gene for IDDM 11.

Materials and methods

cDNA library screening

A cDNA library in λgt10 phage vector derived from human pancreatic tissue (Clontech, Palo Alto, Calif.) was screened using standard procedures with a SEL-1L 569-bp cDNA probe previously described (Biunno et al. 1997). Recombinant positive phages were isolated and purified following the procedure described in Sambrook et al. (1989) and Sambrook and Gething (1989). cDNA inserts were subcloned into the *Eco*RI site of pBluescript KS II(+) (Stratagene, La Jolla, Calif.) and sequenced by standard techniques.

cDNA sequence and analysis

The ORF sequence was compared to the non-redundant protein database using the BLAST algorithm (Altschul et al. 1994; Altschul et al. 1997) at NCBI at http:www.ncbi.nlm.nih.gov/cgi-bin/BLAST/. The genomic region upstream of the starting methionine was analysed by the promoter prediction neural network software at http:www-hgc.lbl.gov/projects/promoter.html. Multiple alignments were obtained by using ClustalW software at URL location http:www2.ebi.ac.uk/clustalw/. Regions of similarities found with the motif database were determined by Prosite software available at URL location http:expasy.hcuge.ch/sprot/scnpsit1.html. A PEST motif region was determined by PESTfind analysis available at URL location http:www.at.embnet.org/embnet/tools/bio/PESTfind/. The determination of fibronectin type II sequence pattern was obtained by using Prosite Scan Server and Prosite patterns DataBase at URL location http:www.isrec.isbsib.ch/software/PSTSCAN_form.html. A signal cleavage peptide was predicted by SignalP ver1.1 at URL location http:www.cbs.dtu.dk/services/SignalP. TM-PRED software was used to locate candidate transmembrane regions and their orientation (available at http:www.ch.embnet.org/software/TMPRED_form.html). The multiple repeat motif (LG–Y-G-GV) in SEL1L and its similarity to known proteins was determined by the MoST program (Tatusov et al. 1994).

Isolation and characterization of *SEL1L*-containing PAC clones

The PCY PAC2 N human genomic PAC library was screened using a PCR-based strategy for STS marker SEL1L3, a 189 bp amplimer in the 3' untranslated region of the gene (primer sequences and PCR conditions are given in Table 1). Three clones (PCY

Table 1 Oligo sequences, TM and PCR conditions of all *SEL1L*-related primers

Oligo name	Sequence	TM	STS	dbSTS_Id	GenBank accession no.	PCR T_{ann}[1]
CAR (r)	5'-TGGGCTTGGTTAGTACTTGG-3'	57.5	Intron 2 (CA)n repeat	67333	G44759	65
CAL (f)	5'-AAAATTACTGACCTACAAGAGGG-3'	55.5				
RepIN20f	5'-CGTATTGGATTACTGGTGGAAAG-3'	59.6	Intron 20 (CA)n repeat	67332	G44758	65
RepIN20r	5'-GGCAAGGAACTGGGAAAGTTAC-3'	61.1				
Sel1L5F	5'-CAAAGGCTTGTGAATCCATAGC-3'	60.8	SEL1L5'UTR	72140	G49445	60
Sel1L5R	5'-GCTCATCCGAGACTACAATTCC-3'	59.9				
SelPacT7F	5'-CAGCTGTCACTGAAATATAGC-3'	52.4	SEL1LT7	72141	G49446	60
SelPacT7R	5'-CCTCATGTTCATTCTTCAAACC-3'	58.3				
Sel1L3 A	5'-CTGCTTAGAATCTGATGC-3'	48.1	SEL1L3'UTR	72139	G49444	52
Sel1L3B	5'-GGATACAGTAGACATTAC-3'	36.9				
SelPacSPF	5'-CCTATATTCCCAGCTATTTGG-3'	55.3	SEL1LSP6	72142	G49447	60
SelPacSPR	5'-GACATACGGCTATGCTCTGG-3'	58.1				
IBD4	5'-ACGATAGTGAGTCAGGGCAGATTC-3'	63	RACE-5'-PCR			68
IBD5	5'-TCTGCTTCCTGCATCTGCCGTCTC-3'	71	RACE-5'-PCR			
dT-anchor	5'-GACCACGCGTATCGATGTCGACT(16)-3'	75.4	RACE-5'-PCR			68
PCR-anchor	5'-GACCACGCGTATCGATGTCGAC-3'	67.2	RACE-5'-PCR			

[1]Except where stated otherwise, all PCRs were performed in the buffer specified by the *Taq* supplier, with 200 μM dNTPs, 1 μM each primer and 1.5 mM $MgCl_2$. Cycle conditions were 30 or 33 cycles of 94°C 30 s, T_{ann} 60 s, 72°C 60 s, where T_{ann} is the annealing temperature indicated

Table 2 Exon-intron sequence boundaries of *SEL1L* and primer pairs used to amplify individual *SEL1L* exons. *Asterisks* refer to those primers able to amplify the exon 7–8 corresponding genomic region at the same time in the same PCR reaction

5' Primer	Exon	3' Splice acceptor		Exon size (bp)	PCR size (bp)	5' Splice donor		3' Primer
GGGAGTCCATGGTGATTGG	1				397	ATG.CCTCGG	gtcagt	ATCTCGCACTGGGCTGTC
CTAAACCCCCAACCCATTTT	2	....ttgcag	ATGAAG	38	326	....TCCAAG	gtatgt	TTAGAAAGGATATGAGGTCTGCA
AGTAAAAATAACGGTCAAATAGGG	3	....tctcag	ACTACT	232	396	....AACCAG	gtagtc	TAAAGAGGCTGAAGAAGACTGG
CAGGCATCATGAATCTTTGC	4	....ttccag	CTTTGA	168	371	....GTGAAA	gtaagt	TTCTAGCCATCCTTTTTCCACA
CTAATGGTTGATAATTGATA	5	....ttttag	CTGAAG	106	271	....AAGAGA	gtaggt	CACTGCCAGAGACGCAGAAG
AAGGAAATTATTTATGTTGCGTTT	6	....ttgaag	AGCATA	163	391	....CAGACT	gtaagt	CCAATTTTGTGGCTGACCAT
TGTCTGTCTGTGTCTGTCTTGG(*)	7	....tttcag	GCTCTT	54	350	....GCAAAG	gtaata	
	8	....tcacag	GCTCTT	60		....GTTTTG	gtaagt	ATAGCGGCCACAACTTTTTC(*)
TTTGTTCCACGAAGAATGACA	9	....ttacag	GGTTAC	83	268	....ATCATG	gtatct	AATGGCTCTTTTATGTGGACTCA
GCAGCTGGAATTGTGCTTC	10	....tttcag	TTGCTA	155	346	....GCACAG	gtacgt	ACATTTTCATTGCCCTGAGC
CAGAAGCCATTTTACGACCAA	11	....tttcag	GTTGGT	57	281	....CATCAG	gtaact	AAACACCCATCACAGATGAGC
TTGGCTGAAAGTCAAAAGCA	12	....tcatag	AGAGCA	69	292	....GGAAAG	gtactg	TGCTGTTCTGAGGCATTTTC
TCTGCCAAAACTCAAAGGAGA	13	....ttccag	ATGTAT	78	339	....GACATG	gtaagg	TAAACAAAGCCCTCCCGATA
AAAGCCCCCAGATGTCTAA	14	....ttgcag	GGCAAC	63	306	....CAAGTT	gtaagc	GAGCAGTCAGCTGGAGAGGT
ATCTCTTAACTTTTGAGTATGTCTC	15	....ctacag	AATTAT	88	260	....ACTATA	gtaagt	TGATGTGGAAATGGGTGCTA
GATTTGTACATCTTGTTTTTCATTG	16	....ttacag	ATGGCA	149	379	....GTGGAG	gtgagg	GTGTAGGGAGTCGATCAGCC
GCTGCCAGAAATTGAAGGAG	17	....acacag	TTGTTT	166	417	....ATCAGA	gtaagg	CTGGAACTGAACACACAGGC
CTTATGCTTCACCTATCCTTC	18	....ctgtag	GAGAAG	75	211	....CTGAAG	gtagtg	ACCCAGGACCTCCTAGTGG
AAGCATGTCAATGGGAGGAG	19	....tttaag	GCTATA	173	371	....AAACAG	gtgagt	CAGTTGGATTTCAGGCCACT
GGCATTTTTGTTTAGGTGAGGG	20	....ttatag	GATATT	129	354	....ACAAAC	gtaagt	CAAAAGTTGAGAAGCACAGGG
TCCCAGTTCCTTGCCTGAC	21	....attaag	ATTCGA.....TAA		334			CTGCTGGGAACACTTGCAT

PAC2 N 863o15, 996o23 and 1000f15) were identified. Clones 996o23 and 1000f15 were found to contain the entire *SEL1L* gene; the latter was used for sequencing.

Physical and metric mapping

The pre-existing HAPPY map of chromosome 14 has been described (Dear et al. 1998). PAC clones containing markers to be placed on the HAPPY map were identified by PCR screening of either the PCY PAC2 N or the RPCI1 human PAC libraries using standard protocols. Primer sequences for the genetic markers used here can be found at http:www-genome.wi.mit.edu/, http:www-genome.wi.mit.edu/, and at GenBank (http:www2.ncbi.nlm.nih.gov/cgi-bin/genbank); primer sequences and PCR conditions for SEL1L-related markers are given in Table 1.

The PAC clones found to contain the genetic or SEL1L-related markers were placed on the HAPPY map by screening toothpicked single colonies by PCR for the presence of HAPPY markers and for other genetic and SEL1L-related markers (primer sequences and PCR conditions for all HAPPY markers can be found at http:www.mrc-lmb.cam.ac.uk/happy/happy-home-page.html).

The derivation and mapping of new HAPPY markers from clones containing D14S128 and D14S287 followed the procedures described in Dear et al. (1998).

Determination of exon-intron sequence boundaries

DNA from the PCY PAC2 N clone 1000f15 was prepared according to the Qiagen "midi" protocol. Approximately 2 μg of DNA template was used for each cycle of sequencing using the "Big Dye" terminator reaction (Perkin Elmer Applied Biosystems). Primers were derived from the cDNA human sequence by the RNASPL program software developed by Soloviev and Salamov (1997) and available at URL location http:dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html. Oligonucleotides (22–24 bases; Tm 58 C–64°C, and C/G clamp at the 3' end) were designed from regions of the sequence which did not contain potential splicing sites, in the 5' to 3' direction with respect to the cDNA sequence. Cycle sequencing reaction conditions were adjusted as described at URL location http:www.tigem.it/TIGEM/SEQCORE/protocols.html. The second set of oligonucleotides used to determine the 5' splicing sequence were designed in the 3' to 5' cDNA direction, the choice of the primers being based on the previous sequencing results. Overall, the method had a 90% success rate. The nucleotide sequence of each corresponding exon and its surrounding intron genomic sequence is deposited in GenBank with the corresponding accession nos.: AF198631; AF198632; AF198633; AF198634; AF198635; AF198636; AF198637; AF198638; AF198639; AF198640; AF198641; AF198642; AF198643; AF198644; AF198645; AF198646; AF198647.

Identification of polymorphic markers

Novel microsatellites discovered in this study were amplified from DNA samples derived from lymphocytes from 50 randomly chosen donors (see Table 2 for PCR conditions). Alleles were scored on 6% polyacrylamide gels, and allele frequencies were used to estimate the heterozygosity of each locus.

Results

Sequence analysis

Full-length cDNA corresponding to *SEL1L* was obtained by screening the human cDNA library as described in Materials and methods. The coding region consists of 2385 bp (GenBank accession no. AF052059) sequenced by a walking primer approach.

Comparison between our cDNA sequence and that of Harada et al. (1999) reveals two silent nucleotide differences (we find C versus T at position 237, and G versus A at 348), and one non-silent difference (G versus A at position 649, changing valine to methionine). A further non-silent mismatch at cDNA position 2189 (C versus A; proline versus glutamine) was observed by comparing our sequence with EST R98036 (sequence ID no. yr30b02.s1).

In addition, upstream of the starting methionine, a genomic sequence of 1271 bp (GenBank accession no. AF157516) was derived as described in Materials and methods. This sequence contains the essential features of a promoter region: the absence of a TATA box, the presence of a CAP site at nucleotide position 1085 (position +1 relative to the start of transcription), and a CAAT box between nucleotides 961 and 968 (–115 to –125). Furthermore, several GC, AP2 and NKX25 boxes were detected throughout the minimal promoter region defined between nucleotide position –382 and position +1.

BLAST (Altschul et al. 1994; Altschul et al. 1997) and CLUSTAL (Higgins et al. 1992; Higgins and Sharp 1988; Higgins et al. 1996) analyses using the entire ORF revealed 40% and 92% amino acid identity with the corresponding *C. elegans sel-1* gene (GenBank accession no. CEU50829) and the murine homolog *m-SEL-1L* (GenBank accession no. AF063095), respectively (Fig. 1). These proteins share several motifs: a signal peptide sequence, a signal for cleavage in the endoplasmic reticulum, and a tandemly repeated motif (LG–Y-G-GV) periodically distributed through the protein (Fig. 1).

By using the "Motif Search Tool" (MoST) program (Tatusov et al. 1994) we have found that this motif is shared with human G2/mitotic-specific cyclin F protein (region 87–122, ID CG2F-Human) which is involved in the control of the cell cycle during S and G2 phases (Bai et al. 1994; Kraus et al. 1994), and with the yeast STK5 protein (regions between amino acids 258–293, 294–329, 330–365, 486–521, ID SKT5-Yeast), which has been suggested to play a role in protoplast regeneration and in the resistance to killer toxin of *K. lactis* (Kawamoto et al. 1993). The role and distribution of these motifs in SEL1L is currently unknown. In addition, we located three transmembrane helices (between amino acids 286–305; 711–729; 753–769) in the protein by using TMpred software (Materials and methods). SEL1L protein contains a fibronectin type II domain that is missing from the other two species.

Furthermore, a polypeptide sequence enriched in proline, glutamic acid, serine and threonine (PEST motif) (Rechsteiner and Rogers 1996; Rogers et al. 1986; Rogers and Rechsteiner 1986), which serves as a proteolytic cleavage site, was identified in the amino terminal portion of the human and most likely the mouse proteins, but is missing in the nematode protein. Finally, the human and mouse proteins share a proline-rich C-terminal tail.

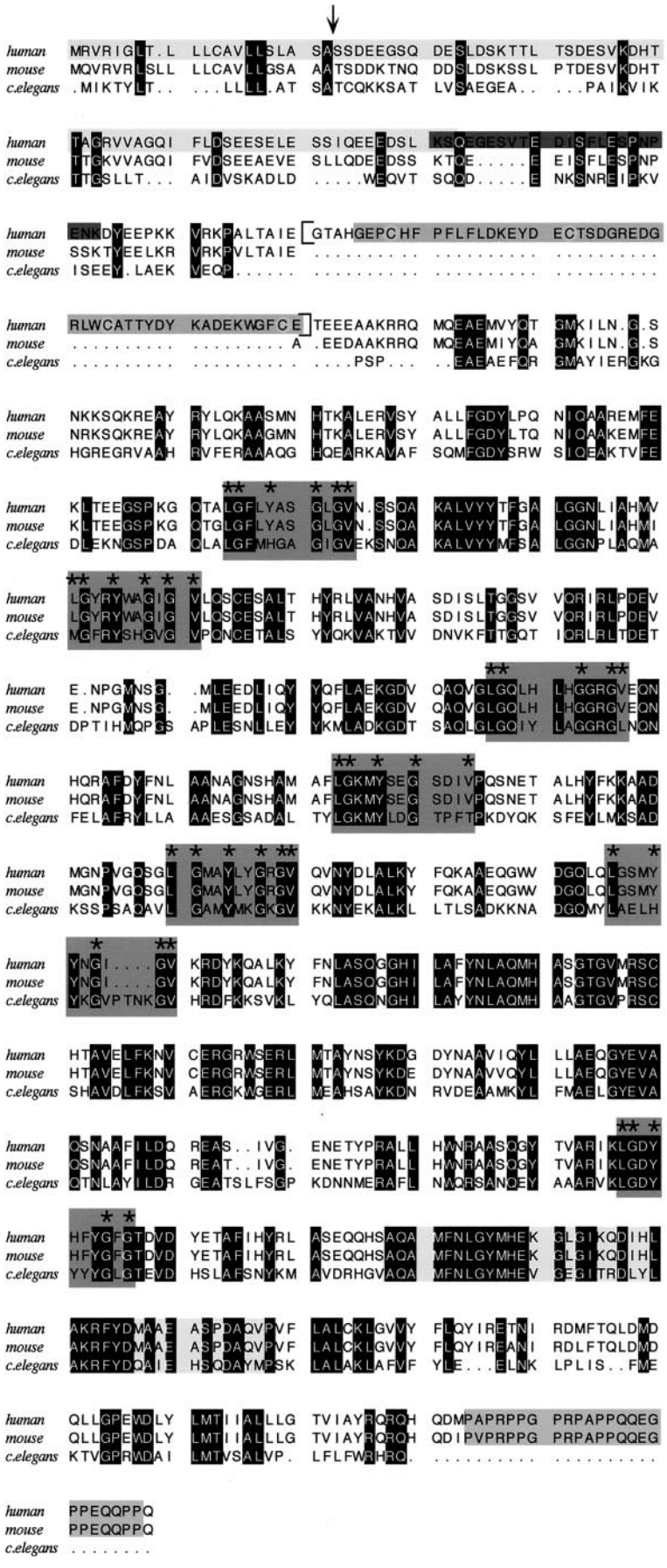

Fig. 1 Amino acid alignment between human, mouse and nematode SEL1 proteins. Regions of identity between all three proteins are shown as *white on black*. From the amino terminal to the carboxy terminal portion of the SEL1 proteins (in *dark grey*), a signal peptide, a PEST motif, a region similar to the fibronectin type II domain (*circles* indicate conserved cysteines), an Hrd3-like motif, and a proline-rich region are shown. *Square brackets* indicate exon 3 in the human. The *arrowhead* indicates the position of signal cleavage. *Grey blocks* indicate the positions of the tandem motifs repeated through the protein (LG–Y-G-GV). Supplementary material in color is available at http://dx.doi.org/10.1007/s004399900217

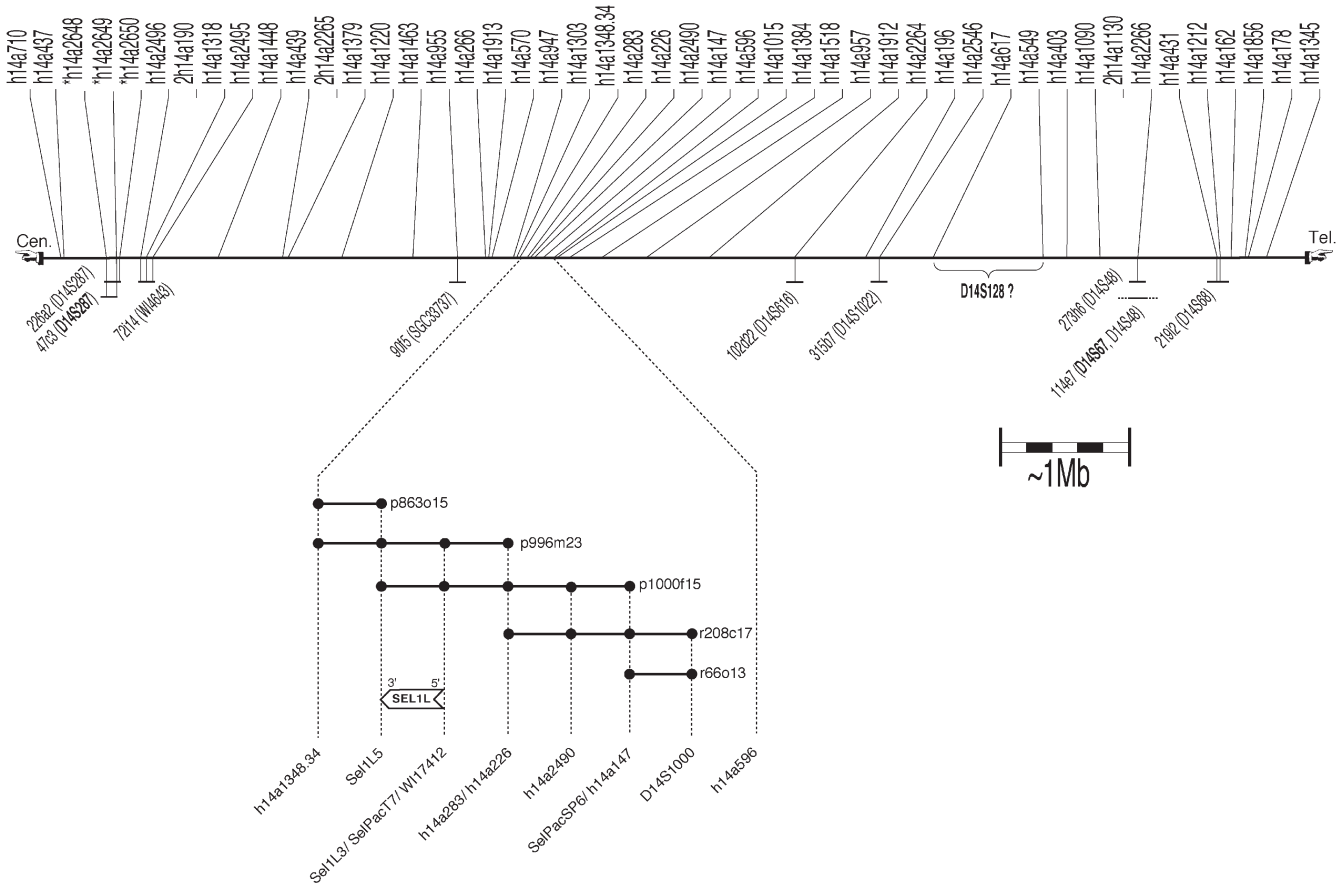

Fig. 2 Fine mapping of *SEL1L*. A part of the HAPPY map of human chromosome 14 is shown at *top*; HAPPY markers are given along the *top* (markers prefixed by asterisks are second-rate markers, assigned tentatively). PAC clones from the RPCI1 library which can be aligned with the map are shown immediately below (clone names are followed by the *bracketed names* of the genetic markers they contain; *vertical lines* connect the PACs with the HAPPY markers they were found to contain; PAC lengths are not to scale). The likeliest location of D14S128 on the map is indicated. *Below* is an expanded view showing a contig of PAC clones surrounding SEL1L. Clones are indicated as *heavy horizontal lines* (clone lengths and marker spacing not to scale), followed by the clone name. Clones prefixed with "p" are from the PCY PAC2 N library; those prefixed with "r" are from the RPCI1 library. Markers are named at the *bottom* (not all markers can be ordered on the basis of this contig). Markers Sel1L5 and Sel1L3 are from the 5' and 3' ends, respectively, of the *SEL1L* gene and delimit its location and orientation. SelPacT7 and SelPacSP6 are end sequences from the T7 and SP6 ends, respectively, of PCY PAC2 N clone 1000f15. RPCI1 clones 58g17, 90k15, 92k6, 116n11, 258i24 and 308a11 (not shown) all had identical marker content to PCY PAC2 N clone 863o15. The *scale bar* refers to distances on the HAPPY map. The distance from D14S67 to *SEL1L* is approximately 4.7 Mb. Shown in *bold* are microsatellite markers for which either an RH or linkage analysis study was previously reported

Physical and metric mapping

To position *SEL1L* more precisely with respect to the available genetic markers, we have placed markers from within the gene, as well as nearby microsatellite markers, on the metric HAPPY map of chromosome 14. This map, produced by a simple in vitro method analogous to radiation hybrid mapping, has been shown to have a resolution of better than 100 kb (Dear et al. 1998) (Fig. 2).

Most *SEL1L*-related and microsatellite markers were placed on the HAPPY map by establishing which of the HAPPY markers were present in PAC clones (identified by PCR screening; see Materials and methods) which also contained the *SEL1L*-related or microsatellite marker. Hence, these markers can be placed on the map to within one PAC-length (<160 kb) relative to the HAPPY markers. The PAC clone which contained D14S67 did not carry any of the HAPPY markers, but did contain D14S48; D14S67 can thus be placed within one PAC-length of D14S48, and hence within two PAC-lengths of the HAPPY marker h14a2266 (Fig. 2).

This approach was unsuccessful in the case of genetic markers D14S287 and D14S128. None of the PAC clones which contained these markers (RPCI1 clones 47c3 and 226a2 in the case of D14S287; RPCI1 clones 183i3, 225i17 and 279a16, and PCY PAC2 N clones 1143p13 and 1130m19 in the case of D14S128) carried any of the other genetic, *SEL1L*-related or HAPPY markers and, thus, could not be directly placed. Therefore, new HAPPY markers were derived from the D14S287-containing RPCI1 PACs 47c3 and 226a2, and from the D14S128-containing PACs RPCI1 183i3 and PCY PAC2 N 1143p13, following the protocols described in Dear et al. (1998). Three markers from the D14S287-containing PACs were mapped using the HAPPY protocol; their position on the map defines the location of these PACs and

hence of D14S287 (Fig. 2). However, neither of the markers from the D14S128-containing PACs mapped successfully, showing no significant linkage to other HAPPY markers. Weak linkages (lod<2) obtained with the long-range B- and C-panels (Dear et al. 1998) imply that these markers (and hence D14S128) lie in the region between h14a617 and h14a1212 (Fig. 2). The absence of other HAPPY markers in any of the D14S128-containing PACs suggest that they lie in one of two relatively large intervals on the HAPPY map: either h14a617–h14a549, or h14a2266–h14a431. Genetic mapping data indicate that the first of these intervals is the likeliest to contain D14S128. The locations of all other genetic markers on the HAPPY map agree with the marker order determined by genetic, physical and radiation hybrid maps of this chromosome. Five of the PAC clones surrounding *SEL1L* form a contig, defining the precise location and orientation of *SEL1L*, and the orientation of the PAC clone PCY PAC2 N 1000f15 (Fig. 2).

Genomic organization of *SEL1L*

By virtue of the location of the *SEL1L* gene in close proximity to the putative *IDDM 11* locus, as well as its pancreas-specific mRNA expression, we undertook to determine the exon-intron boundaries, as described in Materials and methods (see Table 2). The data generated can be used to perform mutation scanning of DNA from IDDM 11 patients. It is noteworthy that this approach identified several very short exons (such as exon 2, 38 bp in size) which would be difficult to identify by other strategies (Hagiwara and Harris 1996; Riley et al. 1990). The *SEL1L* gene is composed of 21 exons spanning 70 kb of genomic DNA (Table 2). Primer pairs able to amplify individual exons are presented in Table 2.

Polymorphic microsatellite repeat identification

While sequencing the genomic region of *SEL1L* we found two (CA)n repeats positioned in intron 2 (CAR/CAL) and intron 20 (repIN20f/repIN20r), about 40 kb apart (Table 1). We have tested CAR/CAL and repIN20f/repIN20r markers on 50 random DNAs extracted from donor lymphocytes and found them to be significantly polymorphic with a size range of 207–229 bp (14 to 25 repeats) and 239–259 bp (16 to 25 repeats), respectively. We observed 12 alleles of marker CAR/CAL and 10 of repIN20f/repIN20r in the population. We have determined the percentage of heterozygosity for allele CAR/CAL and repIN20f/repIN20r measured as 70.6% and 75.8%, respectively (refer to Table 3). This information will be of value to investigators interested in determining if *SEL1L* is a candidate locus for *IDDM 11*.

Table 3 Polymorphic markers identified, GenBank accession numbers, and heterozygosity allele information

Microsatellite	GenBank accession no.	Size range (bp)	Heterozygosity
CAR/CAL	G44759	207–229	70.6%
RepIN20	G44758	239–259	75.8%

Discussion

We have cloned a human cDNA, *SEL1L*, similar to the *C. elegans* protein sel1-l, a regulator enzyme of the Notch pathway. Recently, the murine homolog (*mSEL-1L*) was isolated by Donoviel et al. (1998) and shows an expression profile similar to the human gene. Similar results were obtained by us and presented at URL location http:www.tigem.it/~zollo/insitu.gif. Here we report the genomic organisation of the human gene, its precise mapping and a detailed sequence analysis. On the HAPPY map, SEL1L is located approximately 4.7 Mb centromeric to D14S67, a microsatellite marker located in the chromosomal restricted region of the putative type I insulin dependent diabetes mellitus (*IDDM 11*) locus (Field et al. 1996) and approximately 3 Mb telomeric to the D14S287 marker. These results are in contrast with the data reported by (Donoviel and Bernstein 1999) obtained by RH analysis, in which they show a distance of 1.2 Mb between *SEL1L* and D14S287. The distance reported here is in broad agreement with the sex-average genetic distance of 4.3 cM between D14S67 and D14S1000, the latter of which we find in very close proximity to *SEL1L*.

Based on linkage studies in a subset of type I diabetes families from the UK, Mein et al. (1998) reported a peak of linkage 33 cM proximal to the peak of linkage of the putative *IDDM 11* locus (Field et al. 1996). The data presented would indicate the presence of an alternative locus located at a centromeric position with respect to the previously identified locus.

However, the limitations of a number of markers isolated in the region, and of the sib pair approach used by Mein et al. (1998), means that the locus of Field et al. (1996) cannot be totally excluded.

It is interesting to note that IDDM pathologies are characterised by T lymphocyte-mediated auto-immune destruction of the insulin-producing ß cells in the pancreas as described in a recent overview by (Todd 1999). In addition, it is known that maturation of the thymus and of T cells is dependent upon down-regulation of the notch-like genes (Gray et al. 1999; Hasserjian et al. 1996), whose activity in turn depends on their ligands (Lendahl 1998) and upon the action of negative regulators such as sel1 and sel10 (Greenwald 1998) homologs. The remarkable expression of SEL1L in whole pancreatic cells, as well as its specific presence in the islands of Langerhans, suggest that SEL1L may be involved with the disease, albeit its apparent location centromeric to the putative chromosomal region.

Despite considerable efforts made by several groups, this region of the chromosome remains largely unexplored. Several ESTs of unknown function have been

placed in this chromosomal subregion and few polymorphic markers have been identified (GeneMap 1998; Deloukas et al. 1998). A genomic contig island was constructed and the genomic region encompassing the open reading frame was sequenced in order to determine the exon-intron sequence boundaries. We have found that the gene is composed of 21 relatively small exons spanning about 70 kb of DNA. This contrasts with the recent report of only 20 exons by Harada et al. (1999).

We have found three intragenic nucleotide differences (two silent and one non-silent) between our cDNA sequence and that reported by Harada et al. (1999). Further analysis is needed to determine whether these differences represent genuine poymorphisms. We have also identified two novel and significantly polymorphic (CA)n microsatellites, about 40 kb apart, in introns 2 and 20 (Table 3). These microsatellites provide a tool for further linkage analysis in families affected by IDDM.

Sequence analysis of the *SEL1L* cDNA revealed several interesting homologies and a novel family of motifs. We hypothesise that some of these motifs may play an important role in SEL1L functions related to cell cycle regulation. In addition, the finding of candidate transmembrane domains would further suggest a role of SEL1L on the cellular membrane surface; interestingly, one of these domains (amino acids 286–305) was not described by Harada et al. (1999).

SEL1L protein reveals the presence of PEST signals toward the N-terminus of the protein (Fig. 1).

A significant difference exists between the human protein and those of mouse and nematode. In the latter species the gene lacks the equivalent of human exon 3 (Fig. 1), which contains a region with strong amino acid homology to fibronectin type II. This portion of the protein is missing in the mouse sel1 released sequence, raising the possibility that the published sequence represents an alternatively spliced form. This region corresponds in human to a stretch of 49 amino acids containing four conserved cysteines and highly similar to the type II fibronectin collagen-binding domain. This domain is involved in disulfide bonds to part of the collagen-binding region of fibronectin, a plasma protein that binds cell surfaces. This motif has been found in other proteins such as type IV collagenase (Collier et al. 1988), in blood coagulation factor XII protein (Skorstengaard et al. 1986a; Skorstengaard et al. 1986b), in cation-independent mannose-6-phosphate receptor (Kornfeld 1992) (which also corresponds to the insulin-like growth factor II receptor), in mannose receptor of macrophages (Taylor et al. 1990), and in other transporter proteins. One observation implies that this motif plays a role in cellular adhesion and/ or cell-cell interaction; further experiments are needed to verify this hypothesis.

Observing the alignment shown between the human (AF052059) and mouse translated proteins in (Harada et al. 1999), we found discrepancies related to the presence of the type II fibronectin domain in the murine protein (mSEL-1L). The sequence released (AF063095) by Donoviel et al. (1998) does not contain the type II fibronectin domain. Further investigations are needed to address this point.

We are studying various aspects of SEL1L, both at the structural and functional level, in order to reach a hypothesis regarding the exact role of this protein in cell-cell interaction. At the same time, we are performing association studies in type I diabetic clinical material, such as case control or families, to link polymorphisms and haplotypes of polymorphisms to the *SEL1L* chomosomal region in order to verify if the gene is responsible for diabetes mellitus type I.

Acknowledgements We are grateful to Claudio Gattuso and Alessandro Bulfone for unpublished in situ mouse mRNA expression data, Francesco Cucca and Patrizia Zavattari for heterozygosity allele data, Giuseppe Borsani and Giorgio Casari for helpful and critical discussion, Igor Rogozin for help in computational analysis, and Melissa Smith for help in the preparation of this manuscript. This research was funded by Telethon grants to I.B and M.Z (TIGEM grant 1996–1999), by the Thalassemic Association of Ferrara, and by FIRC (Federazione Italiana Ricerca sul Cancro) sponsoring a fellowship to M.C.

References

Adam J, Myat A, Le Roux I, Eddison M, Henrique D, Ish-Horowicz D, Lewis J (1998) Cell fate choices and the expression of Notch, Delta and Serrate homologues in the chick inner ear: parallels with Drosophila sense organ development. Development 125:4645–4654

Altschul SF, Boguski MS, Gish W, Wootton JC (1994) Issues in searching molecular sequence databases. Nat Genet 6:119–129

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Artavanis-Tsakonas S, Matsuno K, Fortini ME (1995) Notch signaling. Science 268:225–32

Bai C, Richman R, Elledge SJ (1994) Human cyclin F. EMBO J 13:6087–6098

Biunno I, Appierto V, Cattaneo M, Leone BE, Balzano G, Socci C, Saccone S, Letizia A, Della Valle G, Sgaramella V (1997) Isolation of a pancreas-specific gene located on human chromosome 14q31: expression analysis in human pancreatic ductal carcinomas. Genomics 46:284–286

Collier IE, Smith J, Kronberger A, Bauer EA, Wilhelm SM, Eisen AZ, Goldberg GI (1988) The structure of the human skin fibroblast collagenase gene. J Biol Chem 263:10711–10713

Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS (1998) A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. Nat Genet 19:292–296

Crowe R, Henrique D, Ish-Horowicz D, Niswander L (1998) A new role for Notch and Delta in cell fate decisions: patterning the feather array. Development 125:767–775

Dear PH, Bankier AT, Piper MB (1998) A high-resolution metric HAPPY map of human chromosome 14. Genomics 48:232–241

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, Bentolila S, Bihoreau M, Birren BB, Browne J, Butler A, Castle AB, Chiannilkulchai N, Clee C, Day PJ, Dehejia A, Dibling T, Drouot N, Duprat S, Fizames C, Bentley DR, et al (1998) A physical map of 30,000 human genes. Science 282:744–746

Donoviel DB, Bernstein A (1999) SEL-1L maps to human chromosome 14, near the insulin-dependent diabetes mellitus locus 11. Genomics 56:232–233

Donoviel DB, Donoviel MS, Fan E, Hadjantonakis A, Bernstein A (1998) Cloning and characterization of Sel-1 l, a murine homolog of the C. elegans sel-1 gene. Mech Dev 78:203–207

Field LL, Tobias R, Thomson G, Plon S (1996) Susceptibility to insulin-dependent diabetes mellitus maps to a locus (IDDM 11) on human chromosome 14q24.3-q31. Genomics 33:1–8

Fleming RJ (1998) Structural conservation of Notch receptors and ligands. Semin Cell Dev Biol 9:599–607

Gene Map (1998) World Wide Web URL: http://www.ncbi.nlm.nih.gov/genemap98

Grant B, Greenwald I (1996) The Caenorhabditis elegans sel-1 gene, a negative regulator of lin-12 and glp-1, encodes a predicted extracellular protein. Genetics 143:237–247

Grant B, Greenwald I (1997) Structure, function, and expression of SEL-1, a negative regulator of LIN-12 and GLP-1 in C. elegans. Development 124:637–644

Gray GE, Mann RS, Mitsiadis E, Henrique D, Carcangiu ML, Banks A, Leiman J, Ward D, Ish-Horowitz D, Artavanis-Tsakonas S (1999) Human ligands of the Notch receptor. Am J Pathol 154:785–794

Greenwald I (1998) LIN-12/Notch signaling: lessons from worms and flies. Genes Dev 12:1751–1762

Hagiwara K, Harris CC (1996) 'Long distance sequencer' method; a novel strategy for large DNA sequencing projects. Nucleic Acids Res 24:2460–2461

Hampton RY, Gardner RG, Rine J (1996) Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3- methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. Mol Biol Cell 7:2029–2044

Harada Y, Ozaki K, Suzuki M, Fujiwara T, Takahashi E, Nakamura Y, Tanigami A (1999) Complete cDNA sequence and genomic organization of a human pancreas-specific gene homologous to Caenorhabditis elegans sel-1. J Hum Genet 44:330–336

Hasserjian RP, Aster JC, Davi F, Weinberg DS, Sklar J (1996) Modulated expression of notch1 during thymocyte development. Blood 88:970–976

Hatakeyama S, Kitagawa M, Nakayama K, Shirane M, Matsumoto M, Hattori K, Higashi H, Nakano H, Okumura K, Onoe K, Good RA, Nakayama K (1999) Ubiquitin-dependent degradation of IkappaBalpha is mediated by a ubiquitin ligase Skp1/Cul 1/F-box protein FWD1. Proc Natl Acad Sci USA 96:3859–3863

Heitzler P, Simpson P (1991) The choice of cell fate in the epidermis of Drosophila. Cell 64:1083–1092

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383–402

Hubbard EJ, Wu G, Kitajewski J, Greenwald I (1997) sel-10, a negative regulator of lin-12 activity in Caenorhabditis elegans, encodes a member of the CDC4 family of proteins. Genes Dev 11:3182–3193

Joutel A, Tournier-Lasserve E (1998) Notch signalling pathway and human diseases. Semin Cell Dev Biol 9:619–625

Kawamoto S, Sasaki T, Itahashi S, Hatsuyama Y, Ohno T (1993) A mutant allele skt5 affecting protoplast regeneration and killer toxin resistance has double mutations in its wild-type structural gene in Saccharomyces cerevisiae. Biosci Biotechnol Biochem 57:1391–1393

Kitagawa M, Hatakeyama S, Shirane M, Matsumoto M, Ishida N, Hattori K, Nakamichi I, Kikuchi A, Nakayama K, Nakayama K (1999) An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of beta-catenin. EMBO J 18:2401–2410

Kornfeld S (1992) Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. Annu Rev Biochem 61:307–330

Kraus B, Pohlschmidt M, Leung AL, Germino GG, Snarey A, Schneider MC, Reeders ST, Frischauf AM (1994) A novel cyclin gene (CCNF) in the region of the polycystic kidney disease gene (PKD1). Genomics 24:27–33

Larsson C, Lardelli M, White I, Lendahl U (1994) The human NOTCH1, 2, and 3 genes are located at chromosome positions 9q34, 1p13-p11, and 19p13.2-p13.1 in regions of neoplasia-associated translocation. Genomics 24:253–258

Lendahl U (1998) A growing family of Notch ligands. Bioessays 20:103–107

Lewis J (1998) Notch signalling and the control of cell fate choices in vertebrates. Semin Cell Dev Biol 9:583–589

Li L, Krantz ID, Deng Y, Genin A, Banta AB, Collins CC, Qi M, Trask BJ, Kuo W-L, Cochran J, Pierpont MEM, Rand EB, Piccoli DA, Hood L, Spinner NB (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat Genet 16:212–213

Lindsell CE, Shawber CJ, Boulter J, Weinmaster G (1995) Jagged: a mammalian ligand that activates Notch1. Cell 80:909–917

Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Pritchard LE, Cucca F, Barnett AH, Bain SC, Todd JA (1998) A search for type 1 diabetes susceptibility genes in families from the United Kingdom. Nat Genet 19:297–300

Rechsteiner M, Rogers SW (1996) PEST sequences and regulation by proteolysis. Trends Biochem Sci 21:267–271

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith JC, Markham AF (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18:2887–2890

Rogers SW, Rechsteiner MC (1986) Microinjection studies on selective protein degradation: relationships between stability, structure, and location. Biomed Biochim Acta 45:1611–1618

Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234:364–368

Sambrook J, Gething MJ (1989) Protein structure. Chaperones, paperones. Nature 342:224–225

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Skorstengaard K, Jensen MS, Petersen TE, Magnusson S (1986a) Purification and complete primary structures of the heparin-, cell-, and DNA-binding domains of bovine plasma fibronectin. Eur J Biochem 154:15–29

Skorstengaard K, Jensen MS, Sahl P, Petersen TE, Magnusson S (1986b) Complete primary structure of bovine plasma fibronectin. Eur J Biochem 161:441–453

Solovyev VV, Salamov AA (1997) The Gene-Finder computer tools for analysis of human and model organisms genome sequences. In: Rawling C, Clark D, Altman R, Hunter L, Lengauer T, Wodak S (eds) Proceedings of the fifth international conference on intelligent systems for molecular biology. AAAI Press, Halkidiki, Greece pp 294–302

Sundaram M, Greenwald I (1993) Suppressors of a lin-12 hypomorph define genes that interact with both lin-12 and glp-1 in Caenorhabditis elegans. Genetics 135:765–783

Tatusov RL, Altschul SF, Koonin EV (1994) Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. Proc Natl Acad Sci USA 91:12091–12095

Taylor ME, Conary JT, Lennartz MR, Stahl PD, Drickamer K (1990) Primary structure of the mannose receptor contains multiple motifs resembling carbohydrate-recognition domains. J Biol Chem 265:12156–12162

Todd JA (1999) From genome to aetiology in a multifactorial disease, type 1 diabetes. Bioessays 21:164–174