

Identification and characterization of the STIM (stromal interaction molecule) gene family: coding for a novel class of transmembrane proteins

Richard T. WILLIAMS^{*1}, Shehnaaz S. M. MANJI[†], Nigel J. PARKER[†], Manuela S. HANCOCK[†], Leonie VAN STEKELBURG^{*}, Jean-Pierre EID^{*}, Paul V. SENIOR^{*}, Janette S. KAZENWADEL^{*}, Tetyana SHANDALA[‡], Robert SAINT[‡], Peter J. SMITH[†] and Marie A. DZIADEK^{*}

^{*}Department of Anatomy and Cell Biology, University of Melbourne, Parkville, Victoria 3052, Australia, [†]Department of Paediatrics, University of Melbourne, Parkville, Victoria 3052, Australia, and [‡]Department of Genetics, University of Adelaide, North Adelaide, SA 5000, Australia

STIM1 (where STIM is stromal interaction molecule) is a candidate tumour suppressor gene that maps to human chromosome 11p15.5, a region implicated in a variety of cancers, particularly embryonal rhabdomyosarcoma. *STIM1* codes for a transmembrane phosphoprotein whose structure is unrelated to that of any other known proteins. The precise pathway by which *STIM1* regulates cell growth is not known. In the present study we screened gene databases for *STIM1*-related sequences, and have identified and characterized cDNA sequences representing a single gene in humans and other vertebrates, which we have called *STIM2*. We identified a single *STIM* homologue in *Drosophila melanogaster* (*D-Stim*) and *Caenorhabditis elegans*, but no homologues in yeast. *STIM1*, *STIM2* and *D-Stim* have a conserved genomic organization, indicating that the vertebrate family of two *STIM* genes most probably arose from a single ancestral gene. The three *STIM* proteins each contain a single SAM (sterile α -motif) domain and an unpaired EF hand within the highly conserved extracellular region, and have coiled-

coil domains that are conserved in structure and position within the cytoplasmic region. However, the *STIM* proteins diverge significantly within the C-terminal half of the cytoplasmic domain. Differential levels of phosphorylation appear to account for two molecular mass isoforms (105 and 115 kDa) of *STIM2*. We demonstrate by mutation analysis and protein sequencing that human *STIM2* initiates translation exclusively from a non-AUG start site *in vivo*. *STIM2* is expressed ubiquitously in cell lines, and co-precipitates with *STIM1* from cell lysates. This association into oligomers *in vivo* indicates a possible functional interaction between *STIM1* and *STIM2*. The structural similarities between *STIM1*, *STIM2* and *D-STIM* suggest conserved biological functions.

Key words: *Drosophila* homologue, extracellular EF hand, extracellular SAM domain, non-AUG translation start, transmembrane phosphoprotein.

INTRODUCTION

STIM1 (where STIM is stromal interaction molecule; also known as GOK) was identified as a novel human gene that maps to a region of chromosome 11p15.5 known to be involved in adult and childhood tumours [1]. *In vitro* studies have demonstrated a growth-inhibitory function of *STIM1* that is specific to rhabdoid tumour and rhabdomyosarcoma cell lines [2], implicating human *STIM1* as one of the potential tumour suppressor genes within the 11p15.5 locus. Biochemical characterization of *STIM1* has provided experimental evidence of N-linked glycosylation and cell surface localization [3], consistent with initial predictions that *STIM1* is a unique, cell surface, transmembrane glycoprotein [1]. We have demonstrated that *STIM1* is expressed ubiquitously in a wide variety of human primary and transformed cell types, including rhabdoid and rhabdomyosarcoma cell lines [3]. Previous studies have suggested that selective reduction of *STIM1* mRNA expression through methylation of the *STIM1* promoter may contribute to rhabdomyosarcoma tumorigenesis [4]. How-

ever, it remains unclear as to how the specific growth-inhibitory effects of *STIM1* relate to its described biochemical properties.

The mouse *STIM1* homologue, *Stim1*, was identified independently as a stromal-cell-derived gene product that binds to the surface of haematopoietic cells, particularly pre-B lymphoid cells, and can promote subsequent clonal expansion [5]. Murine *Stim1* is 96% identical with human *STIM1* in amino acid sequence, and maps to a region of mouse chromosome 7 syntenic with human 11p15.5 [6]. The α -helical coiled-coil domains within the cytoplasmic region of *STIM1* [1,5] and a SAM (sterile α -motif) domain in the extracellular region [7] may mediate the formation of higher-ordered structures. The phosphorylation of *STIM1* *in vivo*, predominantly on serine residues [3], suggests that *STIM1* may reside and function within a signalling cascade. To date, no other structural characteristics have been identified within these proteins that predict specific protein–protein interactions or biological function.

In this paper, we describe the identification and characterization of a new human *STIM1*-related gene, *STIM2*, repre-

Abbreviations used: CMV, cytomegalovirus; EST, expressed sequence tag; FISH, fluorescence *in situ* hybridization; ORF, open reading frame; RT-PCR, reverse transcription–PCR; S2 cells, Schneider 2 cells; SAM, sterile α -motif; STIM, stromal interaction molecule; anti-*STIM1*-NT and anti-*STIM1*-CT antibodies, antibodies raised to the N-terminal and C-terminal regions respectively of *STIM1*; STS, sequence-tagged site; UTR, untranslated region.

¹ To whom correspondence should be addressed (e-mail R.Williams@anatomy.unimelb.edu.au).

The nucleotide sequence data reported in this study have been deposited in the GenBank®/EMBL/GSDB/DBJ Nucleotide Sequence Databases with the following accession numbers: human *STIM2* cDNA (clone H5F7), AF328905; mouse *Stim2* cDNA, AF328907; *Drosophila melanogaster* *D-Stim* (clone LD06112), AF328906.

senting the second member of a two-gene family. Database searches provide evidence of *STIM1* and *STIM2* homologues in a number of vertebrate species, with no additional *STIM*-related gene products identified in any vertebrate. Additionally, we confirm the existence of a single *STIM* gene in *Drosophila melanogaster* and *Caenorhabditis elegans*. *STIM1*, *STIM2* and *D-STIM* share genomic structure, protein domain organization, and biochemical features. The *STIM* family has evolved from a single gene in lower multicellular eukaryotes, into two related genes in vertebrates. There is no evidence for the existence of *STIM*-related genes in yeast. Finally, we provide evidence that *STIM1* and *STIM2* are capable of interacting with each other *in vivo*, suggesting a reciprocal functional relationship between these two highly related molecules.

EXPERIMENTAL

Identification of *STIM*-like sequences

STIM1-like gene sequences were identified by tblastn [8] searching of NCBI GenBank databases [9] with the full amino acid sequence of human *STIM1* (GenBank accession no. U52426). After identification and sequencing of human *STIM2*, additional tblastn searches were conducted with the full-length *STIM2* amino acid sequence. Dbest [Database of ESTs (expressed sequence tags)], nr (Non-Redundant), gss (Genome Survey Sequence) and htgs (High Throughput Genome Sequences) databases were queried with these sequences. Analysis and annotation of the *STIM* proteins was carried out with SignalP [10], SMART [11], HMMTOP [12] and Coils [13] software packages. Two human EST clones (zc79d12 and zq84h12) identified as putative *STIM2* fragments were purchased from Genome Systems (St. Louis, MO, U.S.A.) and sequenced. RACE (rapid amplification of cDNA ends) was utilized to extend the 5' sequence. A *Drosophila STIM*-related EST was identified (LD06112; accession no. AA247009) and the cDNA clone was acquired and sequenced.

Isolation of full-length human *STIM2* cDNA clones

PCR was used to screen a human foetal brain Rapid-Screen library (Origene, Rockville, MD, U.S.A.) as described by the manufacturer. This library was prepared in the pCMV6-XL3 vector, in which the cDNA inserts are directionally cloned downstream of the CMV (cytomegalovirus) promoter. The primers were designed from the compiled *STIM2* EST sequences: forward, 5' CATGTCCTGAGTCCACCATG (nucleotides 689–709; all nucleotide numbering from clone H5F7); reverse, 5' GTTTATCTTCTCTGTGCAGATGG (nucleotides 860–882). Conditions for the PCR in a PE 9600 thermal cycler were 95 °C for 2 min, and then 35 cycles of 95 °C for 30 s, 62 °C for 30 s and 72 °C for 30 s. Products were analysed by agarose-gel electrophoresis, and the only two positive clones, H5F7 and F5A6, were obtained from Origene.

Nucleotide sequencing

PCR products and cDNA clones were sequenced using the ABI Big-Dye sequencing kit (PE Biosystems, Foster City, CA, U.S.A.) using the manufacturer's protocol for cycle sequencing. Sequencing was performed on either an ABI 310 or an ABI 377 sequencer, and the sequences were assembled and aligned using either Gene Navigator or Autoassembler (PE Biosystems). Endogenous transcripts of human *STIM2*, mouse *Stim2* and *D-Stim* were sequenced after reverse transcription-PCR (RT-PCR) of total RNA prepared from normal human kidney tissue, C57BL6/J mouse skeletal muscle and adult *w¹¹¹⁸* flies respectively.

Total RNA was extracted in TRIzol[®] Reagent (Life Technologies, Rockville, MD, U.S.A.) and purified using chloroform and isopropyl alcohol. Single-stranded cDNA was prepared by oligo(dT)-primed (mouse *Stim2* and *D-Stim*) or random-hexamer-primed (*STIM2*) reverse transcription utilizing M-MLV (Moloney murine leukaemia virus) reverse transcriptase (Life Technologies), and was utilized directly as the template for PCR amplification. Several overlapping PCR fragments (*STIM2* and *D-Stim*) and a single section of mouse *Stim2* were sequenced in both directions. This was sufficient to generate high-quality sequence for *D-Stim*, mouse *Stim2* and most of *STIM2*, while modifications were required to achieve high-quality sequence from the 5' region of *STIM2*. For this 5' region, sequence-specific oligonucleotides were used for cDNA synthesis, and PCR was performed for 50 cycles (instead of 35) with the addition of PCR enhancer (Life Technologies). Nested PCR was needed to generate sufficient product for sequencing.

Chromosomal mapping

Human *STIM2*

Two chromosome 4 sequence-tagged sites (STSs) showing identity to *STIM2* were identified (WI-4086 and SGC33834) by blastn [8] searching of the dbSTS section of GenBank [9] with the *STIM2* cDNA sequence. Their location and positioning relative to known loci was derived from the Whitehead Radiation hybrid map.

Mouse *Stim2*

A mouse EST (GenBank accession no. AA756198) with 86% identity with human *STIM2* (nt 3550–3927; numbered according to clone H5F7) was identified. Primers were designed to amplify a 363 bp fragment of this sequence (forward, 5' TGGGAAGAGTAAACTTGATCGA; reverse, 5' AGAACATTTAAAGATTCAAATA) from C57BL6/J or *Mus spretus* genomic DNA by PCR. The products were purified after 2% (w/v) agarose-gel electrophoresis and sequenced. Three polymorphisms were found between C57BL6/J and *Mus spretus*, one of which created a *DdeI* restriction site in *Mus spretus* at position 3853. This polymorphism was confirmed by restriction digestion of PCR products. Samples of 10 ng of each of the 94 DNA samples of the Jackson Labs BSS panel [14] were amplified by PCR. The products were digested overnight with *DdeI* and analysed by electrophoresis through 2% (w/v) agarose gels. The presence or absence of polymorphic bands (290 and 73 bp) was tabulated for each sample, and the data were sent to Jackson Labs for analysis.

D-Stim

The chromosomal localization of *D-Stim* was mapped by fluorescence *in situ* hybridization (FISH). A digoxigenin-labelled *D-Stim* RNA probe was prepared from full-length *D-Stim* cDNA and hybridized to polytene chromosomes according to the manufacturer's protocol (Nonradioactive In-situ Hybridization Application Manual; Boehringer Mannheim, Mannheim, Germany). Sheep anti-digoxigenin-rhodamine-lissamine (Boehringer Mannheim) amplified with anti-sheep-rhodamine-lissamine (Jackson ImmunoResearch Laboratories, West Grove, PA, U.S.A.) were used for detection of hybridization. Chromosomes were stained with Hoechst 33258.

Site-directed mutagenesis of *STIM2*

The GeneEditor mutagenesis kit (Promega, Madison, WI, U.S.A.) was used to introduce specific mutations into the H5F7

lysate was incubated with 100 µg of anti-STIM2-CT affinity-purified antibodies for 3 h at 4 °C. The immune complexes were captured with 80 µl of Protein G-Sepharose, and the beads were subsequently washed three times with lysis buffer containing a full complement of phosphatase and protease inhibitors [3], and then twice with Tris-buffered saline. The beads were then resuspended in 50 µl of 50 mM Tris/HCl (pH 8.0)/0.1 mM EDTA containing 100 µg of STIM2 immunizing peptide and 2 units of calf intestinal alkaline phosphatase (see below), and incubated at 37 °C for 1 h with occasional mixing. The supernatant containing the eluted, dephosphorylated STIM2 was resolved by SDS/PAGE [3] on 7.5% (w/v) polyacrylamide gels, and the proteins were electroblotted to BioTrace PVDF transfer membranes (Pall Gelman, Ann Arbor, MI, U.S.A.) using 100 mM Caps, pH 11.0, transfer buffer [3]. Coomassie Blue R-250 staining was performed to localize the STIM2 protein band on the membrane.

The STIM2 protein bands (two bands, each approx. 3 µg) were excised from the blot and subjected to conventional N-terminal (Edman) sequence analysis using an Applied Biosystems 494 Procise Protein Sequencing System.

Drosophila cell culture and transfection

Schneider 2 (S2) cells were grown in Schneider's *Drosophila* Medium (Life Technologies) supplemented with 10% (v/v) heat-inactivated fetal bovine serum. The full-length D-STIM cDNA clone LD06112 was subcloned into the pAc5.1 constitutive expression vector (Invitrogen, Groningen, The Netherlands). The stop codon within the D-STIM cDNA prevented the addition of the V5/6 × His epitope tag from this vector. CellFECTIN® Reagent (Life Technologies) was used for transient transfection of S2 cells with pAc5.1/*D-Stim*. S2 cells were seeded in a six-well plate at a density of 0.9×10^6 cells per well in 2 ml of Schneider's medium/10% (v/v) fetal bovine serum, and allowed to attach for 1 h at 25 °C. Then 2 µg of DNA and 9 µl of CellFECTIN reagent in 200 µl of OPTI-MEM® I reduced serum medium (Life Technologies) was incubated at room temperature for 15–45 min before being added to each well. Cells were incubated for 5 h at 25 °C, washed once with PBS, and re-plated into 2.5 ml of fresh Schneider's medium containing 10% (v/v) fetal bovine serum, 50 units/ml penicillin and 50 µg/ml streptomycin (final concentrations).

Metabolic labelling

For metabolic labelling studies, mammalian cells were transfected under standard conditions, with the exception that methionine- and cysteine-free Dulbecco's modified Eagle's medium (Life Technologies) with 1% (v/v) serum was utilized and that the cultures were supplemented with 200 µCi/ml L-[³⁵S]methionine/cysteine (Geneworks, Adelaide, SA, Australia) for the 18 h labelling period. N-linked glycosylation was inhibited in mammalian cells by the addition of 10 µg/ml tunicamycin (Sigma, St. Louis, MO, U.S.A.) during the culture period.

Phosphorylation studies were performed on STIM1- or STIM2-transfected 293T cells 24 h post-transfection. Cells were washed into phosphate-free Dulbecco's modified Eagle's medium (Life Technologies), and preincubated for 45 min at 37 °C, followed by the 2 h labelling period at 37 °C with 100 µCi/ml [³²P]P_i (Geneworks). The phosphatase inhibitor calyculin A (Biomol, Plymouth Meeting, PA, U.S.A.) was added to a final concentration of 100 nM where indicated for the final 1 h of labelling.

Protein modification

Enzymic deglycosylation of STIM2 and D-STIM in cell lysates was performed by incubating 25–50 µg of cell lysate with 2 units of endoglycosidase H (Boehringer Mannheim) in a 20 µl reaction for 1 h at 37 °C. The reaction buffer contained 50 mM sodium citrate, pH 5.5, and 0.5% SDS.

In vitro phosphatase treatment was performed on immunoprecipitated STIM2 protein from *in vivo* [³²P]P_i-labelled 293T cells. The washed STIM2 immunocomplexes were incubated with 2 units of calf intestinal alkaline phosphatase (Roche, Mannheim, Germany) in 50 mM Tris buffer (pH 8.0)/0.1 mM EDTA for 1 h at 37 °C (phosphatase treated), or in buffer alone at 37 °C (mock treated). Boiling the samples in SDS/PAGE sample buffer terminated the reactions.

RESULTS

Identification of human STIM2

Tblastn database searches of the dbEST division of the GenBank database revealed the presence of several overlapping human ESTs that predicted protein coding regions highly similar to, but not identical with, that of human STIM1. In order to identify putative full-length cDNAs representing *STIM1*-related gene(s), a human fetal brain Rapid Screen library was screened by PCR. Two independent clones, H5F7 (3.976 kb) and F5A6 (approx. 5.0 kb), were isolated and sequenced. These provided the full nucleotide sequence of the novel *STIM1*-related gene, that we have designated *STIM2*. While these two clones have an identical open reading frame (ORF), they differ in the lengths and complexity of their 5' and 3' untranslated regions (UTRs). Further database searching identified additional STIM2 ESTs, but no other STIM1-related or STIM2-related sequences. Thus, on the available information, *STIM2* appears to be the second gene to be identified in a two-gene family. The *STIM2* cDNA sequence is identical with that of a cDNA identified recently in independent studies (KIAA 1482; [15]), which represents a third human *STIM2* cDNA (GenBank accession no AB040915; 4.841 kb). Nucleotide sequence alignment of the three cDNAs revealed that H5F7 extends most in a 5' direction, by 237 bases relative to F5A6 and by 315 bases compared with KIAA 1482. The *STIM2* cDNA sequence was used to design oligonucleotide primers for RT-PCR amplification of *STIM2* mRNA from normal human kidney. Direct sequencing of overlapping PCR products demonstrated the presence of endogenous *STIM2* transcripts with a nucleotide sequence identical with that of nt 1–3027 of clone H5F7 (results not shown).

Translation start site of human STIM2

While the ORF of *STIM2* showed 53% sequence identity with human *STIM1* over 577 amino acids, an equivalent AUG translation start site was not present at the corresponding position in the aligned cDNA sequence. Modelling with the Signal P Version 2.0 signal peptide prediction server [10] indicates that the amino acids encoded by nt ~ 501–572 (in the *STIM2* reading frame) have a very high probability of residing within a signal peptide, with a high likelihood of cleavage immediately C-terminal to this sequence. The position of this predicted signal peptide aligns very closely with the signal peptide of *STIM1* when these two sequences are compared [a full amino acid sequence alignment of *STIM1*, *STIM2* (KIAA 1482) and D-STIM (derived from AE003500) is available at www.kazusa.or.jp/huge]. While no AUG start sites are present

near this region in human STIM2 cDNA, several potential non-AUG translation start sites lie within either acceptable or even strong initiation contexts [16]. ESTs corresponding to this region of mouse and rat *Stim2* were identified, and the translated sequences were aligned with that of human *STIM2* (Figure 1A). The SignalP signal peptide prediction server predicts similar signal peptides and cleavage sites for all three proteins. In both rat and mouse *Stim2* cDNAs, AUG codons align with a non-AUG codon in human *STIM2* (UUG; nt 531–533), and all lie within adequate Kozak sequence contexts.

The translation start site of human *STIM2* was determined experimentally using clone H5F7 cDNA into which three specific mutations were introduced (Figure 1B). Translation of the *STIM2* cDNAs was assayed by Western blotting after transient transfection of human G401 and 293T cells. Non-mutated H5F7 cDNA produced a predominant 105 kDa *STIM2* protein in G401 cells that is identical in size with endogenously synthesized *STIM2*, and two major proteins of 105 and 115 kDa in 293T cells (Figure 1B). An endogenous *STIM2* protein was not readily observed in 293T cells in these experiments. Further studies indicated that the larger 115 kDa isoform represents a highly phosphorylated form of *STIM2* (see below). Introduction of a stop codon at bases 522–524 (mutation #1), three codons immediately upstream of the putative UUG start codon, resulted in the translation of a *STIM2* protein species identical in size with that produced from the non-mutated cDNA in both cell lines (Figure 1B). Modification of the sequence around this tentative start region (mutations #2 and #3) resulted in the production of a *STIM2* protein species having an electrophoretic mobility intermediate between those of the 105 and 115 kDa isoforms (Figure 1B), indicating usage of a different translation start site. The size of the product in mutant #2 and #3 is consistent with translation from a downstream, normally silent, start site that results in the production of a non-signal-peptide-containing *STIM2* protein, although the possibility of translation initiation from an upstream site cannot be eliminated. Taken together, these data demonstrate that *STIM2* protein produced in cells after transfection of *STIM2* cDNA initiates translation from a single unconventional start site [either UUG (nt 531–533) or CUG (nt 534–536)] in an adequate Kozak context. The similarity in the sizes of endogenously produced *STIM2* and the recombinant forms indicates the strong likelihood that this non-AUG start site is used exclusively *in vivo*. N-terminal sequencing of immunoaffinity-purified *STIM2* protein from transfected 293T cells identified the first ten N-terminal residues as XELVPRHLRG (single-letter code). This is in precise agreement with the sequence of the predicted mature protein (see Figure 1A), and strongly supports the localization of the translation start site as determined by the mutagenesis studies. These studies indicate a signal peptide of 14 residues in *STIM2*, in contrast with the 22 amino acids predicted for *STIM1*.

STIM gene family members in other vertebrates and invertebrates

STIM family members were identified by Tblastn searches of GenBank databases. Murine *Stim1* has been identified previously (accession no. U47323 [5]). ESTs were identified for rat (*Rattus norvegicus*; accession nos AA996745 and AI763957), bovine (*Bos taurus*; AV609285 and AW669469) and swine (*Sus scrofa*; AW787215 and BE663170) *Stim1* genes. ESTs representing *STIM2* homologues were identified in mouse (*Mus musculus*; accession nos AA088943, AI194208, AW106055, AW910374, BE652414, BE951006 and BF463756), rat (AA944338, BF286659 and BF286660) and bovine (BE482998) databases. The mouse

Stim2 cDNA sequence was compiled from overlapping ESTs (listed above) and an additional fragment (nt 774–1210; numbered according to H5F7) amplified by RT-PCR from adult skeletal muscle RNA. These sequences covered the full coding region, except for a short stretch between nt 718 and 774. The predicted amino acid sequence indicates that mouse *Stim2* is approx. 93% identical with human *STIM2* over 96% of the coding region (results not shown).

EST sequences representing *STIM2* homologues were identified in amphibian (*Xenopus laevis*; accession nos AW633493, AW639117 and BF427891) and avian (*Gallus gallus*; AI981296) species, but no *STIM1* homologues were identified in the relatively limited EST sequences available. Two genomic sequence clusters were identified in the pufferfish, *Tetraodon nigroviridis* [17], that code independently for *STIM1*- and *STIM2*-like proteins. Several overlapping fragments were identified that corresponded closely to exons 2–6 of human *STIM1* (accession nos AL266411, AL268327, AL209250, AL207878 and AL197593). Similarly, a single fragment corresponding to exon 2 (AL235887; exon numbering based on human *STIM1*) and overlapping fragments (AL177867, AL250006, AL244891 and AL235692) corresponding to exons 3–7 of human *STIM2* were found.

A single *Drosophila melanogaster* *STIM* family member was identified on the basis of significant identity between the human *STIM1* amino acid sequence and the translated amino acid sequence of several overlapping *Drosophila* ESTs. A 2.1 kb cDNA clone (clone LD06112) was sequenced and shown to contain a 1710 bp ORF with significant identity with both human *STIM1* and *STIM2*. The predicted 570-amino-acid protein, designated D-*STIM*, is considerably smaller than either human *STIM1* or *STIM2*. A best-fit alignment between D-*STIM* and the human *STIMs* indicates an additional approx. 70–80-amino-acid stretch at the N-terminus and a truncated C-terminus. This cDNA sequence was found to be virtually identical with the predicted stromal cell protein homologue transcript CT26146 identified in the Celera *Drosophila* genome-sequencing project [18] (derived from predicted gene CG9126; accession no. AE003500; Figure 2A). This 3223 bp CT26146 transcript predicts an identical coding region over the first 556 amino acids, with an alternative 3' coding region producing a further 510 amino acids (CG9126 gene product CP26146; accession no. AAF48542) rather than the very short 14-amino-acid tail identified in clone LD06112 (depicted schematically in Figure 2A). The 3' region of CT26146 predicts an amino acid sequence that has no detectable similarity with either human *STIM1* or *STIM2*, and also clusters independently with the 5' ends of several separate *Drosophila* ESTs, suggesting that CT26146 represents a hybrid transcript.

To determine the actual transcript size and sequence in *Drosophila*, we sequenced RT-PCR products amplified from wild-type *w¹¹¹⁸* adult fly RNA. The sequences were identical with that of clone LD06112 over the entire predicted coding region (including the stop codon), except for a single amino acid substitution at position 38 (Arg to Cys) of *w¹¹¹⁸* flies compared with the Celera CT26146 and LD06112 sequences (results not shown). Therefore clone LD06112 represents the true endogenous D-*STIM* transcript. Clone LD06112 produced a 66 kDa protein by coupled *in vitro* transcription and translation, and was translated into a protein of approx. 68 kDa 24 h after transient transfection into S2 cells, that was similar in size to the major endogenous immunoreactive protein (Figure 2B). An additional higher-molecular-mass D-*STIM* isoform was evident 48 h after transfection. Both isoforms were specifically immunoprecipitated and immunoblotted with the Pan-*STIM* antibody (Figure 2B). The observed molecular sizes of these proteins are consistent

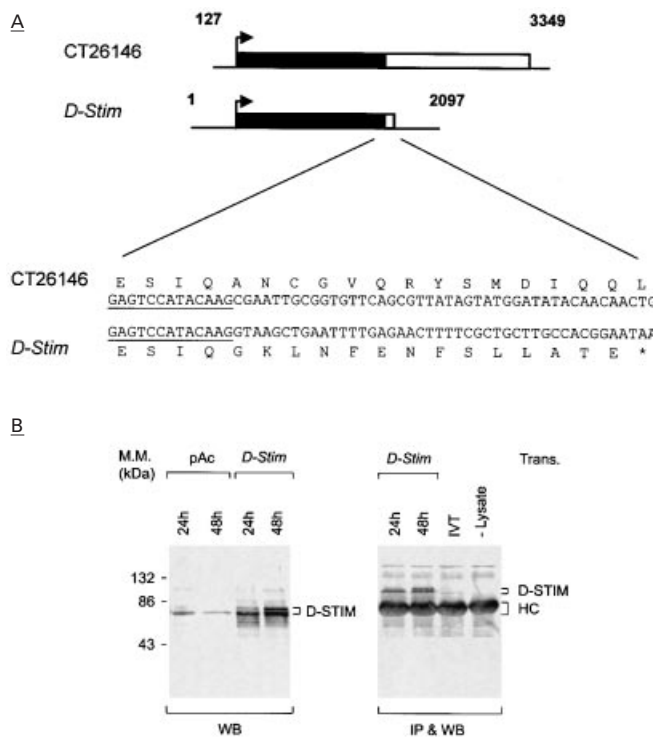


Figure 2 Identification of *D-Stim* and its gene product, D-STIM

(A) Schematic alignment of the predicted Celera stromal cell protein homologue transcript (CT26146) with the experimentally determined full cDNA sequence of *D. melanogaster Stim*. The boxed areas represent coding regions of the cDNAs, while filled and open regions indicate sequence identity and divergence respectively. The cDNA sequence and the translated amino acid sequences in the region of sequence divergence are also indicated. Nucleotides are numbered according to *D-Stim* (clone LD06112). (B) Identification of D-STIM. *D. melanogaster* S2 cells were transiently transfected (Trans.) either with empty pAc5.1 expression vector (pAc) or with *D-Stim* in the pAc5.1 expression vector (*D-Stim*), and cellular lysates prepared either 24 or 48 h post-transfection were immunoblotted with the Pan-STIM antibody (WB). In parallel, the indicated *D-Stim*-transfected S2 cell lysates, *in vitro* translated D-STIM (IVT) or lysis buffer control (—Lysate) were immunoprecipitated with the Pan-STIM antibody, and the immunocomplexes were immunoblotted with the Pan-STIM antibody (IP & WB). The positions of D-STIM protein and the heavy chain (HC) of the sheep immunoglobulins are indicated.

with post-translational modification of a 570-amino-acid sequence predicted for D-STIM. We found no evidence of an endogenous immunoreactive protein that would correspond to the 1066-amino-acid CP26146 protein. Taken together, these data indicate that the Celera CT26146 sequence for *D-Stim* does not represent the true endogenous sequence of this transcript. No additional sequences could be identified in the completed sequence of the euchromatic, gene-rich, portion of the *Drosophila* genome [18] that had identity with mammalian *STIM1* or *STIM2*, indicating the existence of only a single *STIM* homologue in the *Drosophila melanogaster* genome.

From tblastn searching of the *Caenorhabditis elegans* genomic and EST databases, and the subsequent annotation of genes and derived gene products [19], it is clear that *C-Stim* (gene Y55B1BM.b; accession no. AC024823) represents the single *bona fide* *STIM* homologue in the nematode. The *C-Stim* gene product (accession number AAF59595) is predicted from the available genomic sequence and two *C. elegans* EST sequences (accession nos AV197217 and AV184336) that correspond to the translation start site, signal peptide plus EF hand, and cytoplasmic portion of the protein respectively. A sequence alignment

of the highly conserved EF hand regions (see below) of the available vertebrate and invertebrate *STIM* homologues is presented in Figure 3(A) to illustrate the high degree of sequence conservation within this gene family, particularly across the 400 million year evolutionary gap between pufferfish and humans [17]. No additional *STIM*-like genes can be identified in databases representing yeast or prokaryotic genomes.

Domain structural comparison of human *STIM2* and *STIM1* and *Drosophila* D-STIM

The *STIM2* protein predicted from the nucleotide sequence contains 746 amino acids, which is processed to a mature 732-residue protein through signal peptide cleavage, as verified experimentally by N-terminal sequencing. *STIM2* thus has an additional 61 amino acids when compared with *STIM1*. A BLAST comparison of human *STIM1* and *STIM2* reveals 53% amino acid identity and 66% similarity over 577 amino acids (approx. 85% of the length of *STIM1*), with significant sequence divergence located in the extreme C-terminal regions only. The D-STIM protein of 570 amino acids, including a predicted signal peptide of 23 residues, is equally similar to both *STIM1* (33% identical; 50% conserved) and *STIM2* (31% identical; 46% conserved). Overall, the three *STIM* family members are predicted to represent type I transmembrane proteins, with a single transmembrane segment separating an N-terminal extracellular region from a C-terminal cytoplasmic region. The predicted *STIM2* and D-STIM proteins share several structural features with human *STIM1* [1] (Figure 3B). All three *STIMs* contain a pair of cysteine residues spaced eight amino acids apart at similar positions near the N-terminus, in addition to a previously unnoticed single helix-loop-helix region which contains several acidic residues and conforms to the consensus for an EF-hand calcium-binding motif [20]. A SAM domain [21], which forms a five-helical bundle structure and has been noted previously in *STIM1* [7], is situated in the exoplasmic region of all three proteins, close to the predicted membrane-spanning domain. An N-linked glycosylation site delineates the N-terminal limit of the SAM domains in all three proteins, while *STIM1* possesses a unique potential N-linked glycosylation site within the SAM domain (Figure 3B). The single-pass transmembrane region is highly conserved in all *STIM* proteins and contains a single cysteine residue. Unique to D-STIM is the presence of a 70–80-amino-acid region between the signal peptide and the cysteine pair, which possesses no obvious structural features or homologues with other proteins (Figure 3B).

As in *STIM1*, the cytoplasmic regions of *STIM2* and D-STIM contain a significant degree of α -helical structure, a large proportion of which is predicted to form coiled coils [13], which displays weak identity with known structural proteins such as myosin. Further towards the C-terminus, beyond the tail of D-STIM, *STIM2* contains a proline- and histidine-rich motif (PHAPHPSHRPHHPHQHTPHSLPSPDP) at a similar position to a serine- and proline-rich region (SPSAPPGGSPHLDSSRSHSPSPDPDTPSP) in *STIM1*. The two human *STIM* proteins diverge significantly distal to this region, with the exception of similar, yet distinct, lysine-rich tails of 14 residues (five lysines in *STIM1*) and 17 amino acids (nine lysines in *STIM2*). Importantly, none of the three *STIM* proteins contains an identifiable catalytic domain.

Comparative expression of human *STIM1* and *STIM2*

A Northern blot of RNA from various human tissues probed with *STIM2* cDNA demonstrated a single 4.0 kb mRNA species in all tissue samples, with some modest variation in abundance

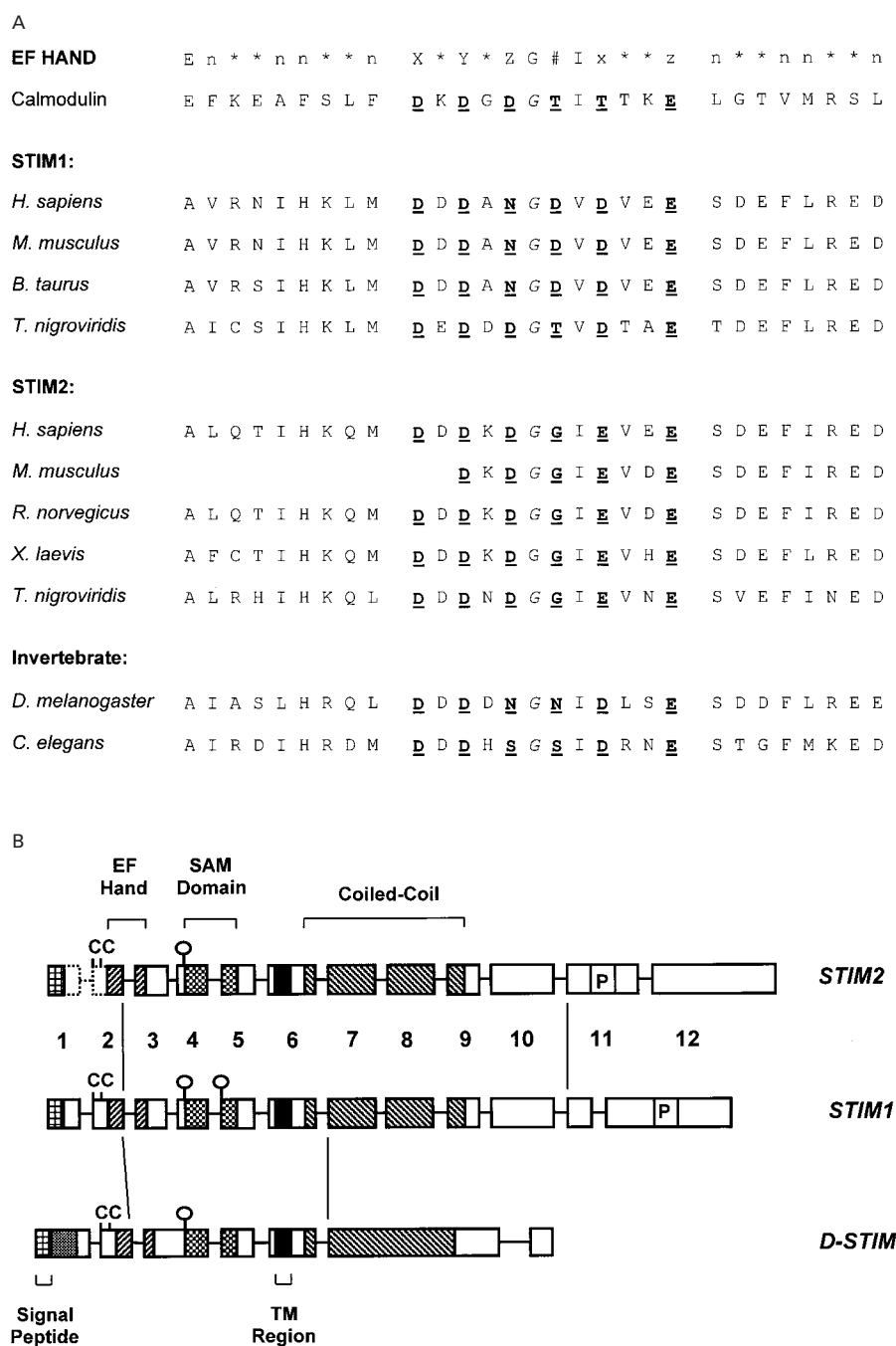


Figure 3 Sequence alignment, domain structure and genomic organization of STIM molecules

(A) Alignment of the available sequences from the EF-hand region of vertebrate STIM1 and STIM2 homologues with those of the two identified invertebrate STIM family members. Sequences are from *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Tetraodon nigroviridis*, *Rattus norvegicus*, *Xenopus laevis*, *Drosophila melanogaster* and *Caenorhabditis elegans*. The minimal consensus of the EF hand is indicated above the alignment [20], with loop residues implicated in ion binding indicated by X, Y, Z, #, x and z; conserved hydrophobic residues in the flanking helices are indicated by n. Under the EF-hand consensus is the sequence of the first EF hand of human calmodulin (accession no. P02593) for comparison. (B) Schematic representation of the domain structure of human STIM2 (top), human STIM1 (centre) and *D. melanogaster* D-STIM (bottom). The transmembrane (TM) region separates an N-terminal extracellular region from a C-terminal cytoplasmic region. Indicated are the signal peptides, the conserved pair of closely spaced cysteine residues (C), and predicted EF-hand and SAM domains. The putative N-linked glycosylation sites are also indicated (O). In the cytoplasmic region are α -helical regions predicted to form coiled-coil structures, and the proline-rich domains (P) unique to the mammalian STIM family members. D-STIM contains amino acid sequence in the extracellular region that is not present in either STIM1 or STIM2 (stippled box). Coding exons, numbered according to human *STIM1*, are represented by boxes drawn to scale, while introns are not to scale. The lines joining exons 2 and 11 indicate that the exon structures of *STIM1* and *STIM2* are absolutely conserved within this region, while the conservation of genomic structure between *D-Stim* and human *STIMs* is restricted to the extracellular and transmembrane regions. The DNA sequence and exon structure of the 5' end of human *STIM2* (putative exons 1 and 2) is not currently available.

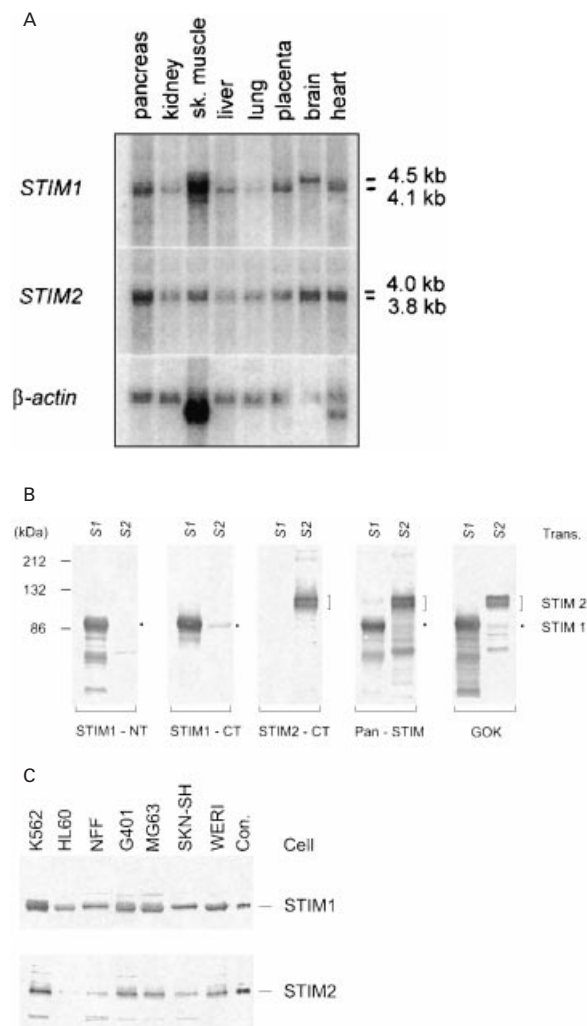


Figure 4 Expression analysis of STIM1 and STIM2

(A) Northern expression analysis of STIM1 and STIM2. A human multiple-tissue Northern blot was probed with cDNA probes specific for human *STIM1* and *STIM2*, and the β -actin control probe. The sizes of the major transcripts are indicated. (B) Characterization of antibodies against human STIM1 and STIM2. Human 293T cells were transfected (Trans.) with either human STIM1 (S1) or STIM2 (S2) cDNAs in eukaryotic expression vectors. Immunoblots were performed on these cellular lysates with STIM1-specific antibodies (STIM1-NT and STIM1-CT), STIM2-specific antibodies (STIM2-CT), antibodies generated against multiple STIM family members (Pan-STIM), or a commercial monoclonal antibody (GOK) raised against the N-terminal region of human STIM1. This N-terminal region is highly conserved at the primary amino acid level in human STIM2. (C) Expression of endogenous STIM1 and STIM2 in human cells. Lysates prepared from a variety of human tumour cell lines (K562 and HL60 leukaemia, G401 rhabdoid, MG63 osteosarcoma, SKN-SH neuroblastoma and WERI retinoblastoma cell lines) and non-transformed neonatal foreskin fibroblasts (NFF) were immunoblotted with antibodies specific for STIM1 (anti-STIM1-CT) and STIM2 (anti-STIM2-CT). Portions of approx. 100 ng of immunoaffinity-purified STIM1 and STIM2 proteins were included as controls (Con.) on the respective blots.

(Figure 4A). This mRNA species corresponds most closely in size to cDNA clone H5F7. The absence of detectable transcripts of 4.8 kb suggests that the larger, 3'-extended STIM2 cDNA clones represent less abundant mRNA species, consistent with the relative paucity of ESTs in this 3' region. Notably, semi-quantitative RT-PCR expression analysis of KIAA 1482 using primers specific to the 3' ends of KIAA 1482 and F5A6 (beyond the 3' end of H5F7) clearly indicates readily detectable expression in a number of human tissues, particularly the adult central

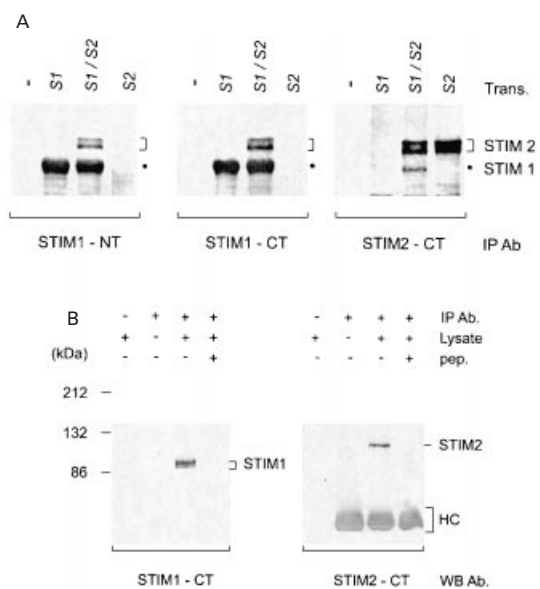


Figure 5 Association of human STIM1 with human STIM2 *in vivo*

(A) Association of STIM1 and STIM2 in transfected 293T cells. Human 293T cells were transfected (Trans.) with empty vector (—), STIM1 (S1), STIM2 (S2) or both STIM1 and STIM2 (S1/S2) expression constructs, and lysates were prepared. Immunoprecipitations (IP) were then performed with the indicated STIM1- and STIM2-specific antibodies, and the resulting immune complexes were immunoblotted with the Pan-STIM antibody to detect both human STIM1 and STIM2. (B) Association of endogenous STIM1 and STIM2 in K562 cells. Lysates from human K562 cells were immunoprecipitated (IP) with no antibody (—) or with STIM1-specific antibodies (+; anti-STIM1-NT) in the absence (—) or presence (+) of the immunizing peptide (pep.). An additional control included antibody, but no lysate (—). The resulting immune complexes were immunoblotted with anti-STIM1-CT and anti-STIM2-CT antibodies. The heavy chain (HC) from the anti-STIM1-NT sheep immunoglobulins is indicated.

nervous system [15]. In comparison, STIM1 mRNA expression varies significantly in abundance in the same tissues, with clear expression of 4.1 kb (most tissues) and 4.5 kb (predominantly brain) transcripts.

An affinity-purified polyclonal antibody was prepared against a peptide modelled on the extreme C-terminus of human STIM2. In Western blotting, the STIM2-specific antibody detected a 105/115 kDa doublet in cells transfected with *STIM2* cDNA (Figure 4B), and showed reactivity against an endogenous STIM2 of 105 kDa (Figure 4C). These proteins did not react with antibodies raised to either the N-terminus (STIM1-NT) or C-terminus (STIM1-CT) of STIM1 [3]. In 293T cells transfected with *STIM1* cDNA, both anti-STIM1 antibodies, but not the anti-STIM2 antibody, detected the 90 kDa STIM1 (Figure 4B). These data demonstrate the specificity of the antibodies against STIM1 and STIM2. Parallel immunoblots were used to characterize two additional STIM family antibodies. A Pan-STIM antibody was prepared by immunizing animals with a mixture of peptides modelled on the highly conserved helix-loop-loop region (putative EF hand) in the N-terminal region of STIM1, STIM2, D-STIM and C-STIM (*C. elegans* STIM). While this antibody was not as sensitive in its ability to detect endogenous STIM expression, it reacted with both STIM1 and STIM2 proteins produced by transfected cells, confirming the reading frame assigned to the STIM2 cDNA close to the N-terminus of the protein (Figure 4B). This antibody also detected both endogenous and transiently transfected D-STIM (Figure 2B). Additionally, we utilized a commercial anti-GOK (STIM1) monoclonal antibody that was generated to an extreme N-

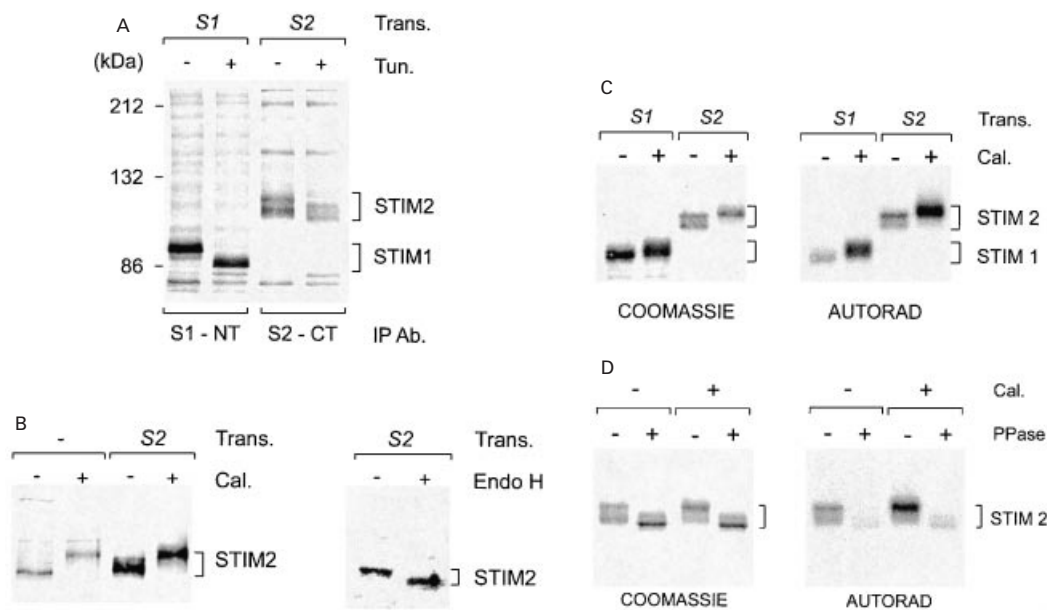


Figure 6 Post-translational modifications of human STIM1 and STIM2

(A) Human STIM1 and STIM2 are both modified by N-linked glycosylation. Human 293T cells were transfected (Trans.) with either STIM1 (*S1*) or STIM2 (*S2*) expression constructs, and the cells were subsequently metabolically labelled overnight with [³⁵S]cysteine and [³⁵S]methionine in either the absence (–) or the presence (+) of tunicamycin (Tun.). The STIM proteins were immunoprecipitated (IP) with their respective antibodies (S1-NT for the STIM1-transfected cells and S2-CT for the STIM2-transfected cells). The immune complexes were resolved by SDS/PAGE, and labelled proteins were visualized by autoradiography. (B) STIM2 is modified by phosphorylation and glycosylation in the human G401 cell line. G401 cells were either mock transfected (–) or transfected with a human STIM2 expression construct (*S2*), and 24 h later the cells were left untreated (–) or were treated with the phosphatase inhibitor calyculin A (Cal.; +) for 1 h. Cell lysates were immunoblotted with the STIM2-specific antibody. Note that the antibody detects endogenous STIM2 with an identical molecular size and response to calyculin A as the transfected STIM2. In a second study, lysates from STIM2-transfected (*S2*) G401 cells were incubated without (–) or with (+) endoglycosidase H (Endo H). (C) *In vivo* phosphorylation of human STIM1 and STIM2. Human 293T cells were transfected with either STIM1 (*S1*) or STIM2 (*S2*) expression vectors, and the following day the cells were incubated with [³²P]_i for a 2 h labelling period. Cells were either non-treated (–) or treated with calyculin A (+) for the second hour of labelling. STIM1 and STIM2 were recovered by immunoprecipitation with their respective antibodies, and the immune complexes were resolved by SDS/PAGE. The gels were stained with Coomassie Blue (COOMASSIE), dried and subjected to autoradiography (AUTORAD). (D) *In vivo* phosphorylation of human STIM2 mostly accounts for the observed heterogeneity of STIM2 molecular size. Human 293T cells were transfected with a human STIM2 expression construct, and labelled with [³²P]_i in the absence (–) or presence (+) of calyculin A, as in (C). The STIM2 protein recovered in the immune complexes was subjected to *in vitro* dephosphorylation (with alkaline phosphatase (PPase), or was mock phosphatase treated (–). The STIM2 proteins were separated by SDS/PAGE, stained with Coomassie Blue and subjected to autoradiography.

terminal region of human STIM1 that is highly conserved at the primary sequence level between STIM1 and STIM2. This monoclonal antibody detected both STIM1 and STIM2 in transfected cells, confirming the structural similarity between these two STIM molecules (Figure 4B).

Utilizing the affinity-purified anti-STIM2-CT antibodies prepared and characterized as described above, STIM2 expression was surveyed in a variety of human cells (Figure 4C). As noted previously for STIM1 [3], STIM2 is expressed at variable levels in all tumour cell lines examined, as well as in non-transformed primary fibroblast culture. By including known amounts of STIM1 and STIM2 proteins as controls on these immunoblots, it appears that in these cell types STIM2 is expressed mostly at lower-steady state levels than STIM1. These data indicate that, at least in established cell lines, STIM1 and STIM2 are co-expressed in the same cell.

Interaction between STIM1 and STIM2

Co-transfection and immunoprecipitation assays were used to determine whether specific interactions occur between STIM1 and STIM2. Human 293T cells were transiently transfected with STIM1 and STIM2 cDNAs, either alone or together. After 2 days of culture, proteins were immunoprecipitated from cell lysates with anti-STIM1-NT, anti-STIM1-CT or anti-STIM2 antibodies, and immunoblotted with the Pan-STIM antibody (Figure 5A). The three antibodies were shown to specifically

immunoprecipitate their respective ligands with moderate efficiency when cells were transfected with STIM1 or STIM2 cDNAs individually (Figure 5A). After co-transfection, readily detectable levels of STIM2 were co-precipitated specifically with STIM1, and vice versa. Thus the STIM1-specific antibodies co-precipitate STIM2 when it is co-expressed with STIM1, and STIM1 is co-precipitated with STIM2 when co-expressed with it. These data indicate that there is a specific interaction between STIM1 and STIM2 when these proteins are co-expressed.

To determine whether an interaction occurs between endogenously expressed STIM1 and STIM2, immunoprecipitation analyses were performed on lysates prepared from K562 cells, in which STIM1 and STIM2 are abundantly expressed (Figure 5B). Endogenous STIM1 can be immunoprecipitated specifically from these cells with the anti-STIM1-NT antibody, which can be blocked by preincubation of the antibody with immunizing peptide [3]. Immunoblotting of these same immunoprecipitates with the STIM2-specific antibody demonstrated co-precipitation of endogenous STIM2 with STIM1, and an absence of STIM2 immunoreactivity in the peptide-blocked control (Figure 5B). These data indicate that STIM1 and STIM2 interact to form co-precipitable oligomeric associations *in vivo*.

Post-translational modifications of STIM2 and D-STIM

The N-linked glycosylation and phosphorylation of STIM2 were examined to determine whether the electrophoretic isoforms

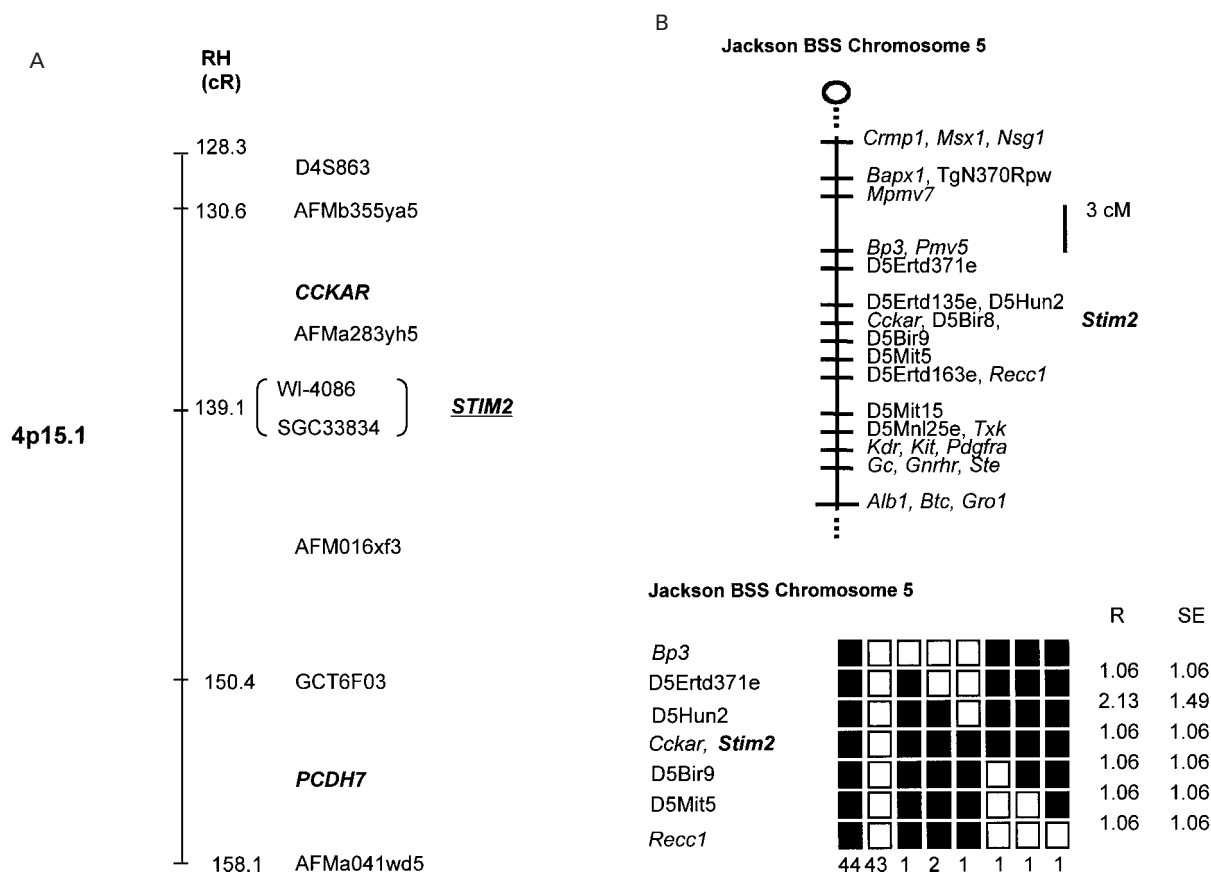


Figure 7 Chromosomal mapping of *STIM2* and *Stim2*

(A) Human *STIM2* chromosomal mapping. *STIM2* was mapped to human chromosome 4p15.1, between *CCKAR* and *PCDH7*, through identification of two mapped STSs, WI-4086 and SGC33834, with identical sequence to *STIM2*. This region of the Whitehead RH map is shown schematically. Numbering indicates the distance (in cR) from the telomere of 4p. (B) Mouse *Stim2* chromosomal mapping. *Stim2* was mapped to a syntenic region of mouse chromosome 5 through analysis of the Jackson BSS backcross. The upper panel indicates the map of this region of mouse chromosome 5, with the centromere towards the top, and a 3 cM scale bar indicated. The lower panel indicates the haplotype analysis of loci on chromosome 5 that are linked to *Stim2*. Loci are listed in order, with the most proximal at the top. Black boxes represent the C57BL/6J/Ei allele, and white boxes represent the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column. The percentage recombination (R) between adjacent loci is given to the right, along with the standard error (SE) for each R.

of *STIM2* are due to post-translational modification. *STIM1* and *STIM2* cDNA constructs were transfected independently into 293T cells that were then labelled overnight with a mixture of [³⁵S]methionine and [³⁵S]cysteine in the presence or absence of tunicamycin. Newly synthesized *STIM1* and *STIM2* proteins were recovered by immunoprecipitation (Figure 6A). In the absence of tunicamycin, *STIM1* exists exclusively as a single species of 90 kDa, which is reduced to a 84 kDa form when N-linked glycosylation is inhibited by tunicamycin, similar to the size of *in vitro*-translated *STIM1* and of *STIM1* digested with endoglycosidase H [3]. In contrast, while inhibition of the N-linked sugar modification of *STIM2* with tunicamycin increased its gel mobility, this treatment failed to shift all isoforms to a single lower-molecular-mass form. Similar results were obtained on immunoblotting cell lysates of *STIM2*-transfected 293T cells cultured in the presence or absence of tunicamycin (results not shown). These results indicate that *STIM2* is modified by N-linked glycosylation, but that this modification does not account for the size difference between the two major isoforms (105 and 115 kDa) of transfected *STIM2* in 293T cells. Enzymic deglycosylation of transfected *STIM2* in G401 cells also confirmed that the dominant 105 kDa form is modified by N-linked glycosylation (Figure 6B).

The phosphorylation of *STIM2* was analysed in G401 cells and 293T cells transiently transfected with *STIM2* cDNA. Cells were cultured in the presence or absence of the phosphatase inhibitor calyculin A, and changes in electrophoretic mobility were assessed by immunoblotting (Figure 6B). A decrease in the electrophoretic mobility of the low level of endogenously produced *STIM2* was seen in G401 cells treated with calyculin A. The apparent molecular size, 115 kDa, of this *STIM2* protein was similar to that of the upper isoform observed in transiently transfected, non-calyculin-treated 293T cells (Figures 6B and 6C). In both cell lines, calyculin A treatment resulted in decreased mobility of virtually all *STIM2* produced from the transfected construct to a single larger isoform of approx. 115 kDa. On immunoblots that were purposely overdeveloped, only a very small degree of smearing was observed above these larger isoforms, indicating that even larger forms are not generated after the inhibition of phosphatase activity. A difference in the proportions of the two dominant isoforms was detected in the *STIM2*-transfected G401 and 293T cells (compare Figures 6B and 6C), with much lower abundance of the upper 115 kDa form in the G401 cells. Taken together, these data suggest that the larger isoform (115 kDa) of *STIM2* is a more highly phosphorylated form of the smaller isoform (105 kDa).

To provide direct evidence for the phosphorylation of STIM2, 293T cells independently transfected with STIM1 and STIM2 expression constructs were labelled with [³²P]P_i, in either the presence or the absence of calyculin A. Phosphorylated STIM proteins were detected by autoradiography after immunoprecipitation (Figure 6C). The level of incorporation of [³²P]P_i in the absence of calyculin A appeared to be similar for STIM1 and STIM2 when compared with the amounts of each protein recovered (Coomassie Blue-stained gel). A 2.5–3-fold increase in incorporation into both proteins was seen after calyculin A treatment, while in the absence of calyculin A the upper isoform of STIM2 incorporated significantly more label than the lower isoform (Figure 6C). Phosphatase treatment of [³²P]P_i-labelled STIM2 resulted in increased mobility of virtually all of the larger isoform to that of the smaller 105 kDa form (Figure 6D). These results demonstrate that STIM2 is modified by phosphorylation *in vivo*, and that the observed heterogeneity in the molecular size of STIM2 in 293T cells arises through variable degrees of phosphorylation.

Enzymic deglycosylation of D-STIM was performed to determine whether the two main D-STIM isoforms observed at 48 h post-transfection represent differentially glycosylated variants. Endoglycosidase H treatment increased the mobility of both D-STIM species (results not shown), confirming that, while D-STIM, like STIM1 [3] and STIM2, is modified by N-linked glycosylation, this modification does not account for the electrophoretic heterogeneity of D-STIM.

Comparative genomic organization of *STIM1*, *STIM2* and *D-Stim*

Alignment of the STIM2 cDNA sequence against the dbSTS division of GenBank revealed two STSs, both of which map to chromosome 4p15.1 (Figure 7A). The *STIM2* gene is located between the genes encoding the cholecystokinin type A receptor (*CCKAR*; telomeric) and BH-protocadherin (*PCDH7*; centromeric). Mouse *Stim2* was mapped to a syntenic region of mouse chromosome 5 by analysing the Jackson Laboratory interspecies backcross panel BSS (Figure 7B). *D-Stim* was mapped to band 14A of the X chromosome by FISH (results not shown). Southern blot hybridization of the D-STIM cDNA to a cosmid spanning this chromosome region [22] positioned *D-Stim* 25–50 kb proximal to *shibire* (*Shi*) (results not shown). This cytogenetic localization has been confirmed in the recent sequencing and annotation of the *Drosophila melanogaster* genome [18].

The intron–exon boundaries of *STIM1* and *STIM2* were determined by alignment of cDNA (blastn) and amino acid (tblastn) sequences with genomic sequences in the human genome database. The genomic organization of the two genes is highly conserved (see Figure 3B). Approx. 85% of the coding region and the entire 3' UTR of human *STIM2* is encoded by 10 exons spaced over 71 kb within a single 170 kb contig (NT-002811; accession no. AC006928) on human chromosome 4. The GC-rich 5' end of *STIM2* and the promoter region are yet to be completed by the Human Genome Sequencing Project. The *STIM1* transcript is encoded by 12 exons spaced over approx. 250 kb [4]. Most importantly, in comparing *STIM1* and *STIM2*, the intron–exon boundaries are exquisitely conserved, with only the position of the most 3' intron differing, suggesting that these two genes have evolved from a common ancestor. In the analysis presented in Figure 3(B), we have compared the genomic structures in terms of translated amino acid sequences, and have overlaid predicted protein domains to highlight similarities. Comparison of the D-STIM cDNA clone with the *Celera Drosophila* genomic database [18] reveals that the *D-Stim* locus is significantly more compact, encompassing seven coding exons

spaced over only 4.2 kb (Figure 3B). The exon structure of *D-Stim* is conserved compared with those of *STIM1* and *STIM2*, particularly with respect to exons coding for the extracellular region (note that a single exon in *D-Stim* corresponds to exons E3 and E4 in mammalian *STIMs*). In contrast, the cytoplasmic region is mostly encoded by a single large exon in *D-Stim* (E6), and by six exons (E7–E12) in *STIM1* and *STIM2*.

DISCUSSION

STIM1 was initially identified as a novel transmembrane protein that had no structural similarity with any other known proteins [1]. *STIM1*-like EST and genomic sequences in available databases enabled us to identify *STIM2* as the only other human gene that is related to *STIM1*. In the present study we have characterized two independent cDNA clones of *STIM2*, and present a structural comparison with *STIM1*. Nagase et al. [15] also identified a *STIM1*-related cDNA as part of their large-scale gene cloning strategy to identify cDNAs that encode large proteins (The Kazusa Project; [23]). This group correctly identified KIAA 1482 (Kazusa gene nomenclature) as a *GOK/STIM1* homologue as part of their extensive HUGE database annotation (www.kazusa.or.jp/huge), but this relationship was not noted in their accompanying publication [15]. This *STIM1*-related sequence is identical with what we have now classified as *STIM2*. Further BLAST analysis with the complete *STIM1* and *STIM2* cDNA sequences has so far shown that all human *STIM1*-related ESTs and those from other vertebrates represent either *STIM1* or *STIM2* transcripts. Thus, on the available evidence, it appears clear that in vertebrates the *STIM* family is a two-gene family.

In contrast, a single *Stim* gene is present in invertebrates, represented by *D-Stim* in *D. melanogaster* and *C-Stim* in *C. elegans*, both of which are approximately equally similar to *STIM1* and *STIM2*. Indeed, in the case of *D-Stim*, for which a full cDNA sequence is available, similarity to the mammalian *STIMs* extends to the conservation of protein domain organization and genomic structure over a significant part of the gene. Taken together, these observations suggest that the *STIM* family has evolved in metazoans from a single ancestral gene, to a dual-gene family in vertebrates. No *STIM*-like genes have been identified in prokaryotes or unicellular eukaryotes. The existence of *STIM* genes only in metazoans is consistent with the cell–cell communication [5] and growth control functions [2] ascribed to *STIM1*.

Human *STIM2*, unlike *STIM1* and rodent *Stim2* counterparts, does not initiate translation from a typical AUG start site that is recognized by the translational machinery in the vast majority of eukaryotic mRNAs [16]. We have demonstrated experimentally that human *STIM2* translation is initiated from either a UUG or an adjacent CUG that lie within weak Kozak consensus sequences. The UUG codon (nt 531–533) corresponds to the position of the presumed AUG start site in both mouse and rat *Stim2*, and on this basis is the most likely start site *in vivo*. Start of translation at this non-AUG site predicts the presence of a signal peptide in *STIM2* that is conserved in *STIM1*. N-terminal sequence analysis of processed *STIM2* corresponds precisely to the sequence of *STIM2* produced after cleavage of the predicted signal peptide. There are clear examples in the literature of both the use of non-AUG start sites and translation start sites in apparently very poor Kozak consensus contexts [16]. However, in almost all cases of non-AUG starts, translation is initiated at a downstream in-frame AUG start in addition to the upstream non-AUG, through a process known as leaky scanning [16,24]. We have no evidence that human *STIM2* is translated from a

normally silent downstream (or upstream) site, except when the predicted UUG/CUG translation start codons are mutated. In this case translation appears to be initiated from an alternative non-AUG site. Human STIM2 thus represents a rare example of a eukaryotic protein with unique translation from a single non-AUG start codon.

While the three STIM2 cDNAs that have been characterized would be expected to encode identical STIM2 polypeptides (we have established this experimentally for clones H5F7 and F5A6; results not shown), they differ in the extent of their 5' and 3' UTRs. Northern blot analysis demonstrated that, in adult tissues, the dominant STIM2 transcript is 4.0 kb in size, which corresponds most closely to cDNA clone H5F7. Most of the 100 STIM2 ESTs now available cluster within the 3' end, with a significant number (16) having 3' ends corresponding to the 3' end of clone H5F7 (nucleotide 3976). Almost the entire region of H5F7 is now covered with overlapping ESTs. However, there is no EST coverage of the first 237 bases of clone H5F7 in any species examined, although we have been able to demonstrate the existence of such transcripts by RT-PCR. While the absence of such ESTs may be due to technical difficulties arising from the significant length of these cDNAs and their high GC content (average of nearly 75% GC over the first 600 bases), the available data suggest that the dominant STIM2 transcript commences around base number 238, corresponding to the 5' ends of clone F5A6 and the three mouse Stim2 ESTs. This would produce transcripts with a 5' terminal oligopyrimidine stretch of 17 T nucleotides in succession, perhaps similar to the 5' TOP (terminal oligopyrimidine) motif in ribosomal and co-regulated mRNAs that are translationally up-regulated in response to mitogens, and/or targeted for binding and regulation by a growing number of *trans*-acting factors [25].

STIM1 and STIM2 are clearly closely related to each other with respect to their primary amino acid sequence and their predicted secondary structure and domain organization. Both proteins are predicted to be single-pass transmembrane proteins, with an exoplasmic N-terminal region and a cytoplasmic C-terminal region. Both STIM1 [3] and STIM2 (the present study) are modified by N-linked glycosylation (placing at least part of these molecules within the endoplasmic reticulum lumen during their lifespan), consistent with the detection of the N-terminal region of STIM1 on the cell surface [3] and the finding that the C-terminal region of STIM1 is the prime site of *in vivo* phosphorylation (R. T. Williams, unpublished work). These findings are also consistent with the identification of murine Stim1 as a stromal cell molecule that is capable of binding to the surface of B lymphoid cells, and the observation that the minimal fragment required for this interaction lies within the predicted extracellular region [5]. Indeed, this minimal fragment of murine Stim1, containing the predicted EF-hand motif, binds pre-B cells in a bivalent-cation-dependent manner [5], suggesting that cation binding and associated conformational changes in this portion of the molecule (a very well characterized phenomenon in EF hands [26]) may be required for such interactions. The recent discovery of EF-hand motifs in the extracellular region of cell surface proteins of the α , β -hydrolase fold family supports this model [27].

Both STIM1 [7] and STIM2 contain predicted SAM domains, which were first identified in yeast and *Drosophila* as single-copy modules [21], and subsequently found in a wide variety of eukaryotic signalling, scaffolding and adaptor molecules, and transcriptional regulators [7,28]. More recently, it has been appreciated that these 70-amino-acid domains mediate a variety of relatively low-affinity homo- and hetero-typic protein-protein interactions in signalling molecules and transcriptional regulators

(reviewed in [28]). To our knowledge, the SAM domains in the STIM family members represent the first putative extracellular SAM domains identified. A logical prediction would be that these regions either modulate extracellular STIM-STIM interactions and/or mediate interactions with as yet unidentified soluble ligands or counter-receptors on the substratum or the surface of other cells.

The N-terminal half of the cytoplasmic region of both STIM molecules is predicted to be almost exclusively α -helical, most of which is predicted to form coiled-coils [13]. We anticipate that homotypic and heterotypic interactions between STIM proteins are mediated by these cytoplasmic coiled-coil regions. The sequences of the three STIM proteins diverge significantly C-terminal to the coiled-coils; STIM1 and STIM2, but not D-STIM, contain unique proline-rich regions that include serine/threonine residues. We have shown that STIM1 is phosphorylated predominantly on serine residues [3], mainly within the proline-rich region (R. T. Williams, unpublished work). As a central paradigm in cell biology, phosphorylation is likely to be a key regulator and modifier of STIM1 and STIM2 function.

The largely conserved domain structure between D-STIM and the mammalian STIM molecules suggests similar molecular functions. Specifically, we would predict that the conserved exoplasmic regions might mediate similar extracellular interactions, while the presence of coiled-coils in all family members would be likely to produce higher-order structures. However, the divergent C-terminal regions of the STIM family members and the additional N-terminal region of the D-STIM molecule provide the opportunity for specific functioning of these proteins. While one must be extremely cautious in drawing functional similarities on the basis of shared sequence and domain organization, it is tempting to speculate that STIM2 may also exhibit some of the growth-inhibitory functions that have implicated STIM1 [2] as a tumour growth suppressor. Human STIM2 maps to chromosome 4p15.1, a region implicated in human cancer. A large region of human chromosome 4p has been implicated in head and neck squamous carcinoma [29], and deletions of 4p15 have been observed in invasive breast carcinoma [30] and metastatic squamous cell carcinomas of the lung [31]. The demonstrated interaction between STIM1 and STIM2 in K562 cells suggests that these proteins function interdependently *in vivo*, while the apparently ubiquitous expression of STIM1 and STIM2 suggests that this may be a common phenomenon. Analysis of the STIM family members in their respective *in vivo* environs will be necessary to determine the functional conservation of these molecules and their shared domains across evolution.

Funding for this project was provided by the National Health and Medical Research Council (NHMRC), the Australian Research Council (ARC), and the Anti-Cancer Council of Victoria (ACCV). We thank Dr. Tim Adams (CSIRO, Melbourne, Australia) for providing 293T cells, and Dr Gary Hime (University of Melbourne), Dr Robert Kretsinger (University of Virginia, Charlottesville, VA, U.S.A.) and Dr James Bowie (UCLA, Los Angeles, CA, U.S.A.) for helpful discussions. We also thank Dr Bernie McInerney (Australian Proteome Analysis Facility, Sydney, Australia) for performing the Edman sequencing of STIM2.

REFERENCES

- Parker, N. J., Begley, C. G., Smith, P. J. and Fox, R. M. (1996) Molecular cloning of a novel human gene (D11S4896E) at chromosomal region 11p15.5. *Genomics* **37**, 253–256
- Sabbioni, S., Barbanti-Brodano, G., Croce, C. and Negrini, M. (1997) *GOK*: A gene at 11p15 involved in rhabdomyosarcoma and rhabdoid tumor development. *Cancer Res.* **57**, 4493–4497

- 3 Manji, S. S. M., Parker, N. J., Williams, R. T., van Stekelenburg, L., Pearson, R. B., Dziadek, M. and Smith, P. J. (2000) STIM1: a novel phosphoprotein located at the cell surface. *Biochim. Biophys. Acta* **1481**, 147–155
- 4 Sabbioni, S., Veronese, A., Trubia, M., Taramelli, R., Barbanti-Brodano, G., Croce, C. M. and Negrini, M. (1999) Exon structure and promoter identification of STIM1 (alias GOK), a human gene causing growth arrest of the human tumor cell lines G401 and RD. *Cytogenet. Cell Genet.* **86**, 214–218
- 5 Oritani, K. and Kincaid, P. W. (1996) Identification of stromal cell products that interact with pre-B cells. *J. Cell Biol.* **134**, 771–782
- 6 Overall, M. L., Parker, N. J., Scarcella, D. L., Smith, P. J. and Dziadek, M. (1998) Murine *Stim1* maps to chromosome 7 and is not imprinted. *Mamm. Genome* **9**, 657–659
- 7 Schultz, J., Ponting, C. P., Hofman, K. and Bork, P. (1997) SAM as a protein interaction domain involved in developmental regulation. *Protein Sci.* **6**, 249–253
- 8 Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
- 9 Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. and Wheeler, D. L. (2000) GenBank. *Nucleic Acids Res.* **28**, 15–18
- 10 Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
- 11 Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. (2000) SMART: A web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234
- 12 Tusnady, G. E. and Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: Applications to topology prediction. *J. Mol. Biol.* **283**, 489–506
- 13 Lupas, A., van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164
- 14 Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S., Thurston, S. J. and Birkenmeier, E. H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5**, 253–274
- 15 Nagase, T., Kikuno, R., Ishikawa, K., Hirosawa, M. and Ohara, O. (2000) Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* **7**, 143–150
- 16 Kozak, M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7**, 563–574
- 17 Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F. et al. (2000) Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat. Genet.* **25**, 235–238
- 18 Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. J., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195
- 19 The *C. elegans* Sequencing Consortium (1998) Genome sequencing of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018
- 20 Kretsinger, R. H. (1996) EF-hands reach out. *Nat. Struct. Biol.* **3**, 12–15
- 21 Ponting, C. P. (1995) SAM: A novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Sci.* **4**, 1928–1930
- 22 Van der Bliek, A. M. and Meyerowitz, E. M. (1991) Dynamin-like protein encoded by the *Drosophila* shibire gene associated with vesicular traffic. *Nature (London)* **351**, 411–414
- 23 Kikuno, R., Nagase, T., Suyama, M., Waki, M., Hirosawa, M. and Ohara, O. (2000) HUGE: a database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res.* **28**, 331–332
- 24 Gray, N. K. and Wickens, M. (1998) Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.* **14**, 399–458
- 25 Meyuhas, O., Avni, D. and Shama, S. (1996) Translational control of ribosomal protein mRNAs in eukaryotes. In *Translational Control* (Hershey, J. W. B., Mathews, M. B. and Sonenberg, N., eds.), pp. 363–388, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 26 Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins. *Trends Biochem. Sci.* **21**, 14–17
- 27 Tsigelny, I., Shindyalov, I. N., Bourne, P. E., Sudhof, T. C. and Taylor, P. (2000) Common EF-hand motifs in cholinesterases and neurologins suggest a role for Ca²⁺ binding in cell surface associations. *Protein Sci.* **9**, 180–185
- 28 Thanos, C. D., Faham, S., Goodwill, K. E., Cascio, D., Phillips, M. and Bowie, J. U. (1999) Monomeric structure of the human EphB2 sterile α motif domain. *J. Biol. Chem.* **274**, 37301–37306
- 29 Pershouse, M. A., El-Naggar, A. K., Hurr, K., Lin, H., Yung, W. K. A. and Steck, P. A. (1997) Deletion mapping of chromosome 4 in head and neck squamous cell carcinoma. *Oncogene* **14**, 369–373
- 30 Richard, F., Pacyna-Gengelbach, M., Schluns, K., Fleige, B., Winzer, K. J., Szymas, J., Dietel, M., Petersen, I. and Schwendel, A. (2000) Patterns of chromosomal imbalances in invasive breast cancer. *Int. J. Cancer* **89**, 305–310
- 31 Petersen, S., Aninat-Meyer, M., Schluns, K., Gellert, K., Dietel, M. and Petersen, I. (2000) Chromosomal alterations in the clonal evolution to the metastatic stage of squamous cell carcinomas of the lung. *Br. J. Cancer* **82**, 65–73