| Date de dépôt/PCT Filing Date: 2002/12/20 |
| Date publication PCT/PCT Publication Date: 2003/07/10 |
| Entrée phase nationale/National Entry: 2004/06/21 |
| N° demande PCT/PCT Application No.: GB 2002/005856 |
| N° publication PCT/PCT Publication No.: 2003/055915 |
| Priorité/Priority: 2001/12/21 (0130721.4) GB |

(54) Titre : PROTEINE TRANSMEMBRANAIRES
(54) Title: HUMAN NETRIN RECEPTOR AND USES THEREOF

1. LPESIPSAPG TLPHFIEEPO DAYIIKSNPF ALRCKARPAM QIFFKCNGEW
2. VHQNEHVEE TLDISESSLKLV REVFINVTRQ QVEDFHGEPED YWCQCVWASH
3. LGTSDSRRKAS VRVAYLRRKNE EQDFQGGREVF IEGMIVLHCR PPEGVPAAEV
4. ERLKNEEPID SEQDENIDMR ADHNLITQAO RLSDSGNYTC MAANIVAKRR
5. SLSATVYVVV NGGWSSTWEN SACNVRSCRG WQKRSRTCTN PAFNLNGAFAC
6. EGMSVQKITC TSLCPVDSGSW EVWSEWSSVC FCEWHLRIRE CTAPPBRNGG
7. KFCCELSQES ENCTDOGLCIL GJENASIAL YSGLGAAVVA VAVLVIGVTL
8. YRRSOSDGYV DVIDSIALTG GFQTFNFKTVC RQGNSLLLNS AHQDPLTVSR
9. TYSQPICLQD PLDKELMTES SLNFPLSDIK VQKQSSFMWS LGVSEARAEYH
10. GKNHSRTFPF GNNHFSFSTM PNKMPYIQON LSSLPTRTEL RTTGVFQHNLG
11. GRLVMPNTGV SLLIPHSGAP EENSWEYIMS INQGEPSLQG DGESVILLSE
12. VTCGPPDMIV TTPFALTIPF CADVSSSEHWN IHLLKRTQGQ WWEEVMSVED
13. ESTSCYCLLD PFCAVVULDS FGTLTGEIP ITDCAVQQLK VAVFHCMSCN
14. SLOYLNKVC VQNTPRCAEQ VSREDHERQGG QLLEEFKLHL FKQNTFSLQT
15. SLYDIPPLPM RIKPFTACQVE VPSRSVWCSN RQPLKCAFSL ERYTPTTQL
16. SCKCRQKLQ GHEQLQVF DILESERETI TFFAQEDSTF PAQGTGPAFK
17. IPSTIRORIC ATFDTPNAKG KDWOMLAQKN SIRNLASYFA TQSSPSAVIL
18. NLWZAHQIDO GULDSLACL EEIGRTHHTKLN SNISESQ

(57) Abrégé/Abstract:
The invention is based on the discovery that the INSP017 protein functions as a trans-membrane protein molecule, preferably as a trans-membrane protein molecule of the netrin receptor family.
Title: HUMAN NETRIN RECEPTOR AND USES THEREOF

Abstract: The invention is based on the discovery that the NSIP017 protein functions as a trans-membrane protein molecule, preferably as a trans-membrane protein molecule of the netrin receptor family.
TRANSMEMBRANE PROTEIN

This invention relates to a novel protein (INSP017), herein identified as a transmembrane protein (in particular, as a member of the netrin receptor family) and to the use of this protein and nucleic acid sequences from the encoding gene in the diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Netrin receptor family background

Migrating neuronal axons require the correct presentation of guidance molecules, often at multiple choice points, to find their target. Netrins are a family of secreted proteins which are involved in both attracting and repelling axons, and as such, function as guidance molecules to allow cell migration and axon path-finding processes in the central nervous system (Where the rubber meets the road: netrin expression and function in developing

Members of the UNC-5 protein family are transmembrane receptors for netrin guidance cues (Hong K, et al. Cell 1999 Jun 25;97(7):927-41) and are believed to play an important role in the development of the mammalian nervous system by acting as dependence receptors for netrin-1 (a diffusible attractant/repellent). INSP17 is believed to be an UNC-5 family member. The netrin 1 receptor DCC has been implicated in directing the guidance of axons toward netrin sources, whereas members of the UNC-5 family are necessary for migrations away from netrin sources (Hong K, et al. (1999); Finger JH, et al. 2002 Dec 1;22(23):10346-56).

Experimental evidence has shown that one member of the UNC-5 family, Unc5h3, plays an important role during cell migration in the developing murine cerebellum as mice homozygous for mutations in Unc5h3 are ataxic and have cerebellar hypoplasia and laminar structure defects (Ackerman, S and Knowles, B. Genomics (1998) Sep 1;52(2):205-8). Mice homozygous for mutations in UNC5H3 are ataxic and have cerebellar hypoplasia and laminar structure defects. The human homologue of the murine UNC5H3 gene localizes to chromosome 4q21-q23 and is closely linked to the Parkinson's disease gene (Ackerman, S and Knowles, B. Genomics (1998) Sep 1;52(2):205-8). Members of the UNC-5 family are type I receptors which contain a cytoplasmically located death domain. In the absence of netrin-1, cells expressing UNC-5 receptors undergo apoptosis, while in the presence of netrin-1 apoptosis is blocked; hence these receptors confer dependence on the presence of ligand in cells in which they are expressed. Recent studies have shown that UNC-5 receptors can undergo proteolytic cleavage by caspases and that this cleavage is required for apoptosis (Llambi, F et al. EMBO J (2001) 20, 2715-2722).

Transmembrane proteins of this kind have thus been shown to play a role in diverse physiological functions, many of which can play a role in disease processes. Alteration of their activity is a means to alter the disease phenotype and as such identification of novel transmembrane proteins is highly relevant as they may play a role in many diseases, particularly inflammatory disease, oncology, and cardiovascular disease. As such,
identification of novel transmembrane proteins is likely to play a role in the development of novel drugs to treat disease.

THE INVENTION

The invention is based on the discovery that the INSP017 protein functions as a transmembrane protein molecule and, moreover, as a trans-membrane protein molecule of the netrin receptor family.

In a first aspect, the invention provides a polypeptide, which polypeptide:

(i) comprises (and optionally consists of) the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and/or SEQ ID NO:34;

(ii) is a fragment thereof having activity as a transmembrane protein, particularly as a netrin receptor, or having an antigenic determinant in common with the polypeptides of (i); or

(iii) is a functional equivalent of (i) or (ii).

Preferably, the polypeptide according to this embodiment comprises the amino acid sequence recited in SEQ ID NO: 32 or SEQ ID NO: 34. More preferably, the polypeptide consists of the amino acid sequence recited in SEQ ID NO: 32 or SEQ ID NO: 34.

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as "the INSP017 exon 1 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as "the INSP017 exon 2 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:6 is referred to hereafter as "the INSP017 exon 3 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:8 is referred to hereafter as "the INSP017 exon 4 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:10 is referred to hereafter as "the INSP017 exon 5 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:12 is referred to hereafter as "the INSP017 exon 6 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:14 is referred to hereafter as "the INSP017 exon 7 polypeptide".
The polypeptide having the sequence recited in SEQ ID NO:16 is referred to hereafter as "the INSP017 exon 8 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:18 is referred to hereafter as "the INSP017 exon 9 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:20 is referred to hereafter as "the INSP017 exon 10 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:22 is referred to hereafter as "the INSP017 exon 11 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:24 is referred to hereafter as "the INSP017 exon 12 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:26 is referred to hereafter as "the INSP017 exon 13 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:28 is referred to hereafter as "the INSP017 exon 14 polypeptide". The polypeptide having the sequence recited in SEQ ID NO:30 is referred to hereafter as "the INSP017 exon 15 polypeptide". The polypeptides having the sequences recited in SEQ ID NO:32 and SEQ ID NO:34 are referred to hereafter as "the INSP017 polypeptide". The polypeptide of SEQ ID NO:32 is a partial sequence of the polypeptide having the sequences recited in SEQ ID NO:34. The term "the INSP017 polypeptide" is also used herein to collectively refer to one or more of the following polypeptides: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34.

By "having activity as a transmembrane protein" we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features within the family of transmembrane proteins.

By "having activity as a netrin receptor" we refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features within the family of netrin receptors, such that the polypeptide's activity is not substantially affected detrimentally in comparison to the function of the full length wild type polypeptide. Like other members of the Unc5-like netrin receptor family, INSP017 contains the following domains: a). (pfam00791), A ZU5 domain. Domain present in ZO-1 and Unc5-like netrin receptors Domain of unknown function; b). (pfam00531), A cytoplasmic Death domain; c). (pfam00090), A Thrombospondin type 1 domain; d). (pfam00047), An
immunoglobulin domain. Suitably, the term "having activity as a netrin receptor" refers to a polypeptide having 1, 2, 3 or 4 of the 4 aforementioned domains. It is particularly preferred that the term by "having activity as a netrin receptor" refers to polypeptides having domain a) and/or b).

Preferably, a polypeptide or nucleic acid sequence according to the invention is not that of the amino acid and nucleic acid sequences respectively of ABK88064 or KIAA1777 (see figures) or of an amino acid or nucleic acid sequence set forth in JP2002153290-A or WO 02/33080.

In a second aspect, the invention provides a purified nucleic acid molecule which encodes a polypeptide of the first aspect of the invention. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP017 exon 1 polypeptide), SEQ ID NO:3 (encoding the INSP017 exon 2 polypeptide), SEQ ID NO:5 (encoding the INSP017 exon 3 polypeptide), SEQ ID NO:7 (encoding the INSP017 exon 4 polypeptide), SEQ ID NO:9 (encoding the INSP017 exon 5 polypeptide), SEQ ID NO:11 (encoding the INSP017 exon 6 polypeptide), SEQ ID NO:13 (encoding the INSP017 exon 7 polypeptide), SEQ ID NO:15 (encoding the INSP017 exon 8 polypeptide), SEQ ID NO:17 (encoding the INSP017 exon 9 polypeptide), SEQ ID NO:19 (encoding the INSP017 exon 10 polypeptide), SEQ ID NO:21 (encoding the INSP017 exon 11 polypeptide), SEQ ID NO:23 (encoding the INSP017 exon 12 polypeptide), SEQ ID NO:25 (encoding the INSP017 exon 13 polypeptide), SEQ ID NO:27 (encoding the INSP017 exon 14 polypeptide), SEQ ID NO:29 (encoding the INSP017 exon 15 polypeptide), SEQ ID NO:31 (encoding the INSP017 polypeptide), SEQ ID NO:33 (encoding the full length INSP017 polypeptide), or is a redundant equivalent or fragment of any of these sequences.

In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention. In a preferred embodiment of this aspect of the invention the vector is the PCR4-TOPO-
INSPO17 vector (see Figure 6).

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably stimulates the activity of, a polypeptide of the first aspect of the invention.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the INSPO17 polypeptide allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in therapy or diagnosis. These molecules may also be used in the manufacture of a medicament for the treatment of cell proliferative disorders particularly brain tumours, nervous system tumours, neoplasm, bone tumor and myeloproliferative disorders particularly myeloid leukaemia, autoimmune/inflammatory disorders, cardiovascular disorders, neurological disorders, such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, psychiatric disorders, such as depression, schizophrenia, brain injury, spinal cord injury, nerve injury, developmental disorders, including disorders of nervous system development, nervous system inflammation including motor neuron disease, amyotrophic lateral sclerosis, multiple sclerosis and inflammatory neuropathy, bone disease, atherosclerosis, glomerulonephritis, cachexia, AIDS, HIV infection, metabolic disorders such as diabetes, infections, reproductive disorders, infertility, embryo implantation failure, pregnancy disorders and birth complication and other pathological conditions, particularly those in which netrin receptors are implicated.
In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out \textit{in vitro}. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a transmembrane protein, preferably as a netrin receptor. A polypeptide of the first aspect of the invention may find utility as a neuroprotectant.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a
vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.
Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.


As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in
purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.
The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the INSP017 polypeptide (preferably INSP034). Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the INSP017 polypeptide (preferably of the amino acid sequence recited in SEQ ID NO:34). Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, i.e. between 5
and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions.

Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

Typically, greater than 80% identity between two polypeptides is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the INSP017 polypeptide (preferably SEQ ID NO:34), or with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 90%, 95%, 98%, 98.5%, 99%, or 99.5% respectively.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Thresher technology that forms one aspect of the search tools used to generate the Biopendium search database may be used (see co-pending PCT patent application PCT/GB01/01105 (published as WO 01/69507)) to identify polypeptides of presently-known function which, while having low sequence identity as compared to the INSP017 polypeptide, are predicted to have secreted molecule activity, by virtue of sharing significant structural homology with the INSP017 polypeptide sequences. By "significant structural homology" is meant that the Inpharmatica Genome Thresher predicts two proteins to share structural homology with a certainty of 10% and above.

The polypeptides of the first aspect of the invention also include fragments of the INSP017 polypeptide (preferably SEQ ID NO:34) and fragments of the functional equivalents of the INSP017 polypeptide (preferably SEQ ID NO:34), provided that those fragments retain netrin receptor activity or have an antigenic determinant in common with the INSP017 polypeptide.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the INSP017
polypeptide (preferably SEQ ID NO:34) or one of its functional equivalents. The fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')2 and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may
be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., Nature 256: 495-497 (1975); Kozbor et al., Immunology Today 4: 72 (1983); Cole et al., 77-96 in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu et al., Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones et al., Nature, 321, 522 (1986); Verhoeyen et al., Science, 239, 1534 (1988); Kabat et al., J. Immunol., 147, 1709 (1991); Queen et al., Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman et al., Proc. Natl Acad. Sci. USA, 88, 34181 (1991); and Hodgson et al., Bio/Technology, 9, 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody
having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. et al., (1990), Nature 348, 552-554; Marks, J. et al., (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34 and functionally equivalent polypeptides. These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical
synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the in vitro or in vivo transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:1. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:4 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:3. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:6 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:5. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:8 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:7. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:10 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:9. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:12 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:11. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:14 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:13. A nucleic acid molecule
which encodes the polypeptide of SEQ ID NO:16 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:15. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:18 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:17. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:20 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:19. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:22 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:21. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:24 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:23. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:26 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:25. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:28 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:27. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:30 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:29. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:32 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:31. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:34 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:33. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encode a polypeptide of SEQ ID NO:2 or SEQ ID NO:4 or SEQ ID NO:6 or SEQ ID NO:8 or SEQ ID NO:10 or SEQ ID NO:12 or SEQ ID NO:14 or SEQ ID NO:16 or SEQ ID NO:18 or SEQ ID NO:20 or SEQ ID NO:22 or SEQ ID NO:24 or SEQ ID NO:26 or SEQ ID NO:28 or SEQ ID NO:30, SEQ ID NO:32 or SEQ ID NO:34. Such nucleic acid molecules may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the
transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

5 The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention. Such a nucleic acid molecule may be a naturally-occurring variant such as a naturally-occurring allelic variant, or the molecule may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

25 Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a
commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

5 The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee et al., Nucleic Acids Res 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); Dervan et al., Science 251, 1360 (1991).

10 The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook et al. [supra]).

15 The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook et al [supra]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5X SSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook et al. [supra]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70%, 80%, 85%, 90%, 95%, or 98% identical over their entire length to a nucleic acid molecule encoding the INSP017 polypeptide (SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1, SEQ ID NO:3, or a nucleic acid molecule that is complementary thereto. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological function or activity as the INSP017 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according
to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the INSP017 polypeptide and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the INSP017 polypeptide is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel et al. (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 or SEQ ID NO:33), are particularly useful probes. Such probes may be labelled with an analytically-detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation
of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Appl. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Appl., 1, 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinderTM libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for
extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be
transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook et al (supra) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook et al., (supra). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco
mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al., [supra]. Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook et al., 1989 [supra]; Ausubel et al., 1991 [supra]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible
promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportITM plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences.

Resistant clones of stably transformed cells may be proliferated using tissue culture
techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as Drosophila S2 and Spodoptera Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, Phytochemistry 30, 3861-3863 (1991).

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, staphylococci, E. coli, Streptomyces and Bacillus subtilis cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, S. cerevisiae) and Aspergillus cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase
(Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes that can be employed in tk- or aprt± cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing
labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble
proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. et al. (1992), Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free
preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

The INSP017 polypeptides of the present invention may modulate a variety of physiological and pathological processes, including neurological processes (e.g. axon guidance in the CNS and PNS, neuron disorders) and the development or modulation of the nervous system. Thus, the biological activity of the INSP017 polypeptides can be examined in systems that allow the study of such modulatory activities, using a variety of suitable assays.

For example, for observing cell growth inhibition, one can use a solid or liquid medium. In a solid medium, cells undergoing growth inhibition can easily be selected from the
subject cell group by comparing the sizes of colonies formed. In a liquid medium, growth inhibition can be screened by measuring culture medium turbidity or incorporation of labelled thymidine in DNA. Typically, the incorporation of a nucleoside analog into newly synthesised DNA may be employed to measure proliferation (i.e., active cell growth) in a population of cells. For example, bromodeoxyuridine (BrdU) can be employed as a DNA labelling reagent and anti-BrdU mouse monoclonal antibodies can be employed as a detection reagent. This antibody binds only to cells containing DNA which has incorporated bromodeoxyuridine. A number of detection methods may be used in conjunction with this assay including immunofluorescence, immunohistochemical, ELISA, and colorimetric methods. Kits that include bromodeoxyuridine (BrdU) and anti-BrdU mouse monoclonal antibody are commercially available from Boehringer Mannheim (Indianapolis, IN).

A preferred method for identifying an agonist or antagonist compound of a polypeptide of the present invention comprises:

(a) contacting a cell expressing on the surface thereof the polypeptide according to the first aspect of the invention, the polypeptide being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and

(b) determining whether the compound binds to and activates or inhibits the polypeptide by measuring the level of a signal generated from the interaction of the compound with the polypeptide.

A further preferred method for identifying an agonist or antagonist of a polypeptide of the invention comprises:

(a) contacting a cell expressing on the surface thereof the polypeptide, the polypeptide being associated with a second component capable of providing a detectable signal in response to the binding of a compound to the polypeptide, with a compound to be screened under conditions to permit binding to the polypeptide; and

(b) determining whether the compound binds to and activates or inhibits the polypeptide
by comparing the level of a signal generated from the interaction of the compound with the polypeptide with the level of a signal in the absence of the compound.

In further preferred embodiments, the general methods that are described above may further comprise conducting the identification of agonist or antagonist in the presence of labelled or unlabelled ligand for the polypeptide.

In another embodiment of the method for identifying agonist or antagonist of a polypeptide of the present invention comprises:

determining the inhibition of binding of a ligand to cells which have a polypeptide of the invention on the surface thereof, or to cell membranes containing such a polypeptide, in the presence of a candidate compound under conditions to permit binding to the polypeptide, and determining the amount of ligand bound to the polypeptide. A compound capable of causing reduction of binding of a ligand is considered to be an agonist or antagonist. Preferably the ligand is labelled.

More particularly, a method of screening for a polypeptide antagonist or agonist compound comprises the steps of:

(a) incubating a labelled ligand with a whole cell expressing a polypeptide according to the invention on the cell surface, or a cell membrane containing a polypeptide of the invention,

(b) measuring the amount of labelled ligand bound to the whole cell or the cell membrane;

(c) adding a candidate compound to a mixture of labelled ligand and the whole cell or the cell membrane of step (a) and allowing the mixture to attain equilibrium;

(d) measuring the amount of labelled ligand bound to the whole cell or the cell membrane after step (c); and

(e) comparing the difference in the labelled ligand bound in step (b) and (d), such that the compound which causes the reduction in binding in step (d) is considered to be an agonist or antagonist.
The polypeptides may be found to modulate a variety of physiological and pathological processes in a dose-dependent manner in the above-described assays. Thus, the "functional equivalents" of the polypeptides of the invention include polypeptides that exhibit any of the same modulatory activities in the above-described assays in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the polypeptides of the invention, preferably the "functional equivalents" will exhibit substantially similar dose-dependence in a given activity assay compared to the polypeptides of the invention.

In certain of the embodiments described above, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be
detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound \([X]\) is "substantially free of" impurities \([Y]\) when at least 85% by weight of the total \(X+Y\) in the composition is \(X\). Preferably, \(X\) comprises at least about 90% by weight of the total of \(X+Y\) in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective
amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targeted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).
Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.
In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated in situ from expression in vivo.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, et al., Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible
modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2'-O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur \textit{in vivo} or \textit{ex vivo}. \textit{Ex vivo} gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, \textit{in vivo} gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell.
transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, H. pylori, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried
condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki et al., Nature, 324, 163-166 (1986); Bej, et al., Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer et al., J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;

b) contacting a control sample with said probe under the same conditions used in step a);

c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.
A further aspect of the invention comprises a diagnostic method comprising the steps of:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to the invention from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita et al., Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of
this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers et al., Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton et al., Proc. Natl. Acad. Sci. USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by in situ analysis (see, for example, Keller et al., DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence in situ hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck et al., Science, 250, 559-562 (1990), and Trask et al., Trends, Genet., 7, 149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science (1996), Vol 274, pp 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al); Lockhart, D. J. et al. (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process.

The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or
any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex
formation. The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

(a) a nucleic acid molecule of the present invention;

(b) a polypeptide of the present invention; or

(c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.
In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly cell proliferative disorders particularly brain tumours, nervous system tumours, neoplasm, bone tumor and myeloproliferative disorders particularly myeloid leukaemia, autoimmune/inflammatory disorders, cardiovascular disorders, neurological disorders, such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, psychiatric disorders, such as depression, schizophrenia, brain injury, spinal cord injury, nerve injury, developmental disorders, including disorders of nervous system development, nervous system inflammation including motor neuron disease, amyotrophic lateral sclerosis, multiple sclerosis and inflammatory neuropathy, bone disease, atherosclerosis, glomerulonephritis, cachexia, AIDS, HIV infection, metabolic disorders such as diabetes, infections, reproductive disorders, infertility, embryo implantation failure, pregnancy disorders and birth complication and other pathological conditions, particularly those in which netrin receptors are implicated.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to INSP017 polypeptide.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

**Brief description of the Figures**

Figure 1: Polypeptide sequence of INSP017 (original prediction, i.e. SEQ ID NO:32).

Figure 2: Top ten BLAST hits at NCBI-nr using INSP017 (SEQ ID NO:32) as query.

Figure 3: Alignment for the top BLAST hit with the Polypeptide sequence of INSP017 (SEQ ID NO:32).

Figure 4: Predicted nucleotide sequence of INSP017 with translation
Figure 5: INSP017 full length cDNA identified by RACE PCR. The shaded sequence represents original prediction of INSP017 (SEQ ID NO.34)

Figure 6: Map of PCR4-TOPO-INSP017

Figure 7: Predicted (SEQ ID NO. 32) and cloned (SEQ ID NO.34) sequences. Shaded sequence in the cloned sequence is absent from the original INSP017 prediction.

Figure 8: Predicted (SEQ ID NO. 32) and cloned (SEQ ID NO. 34) INSP017 sequences blasted against each other.

Figure 9. Tblastn of cloned INSP017 sequence (SEQ ID NO: 34) against NCBI-nr.

Figure 10. ESTs supporting our prediction. Tblastn of the cloned INSP017 sequence (SEQ ID NO: 34) sequence against NCBI-est.

Examples

Example 1 INSP017

The polypeptide sequence derived from combining SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, and SEQ ID NO:30 which represent the translation of consecutive exons from INSP017 (Figure 1) were used as a query in a BLAST homology search against the NCBI-nr database, the top ten results are shown (Figure 2). The top match represents the sequence of a UNC-5 netrin receptor (Figure 3).

It is believed that INSP017 functions as a transmembrane protein, preferably as a netrin receptor. These experimental predictions will be confirmed subsequently by a directed experimental test. For example, a number of different assays may be used to test for netrin receptor activity.

Example 2.

1 Cloning of INSP017

1.1 cDNA libraries
Human cDNA libraries (in bacteriophage lambda (λ) vectors) were purchased from Stratagene or Clontech or prepared at the Serono Pharmaceutical Research Institute in λ ZAP or λ GT10 vectors according to the manufacturer’s protocol (Stratagene). Bacteriophage λ DNA was prepared from small scale cultures of infected *E.coli* host strain using the Wizard Lambda Preps DNA purification system according to the manufacturer’s instructions (Promega, Corporation, Madison WI.) The list of libraries and host strains used is shown in Table 1. Eight pools (A-H) of five different libraries (100 ng/μl phage DNA) were used in subsequent PCR reactions.

1.2 PCR of virtual cDNAs from phage library DNA

A cDNA encoding 3 exons of INSP017 (Figure 4) was obtained as a PCR amplification product of 357 bp using gene specific cloning primers (CP1 and CP2, Figure 4 and Table 2). The PCR was performed in a final volume of 50 μl containing 1X AmpliTaq™ buffer, 200 μM dNTPs, 50 pmoles each of cloning primers primers, 2.5 units of AmpliTaq™ (Perkin Elmer) and 100 ng of each phage library pool DNA using an MJ Research DNA Engine, programmed as follows: 94 °C, 1 min; 40 cycles of 94 °C, 1 min, *x* °C, and *y* min and 72 °C, (where *x* is the lowest Tm – 5 °C and *y* = 1 min per kb of product); followed by 1 cycle at 72 °C for 7 min and a holding cycle at 4 °C.

The amplification products were visualized on 0.8 % agarose gels in 1 X TAE buffer (Life Technologies) and PCR products migrating at the predicted molecular mass were purified from the gel using the Wizard PCR Preps DNA Purification System (Promega). PCR products eluted in 50 μl of sterile water were either subcloned directly or stored at – 20 °C.

1.3 Gene specific cloning primers for PCR

Pairs of PCR primers having a length of between 18 and 25 bases were designed for amplifying the full length sequence of the virtual cDNA using Primer Designer Software (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, USA). PCR primers were optimized to have a Tm close to 55 ± 10 °C and a GC content of 40-60%. Primers were selected which had high selectivity for the target sequence INSP017 (little or no none specific priming).
1.4 **Subcloning of PCR Products**

PCR products were subcloned into the topoisomerase I modified cloning vector (pCR 4 blunt TOPO) using the Zero blunt cloning kit purchased from the Invitrogen Corporation (cat. No. K2875) using the conditions specified by the manufacturer. Briefly, 4 µl of gel purified PCR product from the human library pool A amplification was incubated for 15 min at room temperature with 1 µl of TOPO vector and 1 µl salt solution. The reaction mixture was then transformed into *E. coli* strain TOP10 (Invitrogen) as follows: a 50 µl aliquot of One Shot TOP10 cells was thawed on ice and 2 µl of TOPO reaction was added. The mixture was incubated for 15 min on ice and then heat shocked by incubation at 42 °C for exactly 30 s. Samples were returned to ice and 250 µl of warm SOC media (room temperature) was added. Samples were incubated with shaking (220 rpm) for 1 h at 37 °C. The transformation mixture was then plated on L-broth (LB) plates containing ampicillin (100 µg/ml) and incubated overnight at 37 °C. Ampicillin resistant colonies containing cDNA inserts were identified by colony PCR.

1.5 **Colony PCR**

Colonies were inoculated into 50 µl sterile water using a sterile toothpick. A 10 µl aliquot of the inoculum was then subjected to PCR in a total reaction volume of 20 µl as described above, except the primers used were T3 and T7. The cycling conditions were as follows: 94 °C, 2 min; 30 cycles of 94 °C, 30 sec, 47 °C, 30 sec and 72 °C for 1 min); 1 cycle, 72 °C, 7 min. Samples were then maintained at 4 °C (holding cycle) before further analysis.

PCR reaction products were analyzed on 1 % agarose gels in 1 X TAE buffer. Colonies which gave the expected PCR product size (357 bp cDNA + 105 bp due to the multiple cloning site or MCS) were grown up overnight at 37 °C in 5 ml L-Broth (LB) containing ampicillin (50 µg /ml), with shaking at 220 rpm at 37 °C.

1.6 **Plasmid DNA preparation and Sequencing**

Miniprep plasmid DNA was prepared from 5 ml cultures using a Qiaprep Turbo 9600 robotic system (Qiagen) or Wizard Plus SV Miniprep kit (Promega cat. no. 1460) according to the manufacturer’s instructions. Plasmid DNA was eluted in 100 µl of sterile
water. The DNA concentration was measured using an Eppendorf BO photometer. Plasmid DNA (200-500 ng) was subjected to DNA sequencing with T7 primer and T3 primer using the BigDyeTerminator system (Applied Biosystems cat. no. 4390246) according to the manufacturer’s instructions. Sequencing reactions were purified using Dye-Ex columns (Qiagen) or Montage SEQ 96 cleanup plates (Millipore cat. no. LSKS09624) then analyzed on an Applied Biosystems 3700 sequencer.

1.7 Identification of the full length coding sequence of INSPO17 using RACE PCR.

The predicted sequence of the INSPO17 ORF did not contain a start codon, and also lacked the stop codon at the 3’ end of the coding sequence (Figure 4). RACE PCR was therefore used to extend the 5’ and 3’ ends of the prediction to identify the full coding sequence. RACE PCR was performed on RACE ready cDNA from testis (Invitrogen cat. No. L1510-13) using a Gene Racer kit (Invitrogen) according to the manufacturer’s instructions. For amplification of 5’ ends, the first PCR was performed in a 50 μl reaction volume containing 1 μl RACE Ready cDNA, 5 μl of 10X High Fidelity buffer, 1 μl of dNTPs (10 mM), 2 μl of 50 mM MgSO4, 3 μl of GeneRacer 5’1 primer (10 μM), 1 μl of reverse gene specific primer (63765-GR1-5’) (10 μM) and 2.5 units (0.5 μl) of Platinum Taq DNA polymerase Hi Fi (Invitrogen). The cycling conditions were as follows: 94°C, 2 min; 5 cycles of 94°C 30 s and 72°C 2min; 5 cycles of 94°C, 30 s and 70°C, 5 min; 25 cycles of 94°C, 30 s, 65°C 30 s and 68°C 5 min; a final extension at 68°C for 10 min and a holding cycle of 4°C. One μl of the amplification reaction was then used as a template for a nested PCR which was performed in a final reaction volume of 50 μl with the same reagents as above except for the primers. The primers for the nested PCR were 1 μl of GeneRacer 5’ nested primer (10 μM) and 1 μl of reverse nested gene specific primer (63765-GR1nest-5’) (10 μM). The cycling conditions were 94°C, 2 min; 25 cycles of 94°C, 30 s, 65°C, 30 s and 68°C, 5 min; a final extension at 68°C for 10 min and a holding cycle of 4°C. PCR products were gel purified and subcloned into pCR4-bluntTOPO vector as described above. For amplification of 3’ ends, the same reaction conditions were used as for the 5’ RACE except that the primers were as follows: for PCR1: 3 μl of Gene Racer 3’ primer (10 μM) and 1 μl of 63765-GR1-3’ primer (10°C
μM). For PCR 2 (nested PCR): 1 μl of Gene Racer nested 3’ primer (10 μM) and 1 μl of 63765-GR1nest-3’ primer (10 μM). All primers are listed in Table II.

1.8 Cloning of the full length coding sequence of INSP017 by PCR

The full length cDNA encoding INSP017 identified by RACE PCR is shown in figure 5. The ORF of INSP017 was cloned from human Gene Racer ready testis cDNA by PCR in a 50 μl PCR reaction mixture as containing 1 μl RACE Ready cDNA, 5 μl of 10X High Fidelity buffer, 1 μl of dNTPs (10 mM), 2 μl of 50 mM MgSO4, 0.5 μl of gene specific primer 63765-FL-F (50 pmole), 1 μl of reverse gene specific primer 63765-FL-R (50 pmole) and 2.5 units (0.5 μl) of Platinum Taq DNA polymerase Hi Fi (Invitrogen). The cycling conditions were 94°C, 2 min; 35 cycles of 94°C, 30 s, 65°C, 30 s and 68°C, 5 min; a final extension at 68°C for 10 min and a holding cycle of 4°C. The amplification products were visualized on 0.8 % agarose gels in 1 X TAE buffer (Life Technologies) and PCR products migrating at the predicted molecular mass (2929 bp) were purified from the gel using the Wizard PCR Preps DNA Purification System (Promega). PCR products were eluted in 50 μl of sterile water and subcloned into pCR4 blunt TOPO vector as described in section 1.4. Several ampicillin resistant colonies were subjected to colony PCR as described in section 1.5 except that the extension time in the amplification reaction was 3 min. Colonies containing the correct size insert (2929 bp +105 bp due to the MCS) were grown up overnight at 37 °C in 5 ml L-Broth (LB) containing ampicillin (50 μg /ml), with shaking at 220 rpm at 37 °C. Miniprep plasmid DNA was prepared from 5 ml cultures using a Qiaprep Turbo 9600 robotic system (Qiagen) or Wizard Plus SV Minipreps kit (Promega cat. no. 1460) according to the manufacturer’s instructions and 200-500 ng of mini-prep DNA was sequenced as described in section 1.6 with T3 and T7 primers and INSP017 specific primers (Table 3). The map of the resultant plasmid, pCR4 TOPO-INS017 (plasmid ID. No. 13287) is shown in Figure 6.

2. Identification of cDNA libraries containing INSP017

PCR products obtained with CP1 and CP2 and migrating at the correct size (357 bp) were identified in library pool A (libraries 6, 7, 8, 9 and 10) and in testis cDNA.
<table>
<thead>
<tr>
<th>Library</th>
<th>Tissue/cell source</th>
<th>Vector</th>
<th>Host strain</th>
<th>Supplier</th>
<th>Cat. no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>human fetal brain</td>
<td>Zap II</td>
<td>XL1-Blue MRF*</td>
<td>Stratagene</td>
<td>936206</td>
</tr>
<tr>
<td>2</td>
<td>human ovary</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1098a</td>
</tr>
<tr>
<td>3</td>
<td>human pituitary</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1097a</td>
</tr>
<tr>
<td>4</td>
<td>human placenta</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1075b</td>
</tr>
<tr>
<td>5</td>
<td>human testis</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1010b</td>
</tr>
<tr>
<td>6</td>
<td>human substantia nigra</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>human fetal brain</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>human cortex brain</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>human colon</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1034a</td>
</tr>
<tr>
<td>10</td>
<td>human fetal brain</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1065a</td>
</tr>
<tr>
<td>11</td>
<td>human fetal lung</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1072a</td>
</tr>
<tr>
<td>12</td>
<td>human fetal kidney</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1071a</td>
</tr>
<tr>
<td>13</td>
<td>human fetal liver</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1064a</td>
</tr>
<tr>
<td>14</td>
<td>human bone marrow</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1058a</td>
</tr>
<tr>
<td>15</td>
<td>human peripheral blood monocytes</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1050a</td>
</tr>
<tr>
<td>16</td>
<td>human placenta</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>human SHSY53</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>human U373 cell line</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>human CFPoc-1 cell line</td>
<td>Uni Zap XL1-Blue MRF*</td>
<td>Stratagene</td>
<td>936206</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>human retina</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech</td>
<td>HL1132a</td>
</tr>
<tr>
<td>21</td>
<td>human urinary bladder</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell Type</td>
<td>Host</td>
<td>Transformation</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------</td>
<td>-------</td>
<td>----------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>human platelets</td>
<td>Uni Zap</td>
<td>XL1-Blue MRF'^</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>human neuroblastoma Kan + TS</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>human bronchial smooth muscle</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>human bronchial smooth muscle</td>
<td>GT10</td>
<td>LE392</td>
<td>in house</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>human Thymus</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1127a</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>human spleen 5' stretch</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech HL1134b</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>human peripheral blood monocytes</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1050a</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>human testis</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1065a</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>human fetal brain</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1065a</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>human substantia nigra</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1093a</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>human placenta#11</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech HL1075b</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>human Fetal brain</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech custom</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>human placenta #59</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL5014a</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>human pituitary</td>
<td>GT10</td>
<td>LE392</td>
<td>Clontech HL1097a</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>human pancreas #63</td>
<td>Uni Zap XR</td>
<td>XL1-Blue MRF'^</td>
<td>Stratagene 937208</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>human placenta #19</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech HL1008</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>human liver 5'stretch</td>
<td>GT11</td>
<td>LE392</td>
<td>Clontech HL1115b</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>human uterus</td>
<td>Zap-CMV XR</td>
<td>XL1-Blue MRF'^</td>
<td>Stratagene 980207</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>human kidney large-insert cDNA library</td>
<td>TriplEx2</td>
<td>XL1-Blue</td>
<td>Clontech HL5507u</td>
<td></td>
</tr>
</tbody>
</table>
### Table II

<table>
<thead>
<tr>
<th>Primer</th>
<th>Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP1</td>
<td>3D1</td>
<td>GCG CAT AGC CTA TTT ACG GA</td>
</tr>
<tr>
<td>CP2</td>
<td>3D2</td>
<td>TCC TCT ACC ACA GCG AAC AT</td>
</tr>
<tr>
<td>GeneRacer 5'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeneRacer 5' nested</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63765-GR1-5'</td>
<td>1A9</td>
<td>TGA GTC CGA GAG CCG CTG CTG CCT GAT GAT</td>
</tr>
<tr>
<td>GeneRacer 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeneRacer 3' nested</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63765-GR1nest-5'</td>
<td>1A10</td>
<td>TAT GGT CAG CCC TGG TGT CAA TGT TCT C</td>
</tr>
<tr>
<td>63765-GR1-3'</td>
<td>3G9</td>
<td>AAT GGA GGC TGG TCT TCC TGG ACA GAG</td>
</tr>
<tr>
<td>GeneRacer 3' nested</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63765-GR1nest-3'</td>
<td>3G10</td>
<td>GCC TGC AAT GTT CGC TGT GGT AGA G</td>
</tr>
<tr>
<td>63765FL-F</td>
<td>9D11</td>
<td>GCA GTC CAG CTC ACA GGT TA</td>
</tr>
<tr>
<td>63765FL-R</td>
<td>9D12</td>
<td>CTG GCC ATC ACT CTG TCT CA</td>
</tr>
<tr>
<td>T3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td></td>
<td>TAA TAC GAC TCA CTA TAG GG</td>
</tr>
</tbody>
</table>
Table III
Sequencing primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3</td>
<td>ATT AAC CCT CAC TAA AGG GA</td>
</tr>
<tr>
<td>T7</td>
<td>TAA TAC GAC TCA CTA TAG GG</td>
</tr>
<tr>
<td>63765 F6</td>
<td>GGA AGT GTG GAG CGA ATG GT</td>
</tr>
<tr>
<td>63765 F7</td>
<td>GTC CAG AGC TCG TTC ATG GT</td>
</tr>
<tr>
<td>63765 F8</td>
<td>TTG CGT GTC ATG TGC TCC TG</td>
</tr>
<tr>
<td>63765 F9</td>
<td>CCA AGT GCA GAC ATC AAT CC</td>
</tr>
<tr>
<td>63765 F10</td>
<td>AGC TGA GTA CCA CGG CAA GA</td>
</tr>
<tr>
<td>63765 F11</td>
<td>ACC TGT GGT CCT CCA GAC AT</td>
</tr>
<tr>
<td>63765 F12</td>
<td>TGT GCC GTG AAG CAA CTG AA</td>
</tr>
</tbody>
</table>
Sequence Listing

SEQ ID NO: 1 (INSP017 - Nucleotide sequence exon 1)
1  CTTCGCCGAT CATCGCCCTAC GCAGCTCAGT GAAGGCTGCTA ATGCCCATAG
5  GAAGCTGCTG TTTATGCTTAT CAGAAGAGA TGCAGAAGGAG GACACCTGCT
10  GCAACCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT
15  GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG
20  CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT
25  GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG

SEQ ID NO: 2 (INSP017 - Protein sequence exon 1)
1  LPESIPSAPG TLPHFIEEPD DAIYIKSNPI ALRCKARFPAM QIFFKNGWEW
10  VHHQNEHVEE TLDESSG

SEQ ID NO: 3 (INSP017 - Nucleotide sequence exon 2)
1  GTTGTGAAGGT CCAGCCGAATG TACATTCGGAA GACAGGTGAGA
5  GACCCCTCCAT CGATGGGTGGC GATGGAGGTT CCCTGGCCCA
10  CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT
15  GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG
20  CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT
25  GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG CTGCTGCTGCT GCTGCTGCTG

SEQ ID NO: 4 (INSP017 - Protein sequence exon 2)
1  LKVREVPINV TQQVEDFHG PEDYMCGGVA WSHLGTIERS KASVRIAY

SEQ ID NO: 5 (INSP017 - Nucleotide sequence exon 3)
1  ATTTACGCAA AAACCTTGAA CAGGAAACCA AAGGAAAGGA AGTTCCCAATT
5  GTAAGGACATGA CTGTACTGCA TCGCCGCCTAC CCAGAGAGGAG TCTCTGCTGC
10  CGA

SEQ ID NO: 6 (INSP017 - Protein sequence exon 3)
1  LRKNFEQDPQ GREVPFYMGI VHLCHDPGEV PAAE

SEQ ID NO: 7 (INSP017 - Nucleotide sequence exon 4)
1  GTGGAATGGCC TAAATAATGA AGGCCCATTC GACTCTGAAC AAGACAGGAA
5  TGCCAGACCTG TAAACTGAT CAGCAGACCT GCCGCCCTCT
10  CGAAGCTGCGG AAAATCTGCC TGGACGCAGTG CCAACATCTTG GAATAGGAG
15  AGACAAGCTTG CGCCACGCTCT TGTTGTCTAC G

SEQ ID NO: 8 (INSP017 - Protein sequence exon 4)
1  VLEKNNEEPI DSEQDENIDT RADHNLIIQR ARLSDSGNYT CMAANIVAKR
10  RLSLVVVVY V

SEQ ID NO: 9 (INSP017 - Nucleotide sequence exon 5)
1  TGAATGGCAGC CTGGCTTTCC TGGACAGAGT GGTACACGCG CAAATTCTCG
5  TGGCTGCCAG CATGTGCCAG ACCTGCCGAC ACGAGCAGAC GACGCAGCAC
10  TCTGATGGGT GGGGCTTCGG TGGAGGAAAT GTCGCTGAGA AAAATAACCT
15  GCACATCTCT TGGCTGCTG

SEQ ID NO: 10 (INSP017 - Protein sequence exon 5)
1  MGWSSWNETW SACNWRCGGQ WQKRSRTCTN PAPLNGGAF CEGMSVQKITC
5  TSLCPFV

SEQ ID NO: 11 (INSP017 - Nucleotide sequence exon 6)
1  TGAGTGAGGAC CGGAAGTCG TGGAGCGAAGT GGTCAAGCTG CAGTCCAKAG
5  TGGAGAATCC TGGCGGACTCG GGAGTGCCAC GACAGCCACCC CGAGAAATGG
10  GGGCAATTC TGTGAGAAGTC TAAACGCGGAG ATCCTGAAAAG TCGCAAGATT
15  GCTCTTGGCAT CTCAG

SEQ ID NO: 12 (INSP017 - Protein sequence exon 6)
1  DGSWEAVSWEN SVCSPCCHL RIERCTAPPF RNGKFCEGL SQESENCTDG


SEQ ID NO:13 (INSPIRE - Nucleotide sequence exon 7)
1 GAATTGAGAA TGCCACGCAG ATTAGCTTGT ACTCGGGCTT GGGTGCTGCC
5 GTCTGGTCTC TTCCTCTCAT GCCATCGTGC GGGCAGTACC
10 CCAAGAGGTG CAATCGGTCT CGATACGGGA TTTACGAAAA CGCAGGACT
15 GTCTAGCTGA AGGTGGAGAT GCTGGGTCTC TCCAGGATCT GTAACCTGCT
20 GCTGCGTGCG CGACGTGAGG TGGTACGATC

SEQ ID NO:14 (INSPIRE - Protein sequence exon 7)
1 IEASDIALY SGLGAAAVAV AVLWIGVTLY RRSQSDYGVD VIDSSALTGG
5 FQFNFNKTVR QG

SEQ ID NO:15 (INSPIRE - Nucleotide sequence exon 8)
1 GTAACTCCCT GCTCTGTGAAT TGCTGGCATGC AGCCAGATCT GACAGTGACC
5 CCGACATAC ACGGACCACCAT CGTGTGCGAG GACCCCTCTGG ACAAGGAGCT
10 CATGAAGAAG TTCCACACCT TTAACCCTTT GTCGGACATC AAAGTGAAGG
15 TCCAGAGCTC GTCTATGTTT TCCCTGGAGG TGTCTGAGAG AGCTGAGTAC
20 CACGGCAAGA ATACATTCCG GACCTTTTCC CATGGAAACA ACCACGCTT
25 TAGTACAAGT CATCCCGAGA ATAAAAAGCC ATCAATCCAA AATCTGCTAT
30 CACCTCCCCA AAAGACAGGA CTGGAGCACA CTGGTGTCTT TGGCCATTAT
35 GGGGGCCGCT TAGTAAATGCG AAATACAG

SEQ ID NO:16 (INSPIRE - Protein sequence exon 8)
1 NSSLNNSAMQ POLTVSRTYS GFLCQLQDQLD KELMTESLLF NPLELDIYKV
5 QSMVSYSGY SERAEYHGNK HSRTBFHGNM HSFNMHFRNN KMPYIQNLSS
10 LPRTRELRTT GVFSGHGLGRL VMPNTG

SEQ ID NO:17 (INSPIRE - Protein sequence exon 9)
1 GGGTGAGCTT ACTCATACCA CACCGGTCGA TCCCCAGAGA GAATTCTTGG
5 GACATTCTATA TGTCACATCA CCAAGGTTAACC CCAAG

SEQ ID NO:18 (INSPIRE - Protein sequence exon 9)
1 VSLLLIPHGA PEENSWEIYM SINQGEPS

SEQ ID NO:19 (INSPIRE - Nucleotide sequence exon 10)
1 CCTCCAGTCTA AGTGGCGCTGT AGGTGCTGCTT GACGGTTTGA GTCACCTGAT
5 GTCCACCAGCA GGTATGCTGC ACCACTGTCT TGGATGCAGAT CATCCCCCGC
10 GTGCAGATAG TCAAGCTGGA GATTTGGAAT ATCCATTFGA AGAAGGAGAC
15 ACACGAGGCC AAATGCGAG

SEQ ID NO:20 (INSPIRE - Protein sequence exon 10)
1 LQSDGSEVILL SPEVTCDPPO MIVPFFYFALT IPHCHAVSSE HWNIHLKRT
5 QQKREW

SEQ ID NO:21 (INSPIRE - Nucleotide sequence exon 11)
1 GAAGTGATGT CAGGGGAAAG GGAATCTACA TCCGGCTACT GCCTTTTGGGA
5 CCCCTCTCCG TGATGGCTTCT CTCGAGCTGC TGGGTTGCAAC TATCCCCCTCA
10 CAGGAGACGC AACACAGAGC TGGCGCTGTA AGCAGCTGCA GGGCGGGTT
15 TGGGCGTGA TGGCCTCGTAA ATCCGGGAGG TACAACCTTGA GAGTCTCTCG
20 CTTGCGACAAAT ACCGCTCTTG AATTCCAG

SEQ ID NO:22 (INSPIRE - Protein sequence exon 11)
1 EVMSVEDEST SCYCLLDPFA CVLDDSFGT YAEGPLID ITAVKQLKV
5 FGCHSNSLSD YNLRYVYCDN TPCAFQ

SEQ ID NO:23 (INSPIRE - Nucleotide sequence exon 12)
1 GAAGTGATTTTT CAGATGGAAAG GCATCGAAGG TGGAGAGATCC TGGAGAAACC
51 AAAAAATTGCTG CATTTCAAAAG GGAATACCTT TAGTCTTGAG ATTTCTGTCC
101 TGTGATATTCC CCAATTCCAAC TGGAGATATT AACCACTTAC TGGCTTGCCAG

SEQ ID NO:24 (INSPO17 - Protein sequence exon 12)
5 EVVSDERHOG QQLLEEEKLL HFKGNTSLQ ISVLDIPPEL WRIKPFTACQ

SEQ ID NO:25 (INSPO17 - Nucleotide sequence exon 13)
1 GGAGTCCCGT TCCCTCGGCT GTGTTGCAAT ACCGCAGCCAC CCCCTGCAGT
10 TGGCCTTTCC TGGAGGGCTG TACTGATCCAC TACCAACGCC TGGCTGGCAG
151 AAATCTGACAT CAGGAGCCATG ACAAACAGT CCAAGGCTAG AAGATCAG

SEQ ID NO:26 (INSPO17 - Protein sequence exon 13)
5 EVFPSVRWCS NRQPLHCAFS LEYRTPTTTQ LSCKICIRQL KGHEQILQVQ
15 TSILE

SEQ ID NO:27 (INSPO17 - Nucleotide sequence exon 14)
1 AGTGAACGCAG AAACACCATC TTTCTCAGCA AGAGGACAGC GCACCTTCCC
10 TGGCAGACGT CACAGGCAAAAT CTTCTAGAAT TCCCTACTCC ATCACAGACGC
20 GGATTGCTGC CAGAATTCTGACTACCCATG ACAAAGCCAA GACACTGCCAG
151 ATGTAGACAC AGAAAAACAG CATAACCCAG

SEQ ID NO:28 (INSPO17 - Protein sequence exon 14)
5 SERETITFFA QDSTFPAPQT GPFAKIFYP S IQRICATED TPNAKGKWQ
25 MIAQRKINSIR

SEQ ID NO:29 (INSPO17 - Nucleotide sequence exon 15)
1 GAAATTATCT TATTCTGCTA CAAAAATGAG CACATCTGCT GTCAATTAG
10 GCTTGCTCGGA AGTGGTAGCG TAACTGGAAG TACTGAAAGT TGGGTTCCGG
20 TGCCTTGGC AGAGGAGCCA AAGGATACGG TACCTGGCC AGAGGGCTGC
30 AGATAAGCCAT GCAGAGGCAAG GGAAGCTTGC TGCTGGCATG CCTATTACG
45 GAAAGACCTT GAAACAGACCC CACAGAAGAG GGAAGCTTCC ATGGGAAGCA
50 GAAATGCTGCT GCACCCCTGGC CCACACAGTAG GATCTAGCTG TGGCAGGCTG
55 GAATGCTGCTA AAAAATGGAAG GCGACTGAGT TCTGAACATG AGCAAGACAT
101 TGACAGCCAT GCAGACGCAA AATCGTACAT GTGGATCGAC CAGCCGCTGC
55 AATCGAGCAGA TTACACTGCTG ATGGACGCCA ATACCGGGCT TAAGACAGA
60 AGCCTGGAGC GCATGCTGCTG GTCTACGCTG ATGGAGGCCT GTGTCTCGTG
65 GACAGAGCTG TCGACGCTGG ATGTGCTGGT TGGTAAAGGA TGCGGAAAC
70 GTGGCCGCCG AGCTGACCCAC GCAGCTTCCC TCAATGGTGG GGGCTTCTTG
75 GAGAGAAATGT CAGTGCTGAA AATAACCTCC ATCTTCTCTT GTCTGGAGGA
80 TTGCGAGGCTG GAGAAGCTGGA GGCAAGACTG CTGTGCTGAGT CAGAGGTG
55 AGAATTGTGGA TCCGGGCGAG CTGCCGCTGACA AATGCTGGG CAGC
90 ATACCTGAGA AGGTCTAGAG CCAGAATTGC CAGAACTGCA CAGAYGGCT
95 TTGCACTCTAG GGCATGCGAG ATGCCGACCA TATGGTCTGT TACTGCGGCAG
<table>
<thead>
<tr>
<th>SEQ ID NO:32 (INSPO17 Polypeptide sequence (original))</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LPEISIPSAG TLPHFIEEPD DAYIKNSPFI ALRCKARPAM QIFFKNGEW</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>VHEONEMSEV TLEDSSGLKV REVFINVTRO QVEDFHGPED YWQCCVANSW</td>
</tr>
<tr>
<td>101</td>
<td>LGTSRKSRAAS VRIAYLRKNF EQDPQOREVP IEGMIVLHCR PFPQVPFAEV</td>
</tr>
<tr>
<td>15</td>
<td>EMLKNEEPID SEQDENIITR ADNHLRIROQ RLSDSNGYTC MAANIVKRR</td>
</tr>
<tr>
<td>201</td>
<td>SLSATVQVYV NWWGSSWSWEN SACNVRCCRG WQKRSRTCTN PAPLNGAFC</td>
</tr>
<tr>
<td>251</td>
<td>EGMSVQKRTC TLSVPCDGDSW EWVSEMSVCS PCECEHLIRE CTAPPFRNRG</td>
</tr>
<tr>
<td>301</td>
<td>KFCEGLSQUEST ENCTDGSLCLI GIANSDIAL YSGLGAAVVA VAVLVIGVTL</td>
</tr>
<tr>
<td>351</td>
<td>YRFSQSDYGV DVIDSSALTG GFTQTNFKTV QRGNSLILLS AMQFPLTVSR</td>
</tr>
<tr>
<td>401</td>
<td>TYSQPIICLQD PLDKELMTES LSLFNPLSDJK VKVQSSFMVS LGVESRAEHY</td>
</tr>
<tr>
<td>45</td>
<td>GKNHSRTFPH MNNHSSTSHM PRNMRPYIQN LSSLPRTTEL RRTGVGHLG</td>
</tr>
<tr>
<td>501</td>
<td>GRLVMNPTGV SSLIPPGAP EINESWEYMS INQEGPSLQDS GSEVLLSPE</td>
</tr>
<tr>
<td>551</td>
<td>VTCGPDPMITT TTPFAITTIHP CADVSSEHWN HILKVRQTOQ KKEEVMVSED</td>
</tr>
<tr>
<td>601</td>
<td>ESTSCYLCLI DFFCHVLLDS FGTAYLIGEP ITDCAVKQLK VAVFGMCSDN</td>
</tr>
<tr>
<td>651</td>
<td>SLQYNLVRYC VDTCPEAQCF QSVESEHQQQ QLSEEPKLLL EKGNFTSIQI</td>
</tr>
<tr>
<td>701</td>
<td>SVLQIPPFPLW RIKFTACQQ PVPFVRWSCN ROPHLCAFSL ERYTPPTTQL</td>
</tr>
<tr>
<td>751</td>
<td>SKICIQKLQK GHEQILQVQT SILESERETI TFFAQEDSTF PAQGTFPAFK</td>
</tr>
<tr>
<td>801</td>
<td>IPYSIRQIRIC ATFDTPNAGK KDWQMLAQFRN SINRNLSYFA TQSSPSAVIL</td>
</tr>
<tr>
<td>851</td>
<td>NLWEARHQHD GDSLALCAL EEIGRHTKL SNISESEQ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEQ ID NO:33 (INSPO17 – full length nucleotide sequence)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATGATCTTTTA AAGATGCGCT TTCCAGGGTA CAAGATGCGCA TTCCCTCTTG</td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>AGCTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>101</td>
<td>AGCTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>251</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>301</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>351</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>401</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>45</td>
<td></td>
</tr>
<tr>
<td>501</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
<tr>
<td>551</td>
<td>TCTTCTCTCT CTCCAGGGCT CAGCTACTTGA AGAGATCTCTG</td>
</tr>
</tbody>
</table>
51 TCCTCTCCAC TGTGCTCAGC AATCACACCC TCCTGCCTTT GATTTGTG
5  CAGTTTTCCTG GGAATTAAGT CTTTGTCTAC TGTATGAGA AAGAGGAACT
10 GACAAATGGGC AAGCCTTCCC CGAAATCCATC CCATCAGCTC CTGGGAGACT
15 GCCCTCATTT ATAGAGGAGC CAGATGATGC TTATATATAC AAGACCAACC
20 CAGGTTGCAGCT CAGGTTGACAA GCCAGGCCAG CATGCGAGAT ATTCCTC
30 TGCAACGGCG AGTGCTTCCA TGGAACCGAG CAGTCTCAGT AAGACGCTCT
35 GACGGAGAGG TCACGGTCTGA AGTGGTTCGAA AATGGTCTATC AATTTTAC
40 GACAGCAGGT GGAGAGTGCT CATGGCGCGG AGGCCATATT GTGGCCAGT
45 GTGACTGGTA GCCACCTGGG TACCTTCAAG AGCAGGAAGG CCGCGCTGG
50 CATAGCTCAT TTTACGGAAAA ACTTTGAACA AGACCACAAAG CGAAGGGAG
55 TTGCTATGTA AGGCATGATT GTACTGCCTC GCCGCCCAAC AAGAGGACG
60 CAGAAGGAGG AGGAGGCTCA CAAGGTGGGA AAAGAAAAA GTGACAGAAC
70 AGCCAGCTCG CTCGGACTCA GGAATTACCA CTTCTGGAGG AAGCAACATC
80 GTGCGCAAGA GGAGAAGCCT CTGGCCGACT GTTGGAGGCT ACGTGGAAT
85 AGACCTGGCTC TGCTGGACAG AGTGCTATGC CAGCTATGCT CGCTGDGTGA
90 GAGGATGGCA GAAAGGCTCC CGACACTGCA CCAAACACGC TCTCTCAAT
95 GTGGGGGACT TTTGGAGGAAG ATATCGAATGG CAGAAAAATAA CCTTGCTCTC
100 TCTTTGTCTT GTGGATGCGG GCTGGGGAGTG GTGGACGAA TGGTCGGTCT
105 GCAGTCCAGA GTGTGAACAT TTGCGGATCC GGGAGTGCAC AGCCACCCAG
110 CCGAAAGATG GGGCGAATT CTGTGAAGGT CTAAGCAGAG AATCTGAAAA
115 CGTCACAGAT GTTCTTGGCA TCTTACCAGT TGAATGCG CAGCAGCATG
120 CTGGCTACTC GGCGTTGGGT GCTGCCGTGC TGGCGTGGTC AGTCTCGTG
125 ATGGGTGTCA CCCCCTACAG AGGGAGCGAG AGTGACTION GCGTGGACGT
130 CATGGACTCA CGAGATCCAG TGCTCTAGCTA ATTCTGCGAC GCAAGCCAG
135 CGTACAGTGA GCCGGACATA CAGCGGAGCC ATCGTCTGAG AAGACCGCTCT
140 GGACAAGGAG CTCATGACAG AGTCCCTACT CTCTAACCCC TTGCGGACAG
TCAAGTGAAA AGTCCAGAGC TCGTTCTAGG TTTCCCCTGGG AGTGTCTGAG
AGAGCTGAGT ACCACGGCAA GAATCATCCC AGGACTTTTC CCCATGGGAAA
CAACCACAGC TTTAGTACAA TGCACTCCAG AAATAAATAG CCCTACATCC
AAAATCTGTC ATCACTCCCC ACAAGGACG AACTGAGGAC AACTGCTGTC
TTTGCGGCTAT TAGGGGGCCG CTGATGAAAAG CCAAATACAG GGGTGACTTT
ACTCATAACCA CACGGTCCCA TCCCCAGAGGA GAATTCTTGGG GAGATTTTA
TGAAGGATAA AGGACACAGC AGGGCAAATG GGAGGAATTG
GGAATATCCA TTTAAGAAGG AGGACACAGC AGGGCAAATG GGAGGAATTG
ATGTCACTGGT AAGATGAATC TACATCCTGT TACCTGCTTT TGGACCCTTT
TGCGTGCATG GTGCTCCTGG ACAGCTTGGG GACCTATGGG CTCACCTGAG
AGCCAAATAC ACAGTGTGCC GTGAGCAAC TGAAGGAGGC GGTGTTTGGC
TGCAATGCCT GTAACTCCCT GGATTACAC TTTGAGATTT ACTGCTGTC
CATTACCCTT TGGCATTTTG ACAGAATGGT TTTAGATGAA AGGCCATCA
GTGGACAGCT CCTGGAAGAA CAAAATGGC TGCATTTCAA AGGGAAATACC
TTTATGCTTC AGATAATCTG CCTTGATATT CCCCCATTC CTCGAGAAT
TAAACCCATT ACCTGCTGCCC AGGAAGTCCC GTTCTC CCCG GTGTGTGCA
TGAACC GGCA GCCCTGCAC TGCTGCTTTTC CCGGGAGGCT TATAGGCC
AC TACCCACC AGCTGCTCTG CAAAATCTGC ATCCGCGAGC TCAAAGGCA
TGAAAGATCG CTCCAAGTGC AGACATCAAT CTAAGAGAG GACGCAGAAA
CCATCATTTT CTGGACACAA GAGGACAGCA CTTCCCTGCC AGAAGCTGG
TCAAAGGCT TCATAATTCT AACTCCAT TCAGAGCGGA TTTTGCTTAC
ATTGATACC CCAATGCCA AAGGAAGAGA CTGGACAGAT TTAGAACAGA
AAAACAGCAT CAACAGGAT ATATCTTATT TCCGCTACAC AAGTACCGCA
TCTGCTGTCAT TTTGGAACCT GTGGGAAGCT CGTCATCAAG ATGATGGTGA
CTTGACTGTC GTGGCCTGTG CTTTGAAGAG GATTGGAGG ACACACAGA
AATCTCTCAA CATTTCAGAA TCCCCAGCTTG ATGAAGCGGA CTTCAACTAC
AGCAGGCAA ATGGACTC
SEQ ID NO:34 (INSP017 full length polypeptide sequence)

1  MIFKNGLSRV QDATFPCPLH CAQQSHPPAF DFVAVSWDLS LCSLYEETGT

5  DGNEALPESI PSAPGTLPHF IEEPDDAYII KSNFIALRCK ARPAMQIFFK

10  CNGEWVHQNE HVSEETLDES SGLKVREVFI NVTRQQVEDF HGPEDYWCQC

15  VAWSHLGTSK SRKASVRIAY LRKNFEQDPQ GREVPFEGMI VLHCRPPEGV

20  PAAEVEWLKN EEPIDSEQE NIDTRADHNLI IIRQARLSDS GNYTCAANIM

25  VAKRRSLSAT VVYYVNGGWS SWTWSACNV RCGRGWQKRS RTCTNPAPLN

30  GGAFCCEGMSV QKITCTSLCP VDGSEWVWSE WSVCSECEHH LRIRECTAPP

35  PRNGGKFCEG LSQESENCTD GLCILGIENA SDIALYSGLG AAVVAVAVLV

40  IGVTLYRRSQ SDYGVVDVIDS SALTGGFQTFT NFKTVRQGNS LLLNSAMQPD

45  LTVSRTYSGP ICLQDPDLKE LMTESSLFNPL SDIKKVQGQ SFMVSLGVSE

50  RAHYHGKNHS RTFPHGNHHS FSTMHPRNKM PYIQNLSLPL TRTELRTTG

55  FGHLGGRLVM PNTGVSSLIP HGAIEENSW EIYMSINQGE PSLQSDGSEV

60  LLSPETVTCGP PDMIVTTFPF LTIPHCADVS SEHCWNIHLKK RTQQGKWEEV

65  MSVEDESTSC YCLLDPFACH VLLDSFGTVA LTGEPIEDCA VKQLKVAVFG

70  CMSCNLSDYN LRVYCVDNTP CAFQEVVSDE RHQGGQLLEE PKLHFKGNT

75  FSLQISVLDI PPFLWRIKPF TACQEVFFSR VWCSNRQPLH CAFSLERYTP

80  TTQLSCKIC IRQLKGHEQI LQVQTSILES ERETITFFAQ EDSTFPFAQTG

85  PKAFLKIPYS RQRICATFDT PNAKKGKWQM LAQKNSINRN LSYFATQSSP
901 SAVILNLWEA RHQHDGDLDS LACALEEIGR THTKLSNISE SQLDEADFNY

951 SRQNGL
CLAIMS

1. A polypeptide, which polypeptide
   (i) comprises the amino acid sequence as recited in SEQ ID NO:34 or SEQ ID NO:32
   (and preferably consists of the amino acid sequence as recited in SEQ ID NO:34);
   (ii) is a fragment thereof having transmembrane protein function, particularly netrin
   receptor activity or having an antigenic determinant in common with the
   polypeptide of (i); or
   (iii) is a functional equivalent of (i) or (ii).

2. A polypeptide according to claim 1, which functions as a trans-membrane protein
   molecule and, moreover, as a trans-membrane protein molecule of the netrin receptor
   family.

3. A polypeptide according to claim 2, which consists of the amino acid sequence as
   recited in SEQ ID NO:32 or in SEQ ID NO:34.

4. A polypeptide which is a functional equivalent according to claim 1(iii), which is
   homologous to the amino acid sequence as recited in SEQ ID NO:34.

5. A functional equivalent according to any one of the preceding claims, which has
   greater than 98%, 98.5%, 99% or 99.5% sequence identity with SEQ ID NO:34.

6. A functional equivalent according to any one of the preceding claims, which has
   greater than 98.5% sequence identity with SEQ ID NO:34.

7. A fragment as recited in claim 1 having an antigenic determinant in common with a
   polypeptide of claim 1(i), which consists of 7 or more (for example, 8, 10, 12, 14, 16,
   18, 20 or more) amino acid residues from the sequence of SEQ ID NO:32 or SEQ ID
   NO:34.

8. A purified nucleic acid molecule which encodes a polypeptide according to any one
   of the preceding claims.

9. A purified nucleic acid molecule according to claim 2, which has the nucleic acid
   sequence as recited in SEQ ID NO:31 or SEQ ID NO:33 is a redundant equivalent or
   fragment thereof.
10. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to claim 8 or claim 9.

11. A vector comprising a nucleic acid molecule as recited in any one of claims 8-10.

12. A host cell transformed with a vector according to claim 11.

13. A ligand which binds specifically to, and which preferably stimulates the activity of, a polypeptide according to any one of claims 1-7.

14. A ligand according to claim 13, which is an antibody.

15. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1-7.

16. A compound according to claim 15 that binds to a polypeptide according to any one of claims 1-7 without inducing any of the biological effects of the polypeptide.

17. A compound according to claim 16, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.

18. A polypeptide according to any one of claim 1-7, a nucleic acid molecule according to any one of claims 8-10, a vector according to claim 11, a host cell according to claim 12, a ligand according to claim 13 or 14, or a compound according to any one of claims 15-17, for use in therapy or diagnosis of disease.

19. A method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claim 1-7, or assessing the activity of a polypeptide according to any one of claim 1-7, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.

20. A method according to claim 19 that is carried out in vitro.

21. A method according to claim 19 or claim 20, which comprises the steps of: (a) contacting a ligand according to claim 13 or claim 14 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

22. A method according to claim 19 or claim 20, comprising the steps of:
a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 8-10 and the probe;

b) contacting a control sample with said probe under the same conditions used in step a);

and

c) detecting the presence of hybrid complexes in said samples; wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

23. A method according to claim 19 or claim 20, comprising:

a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 8-10 and the primer;

b) contacting a control sample with said primer under the same conditions used in step a);

and

c) amplifying the sampled nucleic acid; and

d) detecting the level of amplified nucleic acid from both patient and control samples; wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

24. A method according to claim 19 or claim 20 comprising:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to any one of claims 8-10 from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation which is associated with disease in the nucleic acid molecule as an indication of the disease.

25. The method of claim 24, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.
26. The method of either claim 24 or 25, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

27. A method according to any one of claims 19-26, wherein said disease is selected from cell proliferative disorders particularly cell proliferative disorders particularly brain tumours, nervous system tumours, neoplasm, bone tumor and myeloproliferative disorders particularly myeloid leukaemia, autoimmune/inflammatory disorders, cardiovascular disorders, neurological disorders, such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, psychiatric disorders, such as depression, schizophrenia, brain injury, spinal cord injury, nerve injury, developmental disorders, including disorders of nervous system development, nervous system inflammation including motor neuron disease, amyotrophic lateral sclerosis, multiple sclerosis and inflammatory neuropathy, bone disease, atherosclerosis, glomerulonephritis, cachexia, AIDS, HIV infection, metabolic disorders such as diabetes, infections, reproductive disorders, infertility, embryo implantation failure, pregnancy disorders and birth complication and other pathological conditions, particularly those in which netrin receptors are implicated.

28. Use of a polypeptide according to any one of claims 1-7 as a transmembrane protein, preferably as a netrin receptor.

29. A pharmaceutical composition comprising a polypeptide according to any one of claim 1-7, a nucleic acid molecule according to any one of claims 8-10, a vector according to claim 11, a host cell according to claim 12, a ligand according to claim 13 or 14, or a compound according to any one of claims 15-17.

30. A vaccine composition comprising a polypeptide according to any one of claims 1-7 or a nucleic acid molecule according to any one of claims 8-10.
31. A polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-10, a vector according to claim 11, a host cell according to claim 12, a ligand according to claim 13 or 14, a compound according to any one of claims 15-17, or a pharmaceutical composition according to claim 29, for use in the manufacture of a medicament for the treatment of cell proliferative disorders particularly brain tumours, nervous system tumours, neoplasm, bone tumor and myeloproliferative disorders particularly myeloid leukaemia, autoimmune/inflammatory disorders, cardiovascular disorders, neurological disorders, such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, psychiatric disorders, such as depression, schizophrenia, brain injury, spinal cord injury, nerve injury, developmental disorders, including disorders of nervous system development, nervous system inflammation including motor neuron disease, amyotrophic lateral sclerosis, multiple sclerosis and inflammatory neuropathy, bone disease, atherosclerosis, glomerulonephritis, cachexia, AIDS, HIV infection, metabolic disorders such as diabetes, infections, reproductive disorders, infertility, embryo implantation failure, pregnancy disorders and birth complication and other pathological conditions, particularly those in which netrin receptors are implicated.

32. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-7, a nucleic acid molecule according to any one of claims 8-10, a vector according to claim 11, a host cell according to claim 12, a ligand according to claim 13 or 14, a compound according to any one of claims 15-17, or a pharmaceutical composition according to claim 29.

33. A method according to claim 32, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

34. A method according to claim 32, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide,
nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.

35. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-7, or the level of expression of a nucleic acid molecule according to any one of claims 8-10 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.

36. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-7, or a nucleic acid molecule according to any one of claims 8-10 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.

37. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 8-10; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

38. The kit of claim 37, further comprising a third container holding an agent for digesting unhybridised RNA.

39. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 8-10.

40. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-7; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

41. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-7.

42. A method for screening for a compound effective to treat disease, by contacting a
non-human transgenic animal according to claim 41 with a candidate compound and determining the effect of the compound on the disease of the animal.
Figure 1

5  1 LPESIPSAPG TLPFIEEED DAYIKNPI ALRCKARPAM QIFFKCNGEW
51  VHQNEHVSEE TLDSSGLKV REVFINVRTQ QVDEFHGPED YWCQCVAWSH
101  LGTSKSRKAS VRAYLHRNF EQDPQGREV PEGMVHLHC PPEGVPAAEV
151  EWLKNEEPID SEQEDIDTR ADHNLIIRQA RLSDGSNYTC MAANIVAKRQ
201  SLSATVVVYV NGGSSSSTW DANCNVRCGRG WQKRSRTCTN PAPLNGGAFC
10  251  EGMSVQKITC TLSCPVDGWS EVWSEWSVCS PECEHLRIE CTAPPPRNGG
301  KFCGKLQES ENCDTGLCIL GIENASDIAL YSGLGAAVVA VAVLIGVTL
351  YRQSQSDYGV DVIDDSSLALG GFQTFTNFKTV RQGNSLLLMS AMQPDWTVSR
401  TYSGPICALQD PLDKELMTES SLFNPLSDIK VKQVSFMVS LGVESERAYH
451  GKNHSRTFPH GNHNHSIFSTMH PRNKMPYIQN LSSLPTRTEL RTTGVFGLHLG
15  501  GRLVMPTGTV SLIPHGAIIP EENSWEIYMS INQGEPSLQS DGSEWLLSPE
551  VTCGPPDMIV TTPFALTIPH CADVSSEHWN IHLKKRTQQG KWEVMSVED
601  ESTSCYCLLD PFACHVLLDS FGTYALTGEF ITDCAVKQLK VAVFGMCSVN
651  SLDYNLRVVC VDNTPCAFQE VVSDERHQGG QLEEPKLLH FKGNTFSLQI
701  SVLDIPPFWL RIKPFTACQPE VFPSRWVCSN RQPLHCASFL ERYTPTTTQL
20  751  SCKCICRLK GHEQILQVQT SLIESERETI TFFAQEDSTF PAQTPKAFK
801  IPYSIRQIRC ATFDTPNAGK KDWQMLAQKN SINRNLSYFA TQSSPSAVIL
851  NLWEARHQHD GDLDSLACAL EEIGRTHTKL SNISESQ
Figure 2

Query = INS017

888 letters

Database: ncbi-nr

808,980 sequences; 257,180,598 total letters

Searching....................................................done

Score  E

Sequences producing significant alignments: (bits)

Value

15

| gi|1559982|ref[NP_071543.1] (NM_022207) transmembrane receptor | 970 0.0 |
| gi|12857776|dbj|BAB31108.1| (AKO18177) putative [Mus musculus] | 966 0.0 |
| gi|6678505|ref[NP_033498.1] (NM_009472) UNC-5 homolog (C. elegan... | 933 0.0 |
| gi|16933525|ref[NP_003719.2] (NM_003728) unc5 (C.elegans homolog... | 929 0.0 |
| gi|16159681|ref[XP_042940.3] (XM_042940) unc5 (C.elegans homolog... | 929 0.0 |
| gi|3789765|gb|AAC67491.1| (AF055634) transmembrane receptor UNC5... | 929 0.0 |
| gi|11559980|ref[NP_071542.1] (NM_022206) transmembrane receptor | 835 0.0 |
| gi|14424612|gb|AAH09333.1|AAH09333 (BC009333) Similar to transme... | 395 e-108 |

25

| gi|14747123|ref[XP_038314.1] (XM_038314) hypothetical protein XP... | 363 4e-99 |
| gi|10434496|dbj|BAB14276.1| (AK022859) unnamed protein product [... | 363 4e-99 |
Figure 3

>gi|11559902|ref|NP_071543.1| (NM_022207) transmembrane receptor
Unc5H2 [Rattus norvegicus]
gb|AAB57679.1| (U87306) transmembrane receptor UNC5H2 [Rattus norvegicus]
Length = 945

Score = 970 bits (2507), Expect = 0.0
Identities = 478/899 (53%), Positives = 636/899 (70%), Gaps = 23/899 (2%)

Query: 1 LPESIFPSAPGT-LPFIPEPDDAYI1KSNPIALRCKARPAMQIFFKCNQGQHVEHVS 59
LP+S PSAP LPFH+ EP+DAY+K+ P+ L C+A PA QI+FQCNQGQ+ HV++
Sbjct: 35 LPDSFFSAPAEQLPHFLEPEDAYIKNKPVKVELHKRAFPATQIYYFKCNQGQ9 94

Query: 60 ETLDRESSGLKREVFINTQVQVEDFHGPEDYWCQCVAWLHTSKSKRASVRIAYLRK 119
E+LDE+ +GL+ +REV I +VQQVE+ G EDYQNWVAVS GT+KSR+ +RIAYLRK
Sbjct: 95 ESLDEATGLRLREVEQIVETQVQVEELFQLEDYWCQCVAWSSGTSDKRSPAYLRLKR 154

Query: 120 FEQDPQGRFEPIEMGIYLVKHCREFPQEGVPAAEWELKNKHEIPIDSEQDENIDTRADHDNLIRIQ 179
F+S+P+E+EP++ ++L CRPFGQVF AEWEVLKNE+ ID QD N DHNLIRIQ
Sbjct: 155 FDQEPPLAKEVPDHEVLLQCRPFGQVFAEWEVLKNNEDVIDFAQDTNFL1TDHDNLIRIQ 214

Query: 180 ARLSDNSGNYTCMAANIVAKRSSLSATVYVNYNGSSWTEWACNVRCGGRGWQKSRCTC 239
ARLS+ NYTC+A NIVAKRSS +ATV+VYNNGSSW EWS C+ RCGGRGWQKR+RTCT
Sbjct: 215 ARLSDTANYTCSVAKKNIVAKRSSSTTATIVYVNGSSWAEWSPCSNRCGGRGWQKRTCT 274

Query: 240 NAPLNGGACFEGMSQKVTICCSSLCFCVDSWVSEWSWSVCPECHELRERCTAPPFNG 299
NAPLNGGAFEG + QK CT++CEFVG+W WS+WCS EC H R REC APPF+NG
Sbjct: 275 NAPLNGGACFEGQACQTKATCTTVCVDSWTSWCSACEHHWSRECCEAPPPQNG 334

Query: 300 GKFCEGLSGQQSENCTDGLCILG-----------------IENASDIALYSGLGAAV-VAALVL-V 346
G+ C G +S+NCTDGLC+L +E + D+ALY+GL AV V +AVL+ +
Sbjct: 335 GRDCSTLLDSKCNCTDGLCVLNQRLNDPKSRPLFPGDVAQLYAGLVLQAVFVVLAVLMAV 394

Query: 347 GVTLYRSGQSYGVDVIDSS-ALTGGFQTPFKTRQGRQNLSSLNAMQDDTVSR-TYSG 404
GV +YR+ D+ D+ DSS ALTGGF NFKT R N LL+ + PDFT S Y G
Sbjct: 395 GIVIVYRRNCRDFTTDITDSSSALTGGFHPYNFTKARPSNPLLHPAPPDPDTSASYYRG 454
Figure 3 Continued

Query: 405 PI-CLQDPLDKELMTESLFLNLSDKVQSSFMV--SLGVSEARAYHGKNNSTDTPHG 461
P+ LQD DK MT S L +PL +K+V S + G+++ A+ G T+P
Sbjct: 455 PVYQALQDSADKIMNTSLPDPLSLQKIVDSTSTIGSAGLADGALDGVLPGYTPGD 514

Query: 462 NNHSFSTMPHRKMPFYIQLNLSSPLTRTELRTGFGHLGGRLVMTGVSLLIPHAIF 521
+ +H R+ Q+L LP +G FG LGGRL +P TGVSSL+P+GAIP+
Sbjct: 515 FSRDTHFHLRSASLGSMHLLGLPRDPSSSVSTYGFGCLGGRLTIPGTVSSLVPNGAIP 574

Query: 522 ENSWEIYMSINQGEFSLQ-SGDSEVLLSPEVTGPDMMIVTTPFALTIPHCADSSEHWN 580
+++Y+ IN+ E +L S+GS+ +LSP VTCGP +++ P LT+PHCA+v +W
Sbjct: 575 GKYDLYLRLNKTESTLPLSEGSQVTLISPSVCTGPTGLLLDRPVPVTLPHCAEVIAGDIW 634

Query: 581 IHLKRTQGKWEVEWMSVEDE--STSCYCLDDPFACHVLLDSFGYTAYLTGEPIITDCAVKQ 638
LK + QG WEEV+++E +T CYC L+ +CH+LLD GYR TGE + AVK+
Sbjct: 635 FQKLTQAHGQHWEVVEVTLDEETLNTFCYCLEAKSCILLDQLGTYVFTGESYRSAXVR 694

Query: 639 LKVAVFCGMCSCNLSLNYRLVYCFNTPCAOFEEVSDEERHGQQLEFLKLHFGNTFSL 698
L++A+F C SL+Y+LRVYC+++TF A +EV+ ER GG L+EEFL L F + +L
Sbjct: 695 LQALIFAPALCSTLYSLRVCLEDTPAAKKEVLELEKLTLGGYTVEEEKTLLTFKYH 754

Query: 699 QISVLDIPFPFLWRKIKPFATCQEVFSPSRVWSRNRPQHCAFSLERTPTTQLSCKICIRQ 758
+++S+ DIP WR K QE+PF VW +++ LHC F+LER+ +T+ C+C+RQ
Sbjct: 755 RLSLDIPAHVRSKLLAKYQIEFHYHVWNSQKALHCTFTLERHSLASTEFCKCVRQ 814

Query: 759 LKGHEQLQVQTISELERETI-TFFAQEDSTFFAQPFTGPAKFKFYSIRQRICATFTDN 817
+++G QI Q+ T++ E+ +++ + + G QP AFKIP SIRQ+IC +D PN
Sbjct: 815 VEGQCIQFQLHTTLATEAPGSLDLALCSAPGNAAATTQLGYPYAFKIPSLIRQKICNSLAPN 874

Query: 818 AKGKDQMLAQKMSNIRNLSYFTQGSSPSAVLNLWEAHRQDGDDLSLACALEIGRT 876
+++ G DW+++LAQK S++R L+YFAT++SP+ VIL+LWEAR Q DGD+SLA ALE+G+G
Sbjct: 875 SRGNDCWRLAQKLSMDOYNYFTAKSPGTGVDLDLWEARQDDGDCLNSLASEMEGK 933
Figure 4

```
1  CTTCGGGAAT CCATCCCCAT GCAGCTCCTGG GCACAGGCTTC ATTTCATAGA GGAAGCAGAT
   lpe s i p s a p g t l p h i e e p d
5  GATGCTTATA TTAGCCAGAC CACCCCCATT GCACGCTCAGT GGAAAGGCGG GCCAGCCATG
   d a y i i k s n p i a r c k a r p a m
10  CAGATATTTTC TCAAAGTGCA CGAGGAGGTG CTCCAGCAGA AGCAGCACTG CTCGTAAGAG
   q i f f k c n g e w v h q n e h v s e e
15  ACTCTGGAGC AGAACGTCAGT TTTGAAGTGC CGAGAAAGTG TCATCAATGT TACCTAGGCA
   t l d e s s g l k v r e v f i n v t r q
20  CAGGTGAGGG ACTTCCATGG GGGCGAGGAC TATTGGTGCC AGTGTGGAGC GTGGAGGCAAC
   q v e d f h g p e d y w c q c v a w h
25  CTGGGATCT CCGAAGACG GAAGGTTCTCT GTGCGCAAGAG GGGGGGGGGGGGGGGGGG
   C P 1
30  GAAACAGGCC CAAAAGAAGA GGAAGTTGCC ATGGAAGGGCA TGATTGGTACT GCACAGCCGC
   e q d p q g r g e v p i e g m i v l h c r
35  CCACAGGAG GATGCTCTGC TGCCAGGATG GAATGGCTGA AAAATGAAAG GCCCATTGAC
   p p e g v p a a e e v w l k n e e p i d
40  TCCTCAACAG AGCGAGAATC TGACACTCAG GCTGACACCATA ACCTGATCAG CAGGGAGGGA
   s e q d e n i d t r a d h n l i r q a
45  CGGCCGCTCTGG ACTCACGAAA TTACACTCTC ATGGCAAGCCA ACATCGTGCC TAAGAGGAGA
   n y t c m a a n i v a k r r
50  GAGGCGGCG CCATCTTGAG GGTCTAGTAG AGATGGAGCTT GGCTCTCTTG GACAGAGTTG
   s l s a t v v v v y n g g w s s w t e w
55  TCAGCGCTGCA ATGTGCGAGTC AGCAGTTGAAAC GTTCGCCGAG CTGTCACCCAAC
   s a c n v r c g r g w q k r s r t c t n
60  CCAGCTCCTC TCAATGGGTG CGCGCTTGTGT AGGGAAATGT CAGTGACAGA AATAACCTGC
   p a p l n g g a f c e g m s v q k i t c
65  ACTTCTCTTT GTACCTGGGA TGAGGAGCTG GAAATGGGCA TGGACACGAC CACCCCCGGG
   A A T G G G G G G C e p c e h l r i e c t a p p p r n g g
70  AAAGCTCTTG AGGTCTCAGG CAGGAAATCT GAAGACTCAG TTACGATCTA TGCAGCTACT
   k f c e g l s q e s e n c t d g l c i l
75  GCCATGGAGA AGTGCGCGCA CATTGGTTTG TACTCAGGTGT TGCTGTCGTC CTGCCTGGGC
   g i e n a s d i a l y s g l g a a v v a
80  GTGCGAGCTC TGCTTATTGG TCAGCTCTCT TGACAGGAGA AGCGAGTGTA CTAATGGGTTG
   v a v l v i g v t l y r r s q s d y g v
```
2221  GACGTTATA CGCCCACTAC CACCCAGCTG TCCTGCAAAA TCTGCATTCG GCAGCTCAA
eryetpttttqtl slskkicircir klk
5
2281  GGGCATGAAC AGATCTCCCA AGTGCAAGCA TCAATCTTAG AGAATGAACG AGAAAACCATC
gh eqilqvqtsltsilesereti
10
2341  ACTTCTCTCG CACAAAGAGGA CACGACTTTTC CCTGCACAGA CTGGCCCAAA AGGCCTTCAA
  tffaqedstfpaqtgpkafk
2401  ATCCCTACT CCACTGACCA GGGATTTGG GCTACATTTG ATACCCCCAA TGCCAAAGGC
  ipysirqricatfdtpnakg
15
2461  AAGAAGTGGG AGATGTATTG ACAGAAAAAC AGCATCAACA GGAATTTATG TTATTTGTCT
  kdwqmlaqknensinrnltsyfa
2521  ACACAAAGTA GCCCATCTGC TGATATTGTG AACCTGGTGG AAGCTGTCGA TCAGCATGAT
20
2581  GGTGATCTTTG ACTCCCTGGC CTGTCGCCCTTG GAAAGATTTG GAGAGACACA CACGAAAATCTC
gdl dslacale eieigt htkl
2641  TCAAAACATT CAGAATCCCA G

30  Position of cloning primers

35  Position and sense of GeneRacer primers
Figure 6

Molecule: pCR4 blunt TOPO-INSPO17, 6887 bps DNA Circular
File Name: V11C8BEC.cm5,
Description: plasmid ID 13287

Molecule Features:

<table>
<thead>
<tr>
<th>Type</th>
<th>Start</th>
<th>End</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGION</td>
<td>2</td>
<td>216</td>
<td>lac</td>
<td>promoter region</td>
</tr>
<tr>
<td>REGION</td>
<td>205</td>
<td>221</td>
<td>M13 reverse priming site</td>
<td></td>
</tr>
<tr>
<td>REGION</td>
<td>217</td>
<td>294</td>
<td>T3</td>
<td>Lac2a-ccdB gene fusion</td>
</tr>
<tr>
<td>MARKER</td>
<td>243</td>
<td></td>
<td></td>
<td>Poly linker</td>
</tr>
<tr>
<td>REGION</td>
<td>262</td>
<td>294</td>
<td></td>
<td>TOPO cloning site</td>
</tr>
<tr>
<td>GENE</td>
<td>3199</td>
<td>331</td>
<td>C INSP017 ORF</td>
<td>Cloned PCR product</td>
</tr>
<tr>
<td>REGION</td>
<td>3224</td>
<td>295</td>
<td>C</td>
<td>-20M13 forward priming site</td>
</tr>
<tr>
<td>REGION</td>
<td>3225</td>
<td>3740</td>
<td></td>
<td>Lac2a-ccdB gene fusion</td>
</tr>
<tr>
<td>REGION</td>
<td>3225</td>
<td>3242</td>
<td></td>
<td>Poly linker</td>
</tr>
<tr>
<td>MARKER</td>
<td>3225</td>
<td>3225</td>
<td></td>
<td>TOPO cloning site</td>
</tr>
<tr>
<td>GENE</td>
<td>4089</td>
<td>4883</td>
<td>KanR</td>
<td>Ribosome binding site</td>
</tr>
<tr>
<td>REGION</td>
<td>5075</td>
<td>5079</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENE</td>
<td>5087</td>
<td>5947</td>
<td>AmpR</td>
<td></td>
</tr>
<tr>
<td>REGION</td>
<td>6092</td>
<td>6765</td>
<td>pUC ori</td>
<td></td>
</tr>
</tbody>
</table>
pCR4-TOPO INSPI017

1 AGCGCCCAAT AGCACAACCG CCTCTCCCCG CCGTCTGGCC GATTCATTTA TGGGACTGGG
11 ACGAGAAGTT TCCGACTGGA AAAGCGCCAGT GTCGAGGCA CGCGAATTAAC GTCGAGTTAGC
21 TCAGCTATCA GCAGCCGGCA GTCTTTAGCT TAAAGCTTACG GGTCTGATG CCGCTGAGGA
31 GGATCTAAGG CAGTATTGACG AGCTTGGAGT TATGCCGTTT AATTTGAATT TGGGAA
41 GCCTTTTCCG ATTCCGTGATG GTTAGGCTGCCTG CTCCCAATTC
51 CTCTCAAGGG CCAGTCCGAGG AGCTCTATTTT GCGATATCTA TACGATTTCTG CTCTTCTTCA
61 CGAGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
71 TGATTCATATG ACTCCTCGGA GGGCTTACGGG CCTTTGAAGA CGAGATTTCTG CTCTTCTTCA
81 CCCCTGAAAGG GAGGTGTACTT GCCGATGAGG AGCTGATCAA GGCTGAAAGA ATTTTCCCTT
91 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
101 CAGGAGATCC TCCCTGCCTT TGGCGGAAGA AATTGAGA TGGCGGAAGA TGGCGGAAGA
111 TGATTCATATG ACTCCTCGGA GGGCTTACGGG CCTTTGAAGA CGAGATTTCTG CTCTTCTTCA
121 CCCCTGAAAGG GAGGTGTACTT GCCGATGAGG AGCTGATCAA GGCTGAAAGA ATTTTCCCTT
131 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
141 TGGACTGAG ACCTGAGAA AGCTTGGAGT TATGCCGTTT AATTTGAATT TGGGAA
151 CTCTCAAGGG CCAGTCCGAGG AGCTCTATTTT GCGATATCTA TACGATTTCTG CTCTTCTTCA
161 CGAGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
171 TGATTCATATG ACTCCTCGGA GGGCTTACGGG CCTTTGAAGA CGAGATTTCTG CTCTTCTTCA
181 CCCCTGAAAGG GAGGTGTACTT GCCGATGAGG AGCTGATCAA GGCTGAAAGA ATTTTCCCTT
191 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
201 TGGACTGAG ACCTGAGAA AGCTTGGAGT TATGCCGTTT AATTTGAATT TGGGAA
211 CTCTCAAGGG CCAGTCCGAGG AGCTCTATTTT GCGATATCTA TACGATTTCTG CTCTTCTTCA
221 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
231 TGATTCATATG ACTCCTCGGA GGGCTTACGGG CCTTTGAAGA CGAGATTTCTG CTCTTCTTCA
241 TGGACTGAG ACCTGAGAA AGCTTGGAGT TATGCCGTTT AATTTGAATT TGGGAA
251 CTCTCAAGGG CCAGTCCGAGG AGCTCTATTTT GCGATATCTA TACGATTTCTG CTCTTCTTCA
261 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
271 TGATTCATATG ACTCCTCGGA GGGCTTACGGG CCTTTGAAGA CGAGATTTCTG CTCTTCTTCA
281 CCCCTGAAAGG GAGGTGTACTT GCCGATGAGG AGCTGATCAA GGCTGAAAGA ATTTTCCCTT
291 GACGAGTCTC CTGGTGGAGAT TAAGTCTGACG AATTTCCCTT TGGCGGAAGA AATTGAGA
301 TGGACTGAG ACCTGAGAA AGCTTGGAGT TATGCCGTTT AATTTGAATT TGGGAA

ACAATAACCC TGATAAATGC TTCAATAATA TTGAAAAGG AAGAGTATGA GTAATCCAACA
TTTCCTGTCG GCCCTAPFC CTTTTTTGCG GGCAATTTGC CTCTCTGTTTT TTGCTCACCC
AGAACACCTG CGAAAGAATG AAAGATGCTGA AGATCAGTGG GTGCGAGGAG TAGGGTACAT
CGAATCGGAT CTCACAAGCG CGAAGACTCT TGAGAGTTAT CCACGAGAAGA AAGCTTITCC
AATGATGGAG ACTTTTTAAG TTCTGCTATG TGCGCGGCGTA TTATCCCGTA TTAAGCGCCC
GCAAGGCAAA CTACCGACCC GCATAACTA TTCTCAGAAT GCTTTGGTCA AGYACTCCAC
AGTCACAGAA AACATCTTA CGAATAGCCG GACAGTAAAG GAATATAAGA GTGCTGGCCAT
ACAATGAGT CATAACACTG CGCGCAACTT ACTTCTGACA AGCAGTGGAG GACGGAAGGA
GCTTAACCGCT TTTTGGCCAA ACTATGGGGA TCATGTAACT CGCTTTGATC GTTTGGAGACC
GGGACCTGAT GAAGCGCTAC CAACACGGCA CGGGCTGACC AGCAGACGCTG TACCAATGCC
ACCAACGGTG CGGCAACCTAT TAACTGGGCA ACTACTTACT CTAGCTTCCC GCCAACAAAT
ATACTACTGG ATACAGGGCC ATAAAAGCTGC AGAAGCACCT TCTGCGCCGA CCGTCCGCG
TGGCTGTTTT ATTGCTGTAT AACCTGGAAG CGCTGCGGCT GGGTCTCGGC GATATCTTGCC
AGCAYCTGAGG GCCAGATGGTA AGCGCTCCCG TATCGTAGGT ATCTACACGA CGCGAGATCG
GGCAACTTATG GATGAACGGA ATACAGACAT GGCTGAGAAT GCTGCCCACG TGATTAAGCA
TTGGATACTTG CTACAGACAG TTTACTCTTA TATACCTTTAG ATGAGATTAA AACCTTCATT
TTAATTTAAA AAGATCTAGG TGAAGACTCT TTTTGGAAAT CATAGAAACA AAATCCCTTA
AGCAYTAGTTT TCCTTCCACCT GAGGTCGAGA CCCGGTAGAA AAGATCAAAG GATCTCTTTG
AGATCTCCTTT TTCTGCGCCG TATATCTGCTG CTGCAACACA AAAACCACG CGCTACACGC
GGGCTGGTTTG TTGGCCGGATG AAGACGTACC AACCTCTTTT CGGAAGTTAA CTGGCGTTGAG
CAGACCGGACAT ATACACAACTA CTGGCTCTCTT ATACAGCCCG TGGTGGAGCC ACCACTCCAA
GAAACTCTGA GCACAGCCCTA CATAACCTGGC TCTGGCTAACT CGTCTACAG TGGCGCTGCTC
CAGGGACGAT AAGCTGTGCTG TTACGGGTTG GAGCCTCAGA CAATAGGTCG CAGGAAGAGC
GCCGCCGGTG GCCTGAAAGG GGGGTCTCCTG CACACAGCCC AGCTTGGGAC GAAGGACTTA
CCACGAGCTG ATAGATACCAT AGCCTGTGCTG ATGGAAGAAC GCCACGCTTC CCCGAGGGAG
AAAGCCGCGAC AGTATCGGAG TAAAGCGCGG GGCTCGCGAAA GGAGAGCGCA CGAGGAGAGC
TCCAGGGGAA AACGCCTCTGT ATCCTTTATAG TCTGCTCTGG TTGGCCACCC TCTGACTTGA
GGCGTCGATTT TTCTCTGATCT GCTTACGGGCG CGCGGAGCTA TGCGAAACCG CACACAGCCG
GCCCTTTTAA CGGTCTCCTG GCCTTTTGCT CATACGTTCT TCTGCTGGTT
ATCCCTGTAT TCTGGGATAA AGCGTATAC CGCCCTTGAG TGAGCTGTATA CGCTCCTGGG
CAGCCGAACG ACCGAGCGCA CGGAGCTCAGT GAGCGAGGAA CGCGAAG
Figure 7.

> Predicted sequence size: 887 aa. (SEQ ID NO.32)

> HXQKQLVQEGQDFTFGPVPLQGKAEHTRIPREDQKNLRK

LPSLPSAGPTLPFIEEPDADALIKSNPIALRCKAPMQIFFKCGEWHQNEHVSE

TLDESSGLKREVFINVTRQVDFHGPEDYWCQCVAWHLGTSKSRKASVRIAYLKRNF

EQDQGQREVPIEGMVILHCRPFPQGVPAAEVEWLKNEPIDSQEDNITRADHNLNRQA

RLSDGNYTCAANIVAKRSLSATVYWNGWSTEWASONCRGQWQGRSRCTCN

PAPLNCAGFCGMSQVKTCLSLCPVDGSWEVSEWSVCSPECHLRIRECTAPPFRNNG

KFCEGQSQESNCTDGLCILGIENASDIALYSGLGAAVVAVAVLVIGYTVLSRQSDYG

VIVDSSALTTFQGQTFNFKTVRQGNLNLHSAAMQDPLTDSRTYSGPICLQDPLKELMTE

SLFPLSLDIKVKVQSSFMVSGVERAEYHGKHNSRTFPQGNHNSFSTHMFRNKMFIQN

LSLPLRTELEETGFFHGLGRLVMPNTGVSSLIPHGaIPEENSWEIYMINSNQGPPLS

QDGSEVLLSPEVTCPGPDIMIVTFALTIFHCADVSEWHNHLKRTQOGKWEWMSVED

ESTSCYCLLFACHVILLSFEGTYALTGEFITDACAVQLKVALVGCMSCCNLSLVDNYLRC

VDNTPCQAFQEVVSDERHQGGQLEEFKLLHIFKNGTFLSLQISVLIDIPFLWRIKPFTACQE

VPFSRSVWCSNQRPLHCAFSLERYTTTQLSCKICIRCQKHEQILQAVSILESWETI

TFFAQEDSTFFPAQTGPKAFKIPYRQICATFFDDTFNAGKDWQMLAQKNSINRSNYFA

TQSSPSAVILNEARQHDGDOLSLACALEEGRTHTKLNSISEQ

> cloned sequence size: 956 aa. (SEQ ID NO.34)

HPRKQNSLRSQDTPCPCLACAQQDSHPPADEVSNILDSCLYETFQNGEQALPSI

PSAPGTLPIFIEEPDADALIKSNPIALRCKAPMQIFFKCGEWHQNEHVSEETLDES

SGKLMKVQEVFINVTRQVDFHGPEDYWCQCVAWHLGTSKSRKASVRIAYLKRNFQDPQ

GREVPQIGMVILHCRPFPQGVPAAEVEWLKNEPIDSQEDNITRADHNLNRQAARLSD

GNYTCAANIVAKRSLSATVYWNGWSTEWASONCRGQWQGRSRCTCN

PAPLNCAGFCGMSQVKTCLSLCPVDGSWEVSEWSVCSPECHLRIRECTAPPFRNNGKFCEG

LSQESNCTDGLCILGIENASDIALYSGLGAAVVAVAVLTVIGYTVLSRQSDYGVDVIDS

SALTGGFQTFNFKTVRQGNLNLHSAAMQDPLTDSRTYSGPICLQDPLKELMTESSLFPN

LSDIKVKQSSLMSVGLVESQAHEYHGKHNSRTFPQGNHNSFSTHMFRNKMFIQNLSLPL

TRTEILRTTGVFHGLGRLVMPNTGVSSLIPHGAIPEENSWEIYMINSNQGPPLS

QDGSEVLLSPEVTCPGPDIMIVTFALTIFHCADVSEWHNHLKRTQOGKWEWMSVEDSTC
YCLLDPFACHVLLDSFGTYALTGEPTDOCAVKQLKVAVFGCMSCNSLDYNLRVYCVDNTP
CAFQEVSDERHQQIGQLLEEPEKLLHFKGNTFSLQISVLDIPPFLWRIKPTACQEVPHSR
VWCNSRQPLHCASLERYTTTQLSCKICIRQLKGHEQILQVQTSILSERETITFFAQ
EDSTFFAQTPGKAFKIPYSIRQRICATFDTFNAKGKDWOQLAQKNSINRNLSYFATQSSP
SAVILNLWEARKHQHDGDLSLACALEEIGRTHTKLSNISESOQDRADEFVIRONGE
Figure 8.

Query: Cloned sequence (SEQ ID NO. 34)
(956 letters)

Database: predict.pp (SEQ ID NO. 32)
1 sequences; 887 total letters
Searching done

Sequences producing significant alignments:

Score bits (4761), Expect = 0.0
Identities = 887/887 (100%), Positives = 887/887 (100%)

Length = 887

Query: 56 LPEISAPGTLPHFIEEPDDAYIJKSNPIALRCKARPAMQIFFKCNGEWVHQNEHVEE 115
LPEISAPGTLPHFIEEPDDAYIJKSNPIALRCKARPAMQIFFKCNGEWVHQNEHVEE
Sbjct: 1 LPEISAPGTLPHFIEEPDDAYIJKSNPIALRCKARPAMQIFFKCNGEWVHQNEHVEE 60

Query: 116 TLDESSGLKVREVFINVTRQQVEDFHGPEDYWCQCVAWSHLGTSDKRASVRIALRKNF 175
TLDESSGLKVREVFINVTRQQVEDFHGPEDYWCQCVAWSHLGTSDKRASVRIALRKNF
Sbjct: 61 TLDESSGLKVREVFINVTRQQVEDFHGPEDYWCQCVAWSHLGTSDKRASVRIALRKNF 120

Query: 176 EQDPQGREVPiegMIVLHCRPPEGVPAAEVWLKNNE PIDSEQDENIDTRADHNLIIRQA 235
EQDPQGREVPiegMIVLHCRPPEGVPAAEVWLKNNE PIDSEQDENIDTRADHNLIIRQA
Sbjct: 121 EQDPQGREVPiegMIVLHCRPPEGVPAAEVWLKNNE PIDSEQDENIDTRADHNLIIRQA 180

Query: 236 RLSDSGNYTCMAANIVAKRRSLSATVVYVNGWSSWETWSACNVRCGRGQKRSRTCTN 295
RLSDSGNYTCMAANIVAKRRSLSATVVYVNGWSSWETWSACNVRCGRGQKRSRTCTN
Sbjct: 181 RLSDSGNYTCMAANIVAKRRSLSATVVYVNGWSSWETWSACNVRCGRGQKRSRTCTN 240

Query: 296 PAPLNGGAFCEGMSVQKITCTSCLCPVDGSEWVSEWSVCPECEHLRIRECTAPPNRG 355
PAPLNGGAFCEGMSVQKITCTSCLCPVDGSEWVSEWSVCPECEHLRIRECTAPPNRG
Sbjct: 241 PAPLNGGAFCEGMSVQKITCTSCLCPVDGSEWVSEWSVCPECEHLRIRECTAPPNRG 300
Figure 9.

Query: cloned protein (SEQ ID NO:34)

(956 letters)

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PDB

1,242,768 sequences; 395,571,179 total letters

Searching..............................................................done

<table>
<thead>
<tr>
<th>Score</th>
<th>E</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>gb</td>
</tr>
<tr>
<td>emb</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>gb</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
<tr>
<td>ref</td>
</tr>
</tbody>
</table>

>ref|NP_543148.1| unc-5 homolog D; KIAA1777 protein [Homo sapiens]

obj|BAB83663.1| KIAA1777 protein (Unc5H4) [Homo sapiens]

Length = 948

Score = 1870 bits (4845), Expect = 0.0
Identities = 907/919 (98%), Positives = 907/919 (98%), Gaps = 11/919 (1%)

Query: 49 GTDNSGEALPESIPSAPGTLPHFIEEPDDDAYIIKSNPIALRCKARPAMQIFFKCNWEWHQ 108
      GTDNSGEALPESIPSAPGTLPHFIEEPDDDAYIIKSNPIALRCKARPAMQIFFKCNWEWHQ
Sbjct: 30 GTDNSGEALPESIPSAPGTLPHFIEEPDDDAYIIKSNPIALRCKARPAMQIFFKCNWEWHQ 89

Query: 109 NEHVSEETLDESSGLKVREVFINVTRQVQEDFHGPEDYWCQCVASHLGTSGSRKASVRI 168
      NEHVSEETLDESSGLKVREVFINVTRQVQEDFHGPEDYWCQCVASHLGTSGSRKASVRI
Sbjct: 90 NEHVSEETLDESSGLKVREVFINVTRQVQEDFHGPEDYWCQCVASHLGTSGSRKASVRI 149

Query: 169 AYLKNFEQDPQGREGPIEMVILHCRPPEGVPAEVEWLKENNEIPIDSEQDENIDTRAH 228
      AYLKNFEQDPQGREGPIEMVILHCRPPEGVPAEVEWLKENNEIPIDSEQDENIDTRAH
Sbjct: 150 AYLKNFEQDPQGREGPIEMVILHCRPPEGVPAEVEWLKENNEIPIDSEQDENIDTRAH 209

Query: 229 NLIIRQARLSDSNYCTMAANIVAKRRSLTSATVYVVYNGGWVVWEWSACNVRCGRGWQK 288
      NLIIRQARLSDSNYCTMAANIVAKRRSLTSATVYVVYNGGWVVWEWSACNVRCGRGWQK
Sbjct: 210 NLIIRQARLSDSNYCTMAANIVAKRRSLTSATVYVVYNGGWVVWEWSACNVRCGRGWQK 269

Query: 289 RSRCTCNPAPLNNGGAGFCEGSMSVQKITCSTSLCPVDGSWEWSEWSVCSPEEHLRIRECTA 348
      RSRCTCNPAPLNNGGAGFCEGSMSVQKITCSTSLCPVDGSWEWSEWSVCSPEEHLRIRECTA
Sbjct: 270 RSRCTCNPAPLNNGGAGFCEGSMSVQKITCSTSLCPVDGSWEWSEWSVCSPEEHLRIRECTA 329

Query: 349 PPRNNGGFCEGLSQESENCDGLCIL----------GIESADIALYSGLGAAVVAVA 397
      PPRNNGGFCEGLSQESENCDGLCIL GIESADIALYSGLGAAVVAVA
Sbjct: 330 PPRNNGGFCEGLSQESENCDGLCILDKPKPLHEIKPQSGIESADIALYSGLGAAVVAVA 389

Query: 398 VLVIGVTLYRSSQSDYGVDVIDSSALTGGFQTFNFKTRQGNSLLNNSAMPQDLPVSRTY 457
      VLVIGVTLYRSSQSDYGVDVIDSSALTGGFQTFNFKTRQGNSLLNNSAMPQDLPVSRTY
Sbjct: 390 VLVIGVTLYRSSQSDYGVDVIDSSALTGGFQTFNFKTRQGNSLLNNSAMPQDLPVSRTY 449
Query: 458  SGPICLQDQLDKEILMTESLSLFNLIDSVIKKIVQSSFMVSLGVSEHAEYHKGKSNRTPFPXGN 517
SGPICLQDQLDKEILMTESLSLFNLIDSVIKKIVQSSFMVSLGVSEHAEYHKGKSNRTPFPXGN
Sbjct: 450  SGPICLQDQLDKEILMTESLSLFNLIDSVIKKIVQSSFMVSLGVSEHAEYHKGKSNRTPFPXGN 509

Query: 518  NHSFSTMPHPRNYKMQYIQNLSSLPTRLTELETRRTGFHGHLGRLVMPNTGVSSLIIHPHAIPEE 577
NHSFSTMPHPRNYKMQYIQNLSSLPTRLTELETRRTGFHGHLGRLVMPNTGVSSLIIHPHAIPEE
Sbjct: 510  NHSFSTMPHPRNYKMQYIQNLSSLPTRLTELETRRTGFHGHLGRLVMPNTGVSSLIIHPHAIPEE 569

Query: 578  NSWEYMSINQGEPSLSQDGSEVLSSLPEVTCDPIMVTTPFALTIPHCADVSSEHWNIIH 637
NSWEYMSINQGEPSLSQDGSEVLSSLPEVTCDPIMVTTPFALTIPHCADVSSEHWNIIH
Sbjct: 570  NSWEYMSINQGEPSLSQDGSEVLSSLPEVTCDPIMVTTPFALTIPHCADVSSEHWNIIH 629

Query: 638  LKKRTQGGKWEEMMSVEDESTSCYCLLDPFACHVLLDSFGTYALTGEPITDCAVKQLKVA 697
LKKRTQGGKWEEMMSVEDESTSCYCLLDPFACHVLLDSFGTYALTGEPITDCAVKQLKVA
Sbjct: 630  LKKRTQGGKWEEMMSVEDESTSCYCLLDPFACHVLLDSFGTYALTGEPITDCAVKQLKVA 689

Query: 698  VFGCMSCNSLDLINLRVYCVNDTPCPAQEVVSDERHOQGQGQLLEEPKLLHFKGNTFSLQISV 757
VFGCMSCNSLDLINLRVYCVNDTPCPAQEVVSDERHOQGQGQLLEEPKLLHFKGNTFSLQISV
Sbjct: 690  VFGCMSCNSLDLINLRVYCVNDTPCPAQEVVSDERHOQGQGQLLEEPKLLHFKGNTFSLQISV 749

Query: 758  LDIPPFLRIRKPTAQCEVPFFSRWCSNRRQPLHCAFSLEERTPTTTQSLCICIFQKLG 817
LDIPPFLRIRKPTAQCEVPFFSRWCSNRRQPLHCAFSLEERTPTTTQSLCICIFQKLG
Sbjct: 750  LDIPPFLRIRKPTAQCEVPFFSRWCSNRRQPLHCAFSLEERTPTTTQSLCICIFQKLG 809

Query: 818  EQLQVSLESERETITFFAQEDSTFPAQTGPKAFKIPYSIRQRICATFDTPANKGD 877
EQLQVSLESERETITFFAQEDSTFPAQTGPKAFKIPYSIRQRICATFDTPANKGD
Sbjct: 810  EQLQVSLESERETITFFAQEDSTFPAQTGPKAFKIPYSIRQRICATFDTPANKGD 869
Query: 878 WQMLAQKNSINRNLNSYFATQSSPSAVILNLWEARHQHDGDLDLSACAAEEIGRHTKLSN 937

WQMLAQKNSINRNLNSYFATQSSPSAVILNLWEARHQHDGDLDLSACAAEEIGRHTKLSN

Sbjct: 870 WQMLAQKNSINRNLNSYFATQSSPSAVILNLWEARHQHDGDLDLSACAAEEIGRHTKLSN 929

Query: 938 ISESQLDEADFNYSRQNGL 956

ISESQLDEADFNYSRQNGL

Sbjct: 930 ISESQLDEADFNYSRQNGL 948
Figure 10

>gb|BM726542.1|BM726542 UI-E-EJ0-aii-g-02-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone

5

UI-E-EJ0-aii-g-02-0-UI 5'.

Length = 413

Score = 274 bits (700), Expect = 1e-71

Identities = 137/137 (100%), Positives = 137/137 (100%)

Frame = +1

10

Query: 349 PPRRNGKFCGLESQSESCTDGLCILGIENASDIALYGSLGAAVVAVAVLWIGTVLYRR 408

PPRNGKFCGLESQSESCTDGLCILGIENASDIALYGSLGAAVVAVAVLWIGTVLYRR

Sbjct: 1 PPRNGKFCGLESQSESCTDGLCILGIENASDIALYGSLGAAVVAVAVLWIGTVLYRR 180

15

Query: 409 SQSDYGVDVIDSSALTGGQF7FNFKTVQGNSLNLNSAMQPDLTVSTYSGPCICLQPDLD 468

SQSDYGVDVIDSSALTGGQF7FNFKTVQGNSLNLNSAMQPDLTVSTYSGPCICLQPDLD

Sbjct: 181 SQSDYGVDVIDSSALTGGQF7FNFKTVQGNSLNLNSAMQPDLTVSTYSGPCICLQPDLD 360

Query: 469 KELMTESSSLFNPLSDIK 485

KELMTESSSLFNPLSDIK

Sbjct: 361 KELMTESSSLFNPLSDIK 411

Organism: Homo sapiens

Organ: eye

25

Tissue type: fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid

Develop. stage: fetal and adult

>gb|BU729454.1|BU729454 UI-E-CL1-afe-h-03-0-UI.r1 UI-E-CL1 Homo sapiens cDNA clone

30

UI-E-CL1-afe-h-03-0-UI 3'.

Length = 390
Score = 242 bits (617), Expect = 4e-62
Identities = 123/124 (99%), Positives = 124/124 (99%)
Frame = -3

Query: 364 ESENCTDGLCILGIEASDIALYSGLGAAVAVAVLVIGVTLYRRSOSDYGVVDIDSSAL 423
ESENCTDGLCILGIEASDIALYSGLGAAVAVAVLVIGVTLYRRSOSDYGVVDIDSSAL
Sbjct: 388 ESENCTDGLCILGIEASDIALYSGLGAAVAVAVLVIGVTLYRRSOSDYGVVDIDSSAL 209

Query: 424 TGGFQTFFKTVGNSLLLNSAMQPDQTVSRTYSGPICLQDPDLKELMTESSLFINPLSD 483
TGGFQTFFKTVGNSLLLNSAMQPDQTVSRTYSGPICLQDPDLKELMTESSLFINPLSD
Sbjct: 208 TGGFQTFFKTVGNSLLLNSAMQPDQTVSRTYSGPICLQDPDLKELMTESSLFINPLSD 29

Query: 484 IKVK 487
IKV+
Sbjct: 28 IKVE 17

Organism: Homo sapiens
Organ: eye
Tissue type: human retina
Develop. stage: adult
LPESIAPSAPG TLPHFIEEYPD DAYIIKSNPI ALRCKARPAM QIFFKCNGEWH
VHQNEHVSEE TLDESSGLKV REVFINVTRQ QVEDFHGPED YWCQCVAWSH
LGTSKSRKAS VRIAYLRKNF EQDPQGREVP IEGMIVLHRCP PPEGVPAAEV
EWLKNEEPID SEQDENIDTR AHDNLIRQA RLSDSGNYTC MAANIVAKRR
SLSATVVVVYV NGGWSSWTEW SACNVRCGRG WQKRSRTCTN PAPLNGGAFC
EGMSVQKITC TSLCPVDGSW EVWSEWSVCS PECEHLRIRE CTAPPDRNNG
KFCEGLSQES ENCDGLCIL GIENASDIAL YSGLGAAVVA VAFLVIGVTL
YRRSQQSYGV DVIDSSALTG GFOQFNFKTV RQGNSLLNS AMQPDIVDSR
TYSGPICLQD PLDKELMTES SLFNPLSDIK VKVQSSFMVS LGVSELAEHYM
GKNHSRTFPH GNNHFSSTMH PRNKMPYIQN LSSLPTRTEL RTTTVGFGHLG
GRLVMPNTGV SLLIPHGAIP EENSWEIYMS INQGEPSLQS DGSEVLLSPE
VTCGPPDMIV TTPFALTIPH CADVSSEHWN IHLKKRTQGG KWEVEVMSVED
ESTSCYCLLD PFACHVLLDS FGTYALTGEP ITDCAVKQLK VAVFGCMSN
SLDYNLRVYC VDNTPCAFQE VVSDERHQGG QLLEEKPQLH FKGNFTSLQI
SVDIPPPPFLW RIKPFTACQE VPFSRVWCSN RQPLHCAFSL ERYTPFTTQL
SCKICIRQQLK GHEQILQVQVT SILESERETI TFFAQEDSTF PAQTGPKAFK
IPYSIRQRIC ATFDTPNAKG KDQMLAQKSN SRNNSLYFA TQSSPSAVIL
NLWEARHQHD GDLDSLACAL EEIGRTHTKL SNISESQ