Title: NIDOGEN/SNEP-IKE MOLECULE

(57) Abstract: The invention is based on the discovery that the human protein referred to herein as INSP220 protein is a nidogen/snap-like molecule.
RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.


(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, PT, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW).

Published: — with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
Nidogen/snep-like molecule

This invention relates to a novel protein (termed INSP220) herein identified as a nidogen/snep-like molecule 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.

Secreted Proteins

The ability for cells to make and secrete extracellular proteins is central to many biological processes. Enzymes, growth factors, extracellular matrix proteins and signalling molecules are all secreted by cells. This is through fusion of a secretory vesicle with the plasma membrane. In most cases, but not all, proteins are directed to the endoplasmic reticulum and into secretory vesicles by a signal peptide. Signal peptides are cis-acting sequences that affect the transport of polypeptide chains from the cytoplasm to a membrane bound compartment such as a secretory vesicle. Polypeptides that are
targeted to the secretory vesicles are either secreted into the extracellular matrix or are
retained in the plasma membrane. The polypeptides that are retained in the plasma
membrane will have one or more transmembrane domains. Examples of secreted proteins
that play a central role in the functioning of a cell are cytokines, hormones, extracellular
matrix proteins (adhesion molecules), proteases, and growth and differentiation factors.

Secreted proteins of this type have been shown to play a role in diverse physiological
functions, many of which can play a role in disease processes. There remains a need for the
identification of novel polypeptides of this nature to enable new drugs to be developed for
the treatment and prevention of disease, including those diseases mentioned above.

10 THE INVENTION

The invention is based on the discovery that the human protein referred to herein as
INSP220 protein is a nidogen/snep-like molecule.

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

(i) comprises or consists of the amino acid sequence as recited in SEQ ID
NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID
NO:68;

(ii) is a fragment thereof which comprises at least a fragment of the amino
acid sequence as recited in SEQ ID NO:2 or SEQ ID NO:4 or which has an
antigenic determinant in common with the polypeptides of SEQ ID NO:2 or SEQ
ID NO:4; or

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

Preferably, a polypeptide according to the invention is a nidogen/snep-like molecule or
has an antigenic determinant in common with one or more of the polypeptides of (iii).

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as
"the INSP220 protein sequence exon 1". The polypeptide having the sequence recited in
SEQ ID NO:6 is referred to hereafter as "the INSP220 protein sequence exon 2". The
polypeptide having the sequence recited in SEQ ID NO:8 is referred to hereafter as "the
INSP220 protein sequence exon 3". The polypeptide having the sequence recited in SEQ
ID NO:10 is referred to hereafter as "the INSP220 protein sequence exon 4". The
polypeptide having the sequence recited in SEQ ID NO: 12 is referred to hereafter as "the INSP220 protein sequence exon 5". The polypeptide having the sequence recited in SEQ ID NO: 14 is referred to hereafter as "the INSP220 protein sequence exon 6". The polypeptide having the sequence recited in SEQ ID NO: 16 is referred to hereafter as "the INSP220 protein sequence exon 7". The polypeptide having the sequence recited in SEQ ID NO: 18 is referred to hereafter as "the INSP220 protein sequence exon 8". The polypeptide having the sequence recited in SEQ ID NO: 20 is referred to hereafter as "the INSP220 protein sequence exon 9". The polypeptide having the sequence recited in SEQ ID NO: 22 is referred to hereafter as "the INSP220 protein sequence exon 10". The polypeptide having the sequence recited in SEQ ID NO: 24 is referred to hereafter as "the INSP220 protein sequence exon 11". The polypeptide having the sequence recited in SEQ ID NO: 26 is referred to hereafter as "the INSP220 protein sequence exon 12". The polypeptide having the sequence recited in SEQ ID NO: 28 is referred to hereafter as "the INSP220 protein sequence exon 13". The polypeptide having the sequence recited in SEQ ID NO: 30 is referred to hereafter as "the INSP220 protein sequence exon 14". The polypeptide having the sequence recited in SEQ ID NO: 32 is referred to hereafter as "the INSP220 protein sequence exon 15". The polypeptide having the sequence recited in SEQ ID NO: 34 is referred to hereafter as "the INSP220 protein sequence exon 16". The polypeptide having the sequence recited in SEQ ID NO: 36 is referred to hereafter as "the INSP220 protein sequence exon 17". The polypeptide having the sequence recited in SEQ ID NO: 38 is referred to hereafter as "the INSP220 protein sequence exon 18". The polypeptide having the sequence recited in SEQ ID NO: 40 is referred to hereafter as "the INSP220 protein sequence exon 19". The polypeptide having the sequence recited in SEQ ID NO: 42 is referred to hereafter as "the INSP220 protein sequence exon 20". The polypeptide having the sequence recited in SEQ ID NO: 44 is referred to hereafter as "the INSP220 protein sequence exon 21". The polypeptide having the sequence recited in SEQ ID NO: 46 is referred to hereafter as "the INSP220 protein sequence exon 22". The polypeptide having the sequence recited in SEQ ID NO: 48 is referred to hereafter as "the INSP220 protein sequence exon 23". The polypeptide having the sequence recited in SEQ ID NO: 50 is referred to hereafter as "the INSP220 protein sequence exon 24". The polypeptide having the sequence recited in SEQ ID NO: 52 is referred to hereafter as "the
INSP220 protein sequence exon 25". The polypeptide having the sequence recited in SEQ ID NO:54 is referred to hereafter as "the INSP220 protein sequence exon 26". The polypeptide having the sequence recited in SEQ ID NO:56 is referred to hereafter as "the INSP220 protein sequence exon 27". The polypeptide having the sequence recited in SEQ ID NO:58 is referred to hereafter as "the INSP220 protein sequence exon 28". The polypeptide having the sequence recited in SEQ ID NO:60 is referred to hereafter as "the INSP220 protein sequence exon 29".

SEQ ID NO:62 is produced by combining SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, and 60. The polypeptide having the sequence recited in SEQ ID NO:62 is referred to hereafter as "the INSP220 full protein sequence".

Although the Applicant does not wish to be bound by this theory, it is postulated that the first 24 amino acids of the INSP220 polypeptide forms a signal peptide. The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as "the INSP220 mature protein sequence exon 1", and lacks the signal peptide.

SEQ ID NO:64 is produced by combining SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, and 60. The polypeptide having the sequence recited in SEQ ID NO:64 is referred to hereafter as "the INSP220 mature full protein sequence".

The polypeptide having the sequence recited in SEQ ID NO:66 is hereinafter referred to as the "INSP220 full protein sequence and HIS tag" and the polypeptide having the sequence recited in SEQ ID NO:68 is hereinafter referred to as the "INSP220 mature full protein sequence and HIS tag".

In one embodiment of the first aspect of the invention, the polypeptide comprises or consists of the amino acid sequence as recited in: SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68.

In a further embodiment of the first aspect, the polypeptide consists of the amino acid sequence as recited in: SEQ ID NOs:2 and 6; SEQ ID NOs:2, 6 and 8; SEQ ID NOs:2, 6, 8 and 10; SEQ ID NOs:2, 6, 8, 10 and 12; SEQ ID NOs:2, 6, 8, 10, 12 and 14; SEQ ID...
NOs: 2, 6, 8, 10, 12, 14 and 16; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16 and 18; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16 and 20; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20 and 22; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 and 26; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 and 38; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42 and 44; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44 and 46; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52 and 54; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54 and 56; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 and 58; SEQ ID NOs: 2, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58 and 60; SEQ ID NOs: 4 and 6; SEQ ID NOs: 4 and 6; SEQ ID NOs: 4, 6, 8 and 10; SEQ ID NOs: 4, 6, 8, 10 and 12; SEQ ID NOs: 4, 6, 8, 10, 12 and 14; SEQ ID NOs: 4, 6, 8, 10, 12, 14 and 16; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16 and 18; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18 and 20; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20 and 22; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 and 24; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 28; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 342; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 and 38; SEQ ID
NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40 and 42; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, and 44; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52 and 54; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54 and 56; SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 and 58; or SEQ ID NOs: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58 and 60.

The term "nidogen/snep-like molecule" refers to a molecule containing at least one nidogen/snep-like domain.

Preferably, the "nidogen/snep-like molecule" may be a molecule containing a nidogen domain detected with an e-value lower than 0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001 or 0.0000001.

Preferably, the term "nidogen/snep-like molecule" may be a molecule matching the HMM build of the Pfam entry detected with an e-value lower than 0.1, 0.01, 0.001, 0.0001, 0.0002, 0.00001, 0.000001 or 0.0000001.


Nidogen is a dumbbell shaped 150 kD sulphated glycoprotein that is found in all basement membranes and is tightly associated with laminin. The N-terminal globular
domain can self-aggregate, whilst the C-terminal globular domain binds to the short arm of laminin and to collagen IV. It probably has a role in cell-extracellular matrix interactions.

Alpha-tectorin is one of the major non-collagenous components of the tectorial membrane. The tectorial membrane is an extracellular matrix of the inner ear that covers the neuroepithelium of the cochlea and contacts the stereocilia bundles of specialized sensory hair cells. Sound induces movement of these hair cells relative to the tectorial membrane, deflects the stereocilia and leads to fluctuations in hair-cell membrane potential, transducing sound into electrical signals. Defects in alpha-tectorin are the cause of autosomal dominant non-syndromic sensorineural deafness 8 (DFNA8), deafness 12 (DFNA12) (Verhoeven, K. et al, Nat Genet. 1998 May;19(1):60-2) and deafness 21 (DFNB21) (Mustapha, M. et al, Hum Mol Genet. 1999 Mar;8(3):409-12). More recently a novel nidogen/snep-like molecule, which is also a novel marker of kidney stroma cells, has been identified in mouse, denominated as Snep (Leimeister C et al 2004, Dev Dyn. 230(2):371-7). Besides the nidogen at the N-terminus, it has a complement control protein (CCP) domain and three fibronectin type 3 (FN3) domains interspersed by several calcium-binding EGF-like (EGF_CA) domains.

The polypeptides of the present invention may modulate a variety of physiological and pathological processes or disorders. Thus, the biological activity or function of these polypeptides can be examined in systems that allow the study of such modulatory activities, using a variety of suitable assays.

Thus, preferably, the activity of a nidogen/snep-like molecule of the present invention can be confirmed in at least one of the following assays:

a. INSP220 can modulate the proliferation or the survival of normal and cancerous cells, or

b. INSP220 can modulate the proliferation or the survival of normal and cancerous hepatocytes, or

c. INSP220 can modulate the proliferation or the survival of normal and cancerous renal cells, or
d. INSP220 can modulate the apoptosis of kidney cells (e.g. kidney stroma cells), mesenchymal cells or hepatocytes.

In addition, activity of INSP220 can be demonstrated in models of cancer as reviewed by Kamb and Lassota (Drug Discovery Today: Disease Models; Vol. 1, No. 1, 2004, pp. 31-36) and/or in models of glomerulonephritis as reviewed by Alama S.D. (Drug Discovery Today: Disease Models; Vol. 1, No. 4, 2004, pp. 457-463).

The polypeptides of the first aspect of the invention may further comprise a histidine tag. Preferably the histidine tag is found at the C-terminal of the polypeptide. Preferably the histidine tag comprises 1-10 histidine residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 residues).

More preferably the histidine tag comprises 6 histidine residues.

An "antigenic determinant" of the present invention may be a part of a polypeptide of the present invention, which binds to an antibody-combining site or to a T-cell receptor (TCR). Alternatively, an "antigenic determinant" may be a site on the surface of a polypeptide of the present invention to which a single antibody molecule binds.

Generally an antigen has several or many different antigenic determinants and reacts with antibodies of many different specificities. Preferably, the antibody is immunospecific to a polypeptide of the invention. Preferably, the antibody is immunospecific to a polypeptide of the invention, which is not part of a fusion protein. Preferably, the antibody is immunospecific to INSP220 or a fragment thereof. Antigenic determinants usually consist of chemically active surface groupings of molecules, such as amino acids or sugar side chains, and can have specific three dimensional structural characteristics, as well as specific charge characteristics. Preferably, the "antigenic determinant" refers to a particular chemical group on a polypeptide of the present invention that is antigenic, i.e. that elicit a specific immune response.

The polypeptides Q8TEP7JHUMAN (SEQ ID NO: 69) and SEQ ID NO: 70 (from US20050095589 and WO2003/06605), and their encoding nucleic acid sequences are specifically excluded from the scope of this invention.

In a second aspect, the invention provides a purified nucleic acid molecule which encodes a polypeptide of the first aspect of the invention.
The term "purified nucleic acid molecule" preferably refers to a nucleic acid molecule of the invention that (1) has been separated from at least about 50 percent of proteins, lipids, carbohydrates, or other materials with which it is naturally found when total nucleic acid is isolated from the source cells, (2) is not linked to all or a portion of a polynucleotide to which the "purified nucleic acid molecule" is linked in nature, (3) is operably linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated nucleic acid molecule of the present invention is substantially free from any other contaminating nucleic acid molecule(s) or other contaminants that are found in its natural environment that would interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use. In a preferred embodiment, genomic DNA are specifically excluded from the scope of the invention. Preferably, genomic DNA larger than 10 kbp (kilo base pairs), 50 kbp, 100 kbp, 150 kbp, 200 kbp, 250 kbp or 300 kbp are specifically excluded from the scope of the invention. Preferably, the "purified nucleic acid molecule" consists of cDNA only. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP220 protein sequence exon 1), SEQ ID NO:3 (encoding the INSP220 mature protein sequence exon 1), SEQ ID NO:5 (encoding the INSP220 protein sequence exon 2), SEQ ID NO:7 (encoding the INSP220 protein sequence exon 3), SEQ ID NO:9 (encoding the INSP220 protein sequence exon 4), SEQ ID NO:11 (encoding the INSP220 protein sequence exon 5), SEQ ID NO:13 (encoding the INSP220 protein sequence exon 6), SEQ ID NO:15 (encoding the INSP220 protein sequence exon 7), SEQ ID NO:17 (encoding the INSP220 protein sequence exon 8), SEQ ID NO:19 (encoding the INSP220 protein sequence exon 9), SEQ ID NO:21 (encoding the INSP220 protein sequence exon 10), SEQ ID NO:23 (encoding the INSP220 protein sequence exon 11), SEQ ID NO:25 (encoding the INSP220 protein sequence exon 12), SEQ ID NO:27 (encoding the INSP220 protein sequence exon 13), SEQ ID NO:29 (encoding the INSP220 protein sequence exon 14), SEQ ID NO:31 (encoding the INSP220 protein sequence exon 15), SEQ ID NO:33 (encoding the INSP220 protein sequence exon 16), SEQ ID NO:35 (encoding the INSP220 protein sequence exon 17), SEQ ID NO:37 (encoding the INSP220 protein sequence exon 18), SEQ ID NO:39 (encoding the INSP220 protein sequence exon 19), SEQ ID NO:41 (encoding the
INSP220 protein sequence exon 20), SEQ ID NO:43 (encoding the INSP220 protein sequence exon 21), SEQ ID NO:45 (encoding the INSP220 protein sequence exon 22), SEQ ID NO:47 (encoding the INSP220 protein sequence exon 23), SEQ ID NO:49 (encoding the INSP220 protein sequence exon 24), SEQ ID NO:51 (encoding the INSP220 protein sequence exon 25), SEQ ID NO:53 (encoding the INSP220 protein sequence exon 26), SEQ ID NO:55 (encoding the INSP220 protein sequence exon 27), SEQ ID NO:57 (encoding the INSP220 protein sequence exon 28), SEQ ID NO:59 (encoding the INSP220 protein sequence exon 29), SEQ ID NO:61 (encoding the INSP220 full protein sequence), SEQ ID NO:63 (encoding the INSP220 mature full protein sequence), SEQ ID NO:65 (encoding the INSP220 full protein sequence and HIS tag), or SEQ ID NO:67 (encoding the INSP220 mature full protein sequence and HIS tag).

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. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (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.

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 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 inhibits the ability of a polypeptide of the first aspect of the invention to transport small, hydrophobic molecules.

Ligands to a polypeptide according to the invention may come in various forms, including natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies,
structural or functional mimetics of the aforementioned.

Such compounds may be identified using the assays and screening methods disclosed herein.

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 INSP220 polypeptides allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease. Ligands and compounds according to the sixth and seventh aspects of the invention may be identified using such methods. These methods are included as aspects of the present invention.

Compounds identified as agonists of the polypeptides of the invention may be useful for transportation of small hydrophobic molecules either in vitro or in vivo. For example, agonist compounds are useful as components of defined cell culture media, to deliver small, hydrophobic molecules to cells and protect them from degradation by enzymes present in serum.

Another aspect of this invention resides in the use of an INSP220 gene or polypeptide as a target for the screening of candidate drug modulators, particularly candidate drugs active against nidogen/snep-like molecule related disorders.

A further aspect of this invention resides in methods of screening of compounds for therapy of nidogen/snep-like molecule related disorders, comprising determining the ability of a compound to bind to an INSP220 gene or polypeptide, or a fragment thereof.

A further aspect of this invention resides in methods of screening of compounds for therapy of nidogen/snep-like molecule related disorders, comprising testing for modulation of the activity of an INSP220 gene or polypeptide, or a fragment thereof.

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 host cell of the fifth 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.

The moieties of the present invention (i.e. the polypeptides of the first aspect of the invention, a nucleic acid molecule of the second or third aspect of the invention, a vector of the fourth aspect of the invention, a host cell of the fifth aspect of the invention, a ligand of the sixth aspect of the invention, a compound of the seventh aspect of the invention) may have particular utility in the therapy or diagnosis of disorders/diseases (the two terms are used interchangeably herein) including, but not limited to hereditary deafness, disorders of the development of the kidney, the lungs, the pancreas, salivary gland, upper jaw, lower jaw, the brain and the cerebellum, and diseases involving unregulated complement activation for example, auto-immune diseases including psoriasis, adult/acute respiratory distress syndrome, bullous pemphigoid, rheumatoid arthritis, systemic lupus erythematosus, ischemia-reperfusion injury; renal and glomerular diseases; neurodegenerative, cerebrovascular and neuroinflammatory diseases including cerebral ischemia and trauma, Alzheimer's disease, multiple sclerosis; cardiovascular diseases and other pathological conditions.

The assays set forth in the Examples may also be useful for the identification of therapeutically useful moieties.

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 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 a period of time towards a control level is indicative of regression of disease.

One possible 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 nidogen/snep-like molecule.

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 host cell of the fifth 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 host cell of the fifth 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 including, but not limited to hereditary deafness, disorders of the development of the kidney, the lungs, the pancreas, salivary gland, upper jaw, lower jaw, the brain and the cerebellum, and diseases involving unregulated complement activation for example, auto-immune diseases including psoriasis, adult/acute respiratory distress syndrome, bullous pemphigoid, rheumatoid arthritis, systemic lupus erythematosus, ischemia-reperfusion injury; renal and glomerural diseases; neurodegenerative, cerebrovascular and neuroinflammatory diseases including cerebral ischemia and
trauma, Alzheimer's disease, multiple sclerosis; cardiovascular diseases and other pathological conditions.

Preferably, the polypeptides of the present invention are useful for the treatment of a hepatic/biliary disorder, a genitourinary disorder or cancer.

Preferably, cancer is selected among cancer from blood and lymphatic systems, skin cancers, liver cancer, liver carcinoma, cancer of digestive systems, cancers of urinary systems, breast cancer, ovarian cancer, gynaecological cancers, choriocarcinoma, lung cancer, brain tumors, bone tumors, carcinoid tumor, nasopharyngeal cancer, retroperitoneal sarcomas, soft tissue tumors, adenoid cystic carcinoma, choriocarcinoma, prostatic carcinoma, adenocarcinoma, neoplasm metastasis, carcinoma, melanoma, thyroid cancer or cancers of unknown primary site.

Preferably, the hepatic/biliary disorder is selected from postoperative liver disorder, hepatic granuloma, fatty liver, vascular lesion, alcoholic liver disease, liver tumor, chronic liver disease, extrahepatic biliary disorder, hepatitis, liver fibrosis or cirrhosis.

Preferably, the genitourinary disorder is selected from a glomerular disease, urinary incontinence, tubulointerstitial disease, myoneurogenic disorder, toxic nephropathy, obstructive uropathy, urinary tract infection, prostate disease, renovascular disease, disorder of the penis and scrotum, abnormal renal transport syndrome, erectile dysfunction, inherited and congenital renal disorder, urinary calculi, immunologically mediated renal disease, renal failure, kidney disease, trauma to the urinary tract, genitourinary cancer.

Preferably, the glomerular disease is selected from nephritic syndrome, nephrotic syndrome, primary glomerular disease, or secondary renal disease.

Preferably, the primary glomerular disease is selected from minimal change disease, focal segmental glomerulosclerosis, membranous glomerulonephritis, membranoproliferative glomerulonephritis, mesangial proliferative glomerulonephritis, IgA nephropathy, rapidly progressive glomerulonephritis, or fibrillar glomerulonephritis.

Preferably, the nephritic syndrome is selected from hematuria, hypertension, renal insufficiency, edema, acute glomerulonephritis, transient glomerulonephritis,
postinfectious glomerulonephritis, fulminant glomerulonephritis, rapidly progressive glomerulonephritis (RPGN), indolent glomerulonephritis, IgA nephropathy, crescentic glomerulonephritis, Pauci-immune RPGN, immune complex RPGN, Anti-GBM antibody disease autoimmunity, primary renal hematuric-proteinuric syndrome, asymptomatic hematuric-proteinuric syndrome, chronic nephritic-proteinuric syndrome, chronic glomerulonephritis, slowly progressive glomerular 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 host cell of the fifth 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.

The INSP220 polypeptides are nidogen/snep-like molecules and thus have roles in many disease states. Antagonists of the INSP220 polypeptides are of particular interest as they provide a way of modulating these disease states.

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 using in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.
As used herein, "functional equivalent" refers to a protein or nucleic acid molecule that possesses functional or structural characteristics that are substantially similar to a polypeptide or nucleic acid molecule of the present invention. A functional equivalent of a protein may contain modifications depending on the necessity of such modifications for the performance of a specific function. The term "functional equivalent" is intended to include the fragments, mutants, hybrids, variants, analogs, or chemical derivatives of a molecule.

Preferably, the "functional equivalent" may be a protein or nucleic acid molecule that exhibits any one or more of the functional activities of the polypeptides of the present invention.

Preferably, the "functional equivalent" may be a protein or nucleic acid molecule that displays substantially similar activity compared with INSP220 or fragments thereof in a suitable assay for the measurement of biological activity or function. Preferably, the "functional equivalent" may be a protein or nucleic acid molecule that displays identical or higher activity compared with INSP220 or fragments thereof in a suitable assay for the measurement of biological activity or function. Preferably, the "functional equivalent" may be a protein or nucleic acid molecule that displays 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, 100% or more activity compared with INSP220 or fragments thereof in a suitable assay for the measurement of biological activity or function.

Preferably, the "functional equivalent" may be a protein or polypeptide capable of exhibiting a substantially similar in vivo or in vitro activity as the polypeptides of the invention. Preferably, the "functional equivalent" may be a protein or polypeptide capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the polypeptides of the invention would. For example, a "functional equivalent" would be able, in an immunoassay, to diminish the binding of an antibody to the corresponding peptide (i.e., the peptide the amino acid sequence of which was modified to achieve the "functional equivalent") of the polypeptide of the invention, or to the polypeptide of the invention itself, where the antibody was raised against the corresponding peptide of the polypeptide of the invention. An equimolar concentration of the functional equivalent will diminish the
aforesaid binding of the corresponding peptide by at least about 5%, preferably between about 5% and 10%, more preferably between about 10% and 25%, even more preferably between about 25% and 50%, and most preferably between about 40% and 50%.

For example, functional equivalents can be fully functional or can lack function in one or more activities. Thus, in the present invention, variations can affect the function, for example, of the activities of the polypeptide that reflect its possession of a nidogen domain.

A summary of standard techniques and procedures which may be employed in order to utilise the invention are 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).

In a further preferred embodiment, a polypeptide of the invention, that may comprise a sequence having at least 85% of homology with INSP220. These fusion proteins can be obtained by cloning a polynucleotide encoding a polypeptide comprising a sequence having at least 85% of homology with INSP220 in frame to the coding sequences for a heterologous protein sequence.

The term "heterologous", when used herein, is intended to designate any polypeptide other than a human INSP220 polypeptide.

Example of heterologous sequences, that can be comprised in the soluble fusion proteins either at N- or at C-terminus, are the following: extracellular domains of membrane-bound protein, immunoglobulin constant regions (Fc region), multimerization domains,
domains of extracellular proteins, signal sequences, export sequences, or sequences allowing purification by affinity chromatography.

Many of these heterologous sequences are commercially available in expression plasmids since these sequences are commonly included in the fusion proteins in order to provide additional properties without significantly impairing the specific biological activity of the protein fused to them (Terpe K, Appl Microbiol Biotechnol, 60: 523-33, 2003). Examples of such additional properties are a longer lasting half-life in body fluids, the extracellular localization, or an easier purification procedure as allowed by the a stretch of Histidines forming the so-called "histidine tag" (Gentz et al, Proc Natl Acad Sci USA, 86: 821-4, 1989) or by the "HA"tag, an epitope derived from the influenza hemagglutinin protein (Wilson et al., Cell, 37: 767-78, 1994). If needed, the heterologous sequence can be eliminated by a proteolytic cleavage, for example by inserting a proteolytic cleavage site between the protein and the heterologous sequence, and exposing the purified fusion protein to the appropriate protease. These features are of particular importance for the fusion proteins since they facilitate their production and use in the preparation of pharmaceutical compositions. For example, the protein used in the examples (INSP220) can be purified by means of a hexa-histidine peptide fused at the C-terminus of INSP220. When the fusion protein comprises an immunoglobulin region, the fusion may be direct, or via a short linker peptide which can be as short as 1 to 3 amino acid residues in length or longer, for example, 13 amino acid residues in length. Said linker may be a tripeptide of the sequence E-F-M (Glu-Phe-Met), for example, or a 13-amino acid linker sequence comprising Glu-Phe-Gly-Ala-Gly-Leu-Val-Leu-Gly-Gly-Gln-Phe-Met (SEQ ID NO: 71) introduced between the sequence of the substances of the invention and the immunoglobulin sequence. The resulting fusion protein has improved properties, such as an extended residence time in body fluids (half-life), increased specific activity, increased expression level, or the purification of the fusion protein is facilitated.

In a preferred embodiment, the protein is fused to the constant region of an Ig molecule. Preferably, it is fused to heavy chain regions, like the CH2 and CH3 domains of human IgG1, for example. Other isoforms of Ig molecules are also suitable for the generation of fusion proteins according to the present invention, such as isoforms IgG2 or IgG4, or other
Ig classes, like IgM or IgA, for example. Fusion proteins may be monomeric or multimeric, hetero- or homomultimeric.

In a further preferred embodiment, the functional derivative comprises at least one moiety attached to one or more functional groups, which occur as one or more side chains on the amino acid residues. Preferably, the moiety is a polyethylene (PEG) moiety. PEGylation may be carried out by known methods, such as the ones described in WO99/55377, for example.

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 INSP220 polypeptides. 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 I, 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 INSP220 polypeptides. 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 He; 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.

In accordance with the present invention, any substitution should be preferably a "conservative" or "safe" substitution, which is commonly defined a substitution introducing an amino acids having sufficiently similar chemical properties (e.g. a basic, positively charged amino acid should be replaced by another basic, positively charged amino acid), in order to preserve the structure and the biological function of the molecule.

The literature provide many models on which the selection of conservative amino acids substitutions can be performed on the basis of statistical and physico-chemical studies on the sequence and/or the structure of proteins (Rogov SI and Nekrasov AN, 2001). Protein design experiments have shown that the use of specific subsets of amino acids can produce foldable and active proteins, helping in the classification of amino acid "synonymous" substitutions which can be more easily accommodated in protein structure, and which can be used to detect functional and structural homologs and paralogs (Murphy LR et al, 2000). The groups of synonymous amino acids and the groups of more preferred synonymous amino acids are shown in Table 1.

Specific, non-conservative mutations can be also introduced in the polypeptides of the invention with different purposes. Mutations reducing the affinity of the nidogen/snep-like molecule may increase its ability to be reused and recycled, potentially increasing its therapeutic potency (Robinson CR, 2002). Immunogenic epitopes eventually present in the polypeptides of the invention can be exploited for developing vaccines (Stevanovic S, 2002), or eliminated by modifying their sequence following known methods for selecting
mutations for increasing protein stability, and correcting them (van den Burg B and Eijsink V, 2002; WO 02/05146, WO 00/34317, WO 98/52976).

Preferred alternative, synonymous groups for amino acids derivatives included in peptide mimetics are those defined in Table 2. A non-exhaustive list of amino acid derivatives also include aminoisobutyric acid (Aib), hydroxyproline (Hyp), 1,2,3,4-tetrahydroisoquinoline-3-COOH, indoline-2-carboxylic acid, 4-difluoro-proline, L-thiazolidine-4-carboxylic acid, L-homoproline, 3,4-dehydro-proline, 3,4-dihydroxy-phenylalanine, cyclohexyl-glycine, and phenylglycine.

By "amino acid derivative" is intended an amino acid or amino acid-like chemical entity other than one of the 20 genetically encoded naturally occurring amino acids. In particular, the amino acid derivative may contain substituted or non-substituted, linear, branched, or cyclic alkyl moieties, and may include one or more heteroatoms. The amino acid derivatives can be made de novo or obtained from commercial sources (Calbiochem-Novabiochem AG, Switzerland; Bachem, USA).

Various methodologies for incorporating unnatural amino acids derivatives into proteins, using both in vitro and in vivo translation systems, to probe and/or improve protein structure and function are disclosed in the literature (Dougherty DA, 2000). Techniques for the synthesis and the development of peptide mimetics, as well as non-peptide mimetics, are also well known in the art (Golebiowski A et al, 2001; Hruby VJ and Balse PM, 2000; Sawyer TK, in "Structure Based Drug Design", edited by Veerapandian P, Marcel Dekker Inc., pg. 557-663, 1997).

Typically, greater than 30% 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 INSP220 polypeptides, or with active fragments thereof, of greater than 70% or 80%. More preferred polypeptides have degrees of identity of greater than 85%, 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 WO 01/67507) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the INSP220 exon polypeptides or the INSP220 polypeptide (SEQ ID NOs. 62 and 64), are predicted to be nidogen/snep-like molecules, by virtue of sharing significant structural homology with the INSP220 exon polypeptides or the INSP220 polypeptide or INSP220 mature polypeptide.

By "significant structural homology" is meant that the Inpharmatica Genome Threader™ predicts two proteins, or protein regions, to share structural homology with a certainty of at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome Threader™ is calculated as follows. A set of comparisons was initially performed using the Inpharmatica Genome Threader™ exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure classification (www.biochem.ucl.ac.uk/bsm/cath). This resulted in a neural network score between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an attached neural network score and the percentage confidence is a reflection of how successful the Inpharmatica Genome Threader™ was in the training/testing set.

The polypeptides of the first aspect of the invention also include fragments of the INSP220 polypeptides and fragments of the functional equivalents of the INSP220 polypeptides, provided that those fragments also comprise at least a fragment of SEQ ID NO:2 or SEQ ID NO:4 or have an antigenic determinant in common with the INSP220 exon 1 or the INSP220 mature exon 1.

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 INSP220 polypeptides 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.

Nucleic acids according to the invention are preferably 40-4100 nucleotides in length,
preferably 100-3500 nucleotides, preferably 500-3000, preferably 1000-2500, preferably
1500-2000 nucleotides in length. Polypeptides according to the invention are preferably
10-1400 amino acids in length, preferably 50-1200, preferably 100-1000, preferably 150-
500 amino acids in length, preferably 200-400 amino acids in length.

Fragments of the full length INSP220 polypeptides may consist of combinations of 1, 2,
3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
or all 29 neighbouring exon sequences in the INSP220 polypeptide sequences,
respectively. These exons may be combined with further mature fragments according to
the invention. For example, such combinations include exons 1 and 2, 2 and 3, 3 and 4, 1
and 3, 2 and 4, and so on. Such fragments are included in the present invention.

Fragments may also consist of combinations of different domains of the INSP220
protein. For example, exons are combined in order to match to identified domains, for
example to nidogen domains, EGF domains, CCP domains and/or FN3 domains

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.

By "substantially greater affinity" we mean that there is a measurable increase in the affinity for a polypeptide of the invention as compared with the affinity for other related polypeptides in the prior art such as known nidogen/snep-like molecules.

Preferably, the affinity is at least 1.5-fold, 2-fold, 5-fold 10-fold, 100-fold, 10\(^3\)-fold, 10\(^4\)-fold, 10\(^5\)-fold, 10\(^6\)-fold or greater for a polypeptide of the invention than for other related polypeptides in the prior art.

Preferably, there is a measurable increase in the affinity for a polypeptide of the invention as compared with known nidogen/snep-like molecules.

Preferably, there is a measurable increase in the affinity for a polypeptide of the invention as compared with the natural nidogen domain.

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 ah, Immunology Today 4: 72 (1983); Cole et ah, 11-96 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 ah, 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 ah, Nature, 321, 522 (1986); Verhoeyen et ah, Science, 239, 1534 (1988); Kabat et ah, J. Immunol., 147, 1709 (1991); Queen et ah, Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman et ah, Proc. Natl Acad. Sci. USA, 88, 34181 (1991); and Hodgson et ah, 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 ah, (1990), Nature 348,

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, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38 SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68 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. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 36 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 35. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 38 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 37. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 40 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 39. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 42 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 41. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 44 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 43. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 46 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 45. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 48 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 47. A nucleic acid molecule
which encodes the polypeptide of SEQ ID NO:50 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:49. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:52 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:51. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:54 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:53. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:56 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:55. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:58 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:57. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:60 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:59. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:62 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:61. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:64 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:63. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:66 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:65. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:68 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:67. These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes a polypeptide of 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, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68. Such nucleic acid molecules that encode the polypeptide of 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, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68 may include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

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.

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.

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).

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]).

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

"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, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts 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% identical over their entire length to a nucleic acid molecule encoding an INSP220 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, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38 SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, SEQ ID NO:56, SEQ ID NO:58, SEQ ID NO:60, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68), and nucleic acid molecules that are substantially complementary to these 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 a nucleic acid molecule having the sequence given in SEQ ID NO:1, SEQ ID NO:2, 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, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID
NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67 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%, 98.5%, 99% 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 INSP220 polypeptide or the INSP220 mature 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 INSP220 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding these polypeptides.

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 INSP220 polypeptides 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:2, 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, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37 SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:55, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, or SEQ ID NO:67) 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 Marathon™ 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 Applic. 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 Applic. 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 PromoterFinder™ 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 nucleotid 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.

Gene silencing approaches may also be undertaken to down-regulate endogenous expression of a gene encoding a polypeptide of the invention. RNA interference (RNAi) (Elbashir, SM et al, Nature 2001, 411, 494-498) is one method of sequence specific post-transcriptional gene silencing that may be employed. Short dsRNA oligonucleotides are synthesised in vitro and introduced into a cell. The sequence specific binding of these dsRNA oligonucleotides triggers the degradation of target mRNA, reducing or ablating target protein expression.

Efficacy of the gene silencing approaches assessed above may be assessed through the measurement of polypeptide expression (for example, by Western blotting), and at the RNA level using TaqMan-based methodologies.

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, transfested 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

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. The vectors pCR4-TOPO, pCR4-TOPO-INS220, pDONR221, pDONR221-INS220-6HIS, pEAK12d, pDEST12.2, pEAK12d-INS220-6HIS and pDEST12.2-INS220-6HIS are preferred examples of suitable vectors for use in accordance with the aspects of this invention relating to INS220.

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, promotors 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 pSportI™ plasmid (Gibco BRL) and the like may be used.
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 have 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, Y. 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, DJ. 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.

As indicated above, the present invention also provides novel targets and methods for the screening of drug candidates or leads. These screening methods include binding assays and/or functional assays, and may be performed in vitro, in cell systems or in animals.

In this regard, a particular object of this invention resides in the use of an INSP220 polypeptide as a target for screening candidate drugs for treating or preventing nidogen/snep-like molecule related disorders.

Another object of this invention resides in methods of selecting biologically active compounds, said methods comprising contacting a candidate compound with a INSP220 gene or polypeptide, and selecting compounds that bind said gene or polypeptide.
A further other object of this invention resides in methods of selecting biologically active compounds, said method comprising contacting a candidate compound with recombinant host cell expressing a INSP220 polypeptide with a candidate compound, and selecting compounds that bind said INSP220 polypeptide at the surface of said cells and/or that modulate the activity of the INSP220 polypeptide.

A "biologically active" compound denotes any compound having biological activity in a subject, preferably therapeutic activity, more preferably a compound having nidogen domain activity, and further preferably a compound that can be used for treating INSP220 related disorders, or as a lead to develop drugs for treating nidogen/snep-like molecule related disorder. A "biologically active" compound preferably is a compound that modulates the activity of INSP220.

The above methods may be conducted in vitro, using various devices and conditions, including with immobilized reagents, and may further comprise an additional step of assaying the activity of the selected compounds in a model of nidogen/snep-like molecule related disorder, such as an animal model.

Preferred selected compounds are agonists of INSP220, i.e., compounds that can bind to INSP220 and mimic the activity of an endogenous ligand thereof.

A further object of this invention resides in a method of selecting biologically active compounds, said method comprising contacting in vitro a test compound with a INSP220 polypeptide according to the present invention and determining the ability of said test compound to modulate the activity of said INSP220 polypeptide.

A further object of this invention resides in a method of selecting biologically active compounds, said method comprising contacting in vitro a test compound with a INSP220 gene according to the present invention and determining the ability of said test compound to modulate the expression of said INSP220 gene, preferably to stimulate expression thereof.

In another embodiment, this invention relates to a method of screening, selecting or identifying active compounds, particularly compounds active on multiple sclerosis or related disorders, the method comprising contacting a test compound with a recombinant
host cell comprising a reporter construct, said reporter construct comprising a reporter

gene under the control of a INSP220 gene promoter, and selecting the test compounds

that modulate (e.g. stimulate or reduce, preferably stimulate) expression of the reporter
gene.

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


Binding to a target gene or polypeptide provides an indication as to the ability of the

compound to modulate the activity of said target, and thus to affect a pathway leading to

Nidogen/snep-like molecule related disorder in a subject. The determination of binding

may be performed by various techniques, such as by labelling of the candidate

compound, by competition with a labelled reference ligand, etc. For in vitro binding

assays, the polypeptides may be used in essentially pure form, in suspension, immobilized

on a support, or expressed in a membrane (intact cell, membrane

preparation, liposome, etc.).

Modulation of activity includes, without limitation, stimulation of the surface expression

of the INSP220 receptor, modulation of multimerization of said receptor (e.g., the

formation of multimeric complexes with other sub-units), etc. The cells used in the assays

may be any recombinant cell (i.e., any cell comprising a recombinant nucleic acid

encoding a INSP220 polypeptide) or any cell that expresses an endogenous INSP220

polypeptide. Examples of such cells include, without limitation, prokaryotic cells (such as
bacteria) and eukaryotic cells (such as yeast cells, mammalian cells, insect cells, plant cells, etc). Specific examples include *E.coli, Pichia pastoris, Hansenula polymorpha, Schizosaccharomyces pombe, Kluveromyces* or *Saccharomyces* yeasts, mammalian cell lines (e.g., Vero cells, CHO cells, 3T3 cells, COS cells, etc.) as well as primary or established mammalian cell cultures (e.g., produced from fibroblasts, embryonic cells, epithelial cells, nervous cells, adipocytes, etc.).

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.

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

(a) contacting a cell expressing (optionally 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.

Methods for generating detectable signals in the types of assays described herein will be known to those of skill in the art. A particular example is cotransfecting a construct expressing a polypeptide according to the invention, or a fragment such as the LBD, in fusion with the GAL4 DNA binding domain, into a cell together with a reporter plasmid, an example of which is pFR-Luc (Stratagene Europe, Amsterdam, The Netherlands). This particular plasmid contains a synthetic promoter with five tandem repeats of GAL4 binding sites that control the expression of the luciferase gene. When a potential ligand is added to the cells, it will bind the GAL4-polypeptide fusion and induce transcription of the luciferase gene. The level of the luciferase expression can be monitored by its activity using a luminescence reader (see, for example, Lehman et al. JBC 270, 12953, 1995; Pawar et al. JBC, 277, 39243, 2002).

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

(a) contacting a labelled or unlabeled compound with the polypeptide immobilized on any solid support (for example beads, plates, matrix support, chip) and detection of the compound by measuring the label or the presence of the compound itself; or

(b) contacting a cell expressing on the surface thereof the polypeptide, by means of artificially anchoring it to the cell membrane, or by constructing a chimeric receptor 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

(c) 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.
For example, a method such as FRET detection of ligand bound to the polypeptide in the presence of peptide co-activators (Norris et al, Science 285, 744, 1999) might be used.

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

(a) contacting a cell expressing (optionally 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 express a polypeptide of the invention (and which optionally 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, optionally 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.

Similarly, there is provided a method of screening for a polypeptide antagonist or agonist compound which comprises the steps of:

(a) incubating a labelled ligand with a polypeptide according to the invention on any solid support or the cell surface, or a cell membrane containing a polypeptide of the invention.

(b) measuring the amount of labelled ligand bound to the polypeptide on the solid support, whole cell or the cell membrane;

(c) adding a candidate compound to a mixture of labelled ligand and immobilized polypeptide on the solid support, 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 immobilized polypeptide or 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.

Assay methods that are also included within the terms of the present invention are those that involve the use of the genes and polypeptides of the invention in overexpression or ablation assays. Such assays involve the manipulation of levels of these
genes/polypeptides in cells and assessment of the impact of this manipulation event on the physiology of the manipulated cells. For example, such experiments reveal details of signalling and metabolic pathways in which the particular genes/polypeptides are implicated, generate information regarding the identities of polypeptides with which the studied polypeptides interact and provide clues as to methods by which related genes and proteins are regulated.

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.

In another embodiment, this invention relates to the use of a INSP220 polypeptide or fragment thereof, whereby the fragment is preferably a INSP220 gene-specific fragment, for isolating or generating an agonist or stimulator of the INSP220 polypeptide for the
treatment of an immune related disorder, wherein said agonist or stimulator is selected from the group consisting of:

1. a specific antibody or fragment thereof including: a) a chimeric, b) a humanized or c) a fully human antibody, as well as;

2. a bispecific or multispecific antibody,

3. a single chain (e.g. scFv) or

4. single domain antibody, or

5. a peptide- or non-peptide mimetic derived from said antibodies or

6. an antibody-mimetic such as a) an anticalin or b) a fibronectin-based binding molecule (e.g. trinectin or adnectin).

The generation of peptide- or non-peptide mimetics from antibodies is known in the art (Saragovi et al., 1991 and Saragovi et al., 1992).

Anticalins are also known in the art (Vogt et al., 2004). Fibronectin-based binding molecules are described in US6818418 and WO2004029224.

Furthermore, the test compound may be of various origin, nature and composition, such as any small molecule, nucleic acid, lipid, peptide, polypeptide including an antibody such as a chimeric, humanized or fully human antibody or an antibody fragment, peptide- or non-peptide mimetic derived therefrom as well as a bispecific or multispecific antibody, a single chain (e.g. scFv) or single domain antibody or an antibody-mimetic such as an anticalin or fibronectin-based binding molecule (e.g. trinectin or adnectin), etc., in isolated form or in mixture or combinations.

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.

As mentioned above, it is envisaged that the various moieties of the invention (i.e. the polypeptides of the first aspect of the invention, a nucleic acid molecule of the second or
third aspect of the invention, a vector of the fourth aspect of the invention, a host cell of
the fifth aspect of the invention, a ligand of the sixth aspect of the invention, a compound
of the seventh aspect of the invention) may be useful in the therapy or diagnosis of
diseases. To assess the utility of the moieties of the invention for treating or diagnosing a
disease one or more of the following assays may be carried out. Note that although some
of the following assays refer to the test compound as being a protein/polypeptide, a
person skilled in the art will readily be able to adapt the following assays so that the other
moieties of the invention may also be used as the "test compound".

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 [herein, 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%. 98.5% 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., NJ. 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, intrarterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous
applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal. 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 ah, 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 in vivo or ex vivo. 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, 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 Muzychka, 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 (i.e. to prevent infection) or therapeutic (i.e. 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.

Genetic delivery of antibodies that bind to polypeptides according to the invention may also be effected, for example, as described in International patent application WO98/55607.

The technology referred to as jet injection (see, for example, www.powderject.com) may also be useful in the formulation of vaccine compositions.

A number of suitable methods for vaccination and vaccine delivery systems are described in International patent application WO00/29428.

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 radiolabeled 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 WO95/1 1995 (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 W095/25116 (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 certain diseases including, but not limited to, hereditary deafness, disorders of the development of the kidney, the lungs, the pancreas, salivary gland, upper jaw, lower jaw, the brain and the cerebellum, and diseases involving unregulated complement activation for example, auto-immune diseases including psoriasis, adult/acute respiratory distress syndrome, bullous pemphigoid, rheumatoid arthritis, systemic lupus erythematosus, ischemia-reperfusion injury; renal and glomerular diseases; neurodegenerative, cerebrovascular and neuroinflammatory diseases including cerebral ischemia and trauma, Alzheimer’s disease, multiple sclerosis; cardiovascular diseases and other pathological conditions.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to INSP220 polypeptides.

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**: INSP220 Protein Sequence Domain Structure. The sequence is annotated to show the domains of INSP220 as follows: Boxed - Signal peptide; Bold - Nidogen-like (NIDO) domain; Grey - Calcium-binding EGF-like (EGF_CA) domain; Italic & Bold - Complement control protein (CCP) domain; Underlined - Fibronectin type 3 (FN3) domain.

**Figure 2**: Top 10 BLAST hits for INSP220 full protein sequence against NCBI-nr database.

**Figure 3**: Top annotated BLAST alignment for INSP220 full protein sequence against NCBI-nr database.
TABLE 1

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Synonymous Groups</th>
<th>More Preferred Synonymous Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Gly, Ala, Ser, Thr, Pro</td>
<td>Thr, Ser</td>
</tr>
<tr>
<td>Arg</td>
<td>Asn, Lys, Gln, Arg, His</td>
<td>Arg, Lys, His</td>
</tr>
<tr>
<td>Leu</td>
<td>Phe, Ile, Val, Leu, Met</td>
<td>Ile, Val, Leu, Met</td>
</tr>
<tr>
<td>Pro</td>
<td>Gly, Ala, Ser, Thr, Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>Thr</td>
<td>Gly, Ala, Ser, Thr, Pro</td>
<td>Thr, Ser</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly, Thr, Pro, Ala, Ser</td>
<td>Gly, Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Met, Phe, Ile, Leu, Val</td>
<td>Met, Ile, Val, Leu</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala, Thr, Pro, Ser, Gly</td>
<td>Gly, Ala</td>
</tr>
<tr>
<td>Ile</td>
<td>Phe, Ile, Val, Leu, Met</td>
<td>Ile, Val, Leu, Met</td>
</tr>
<tr>
<td>Phe</td>
<td>Trp, Phe, Tyr</td>
<td>Tyr, Phe</td>
</tr>
<tr>
<td>Tyr</td>
<td>Trp, Phe, Tyr</td>
<td>Phe, Tyr</td>
</tr>
<tr>
<td>Cys</td>
<td>Ser, Thr, Cys</td>
<td>Cys</td>
</tr>
<tr>
<td>His</td>
<td>Asn, Lys, Gln, Arg, His</td>
<td>Arg, Lys, His</td>
</tr>
<tr>
<td>Gln</td>
<td>Glu, Asn, Asp, Gln</td>
<td>Asn, Gln</td>
</tr>
<tr>
<td>Asn</td>
<td>Glu, Asn, Asp, Gln</td>
<td>Asn, Gln</td>
</tr>
<tr>
<td>Lys</td>
<td>Asn, Lys, Gln, Arg, His</td>
<td>Arg, Lys, His</td>
</tr>
<tr>
<td>Asp</td>
<td>Glu, Asn, Asp, Gln</td>
<td>Asp, Glu</td>
</tr>
<tr>
<td>Glu</td>
<td>Glu, Asn, Asp, Gln</td>
<td>Asp, Glu</td>
</tr>
<tr>
<td>Met</td>
<td>Phe, Ile, Val, Leu, Met</td>
<td>Ile, Val, Leu, Met</td>
</tr>
<tr>
<td>Trp</td>
<td>Trp, Phe, Tyr</td>
<td>Trp</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Synonymous Groups</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>D-Pro, L-1-thiazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>D-Ala, Gly, Aib, B-Ala, Acp, L-Cys, D-Cys</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met, AdaA, AdaG</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ala, D-Ala, Pro, D-Pro, Aib, beta-Ala, Acp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, AdaA, AdaG, cis-3,4, or 5-phenylproline, Bpa, D-Bpa</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>D-Tyr, Phe, D-Phe, L-Dopa, His, D-His</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>D-Cys, S-Me--Cys, Met, D-Met, Thr, D-Thr</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>D-Met, S-Me--Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val</td>
<td></td>
</tr>
</tbody>
</table>
Examples

Example 1: INSP220

The INSP220 polypeptide sequence was used as a BLAST query against the NCBI non-redundant sequence database. The top ten matches from the BLAST query are shown in Figure 2. Figure 3 shows the alignment of the INSP220 query sequence to the top blast hit.

The cloning of the INSP220 gene from human genomic DNA will allow the high level expression of the INSP220 protein in prokaryotic or eukaryotic expression systems and its subsequent purification and characterisation. For example, recombinant INSP220 may be used to generate INSP220-specific monoclonal or polyclonal antibodies which might then be used in the biochemical characterisation of INSP220. Alternatively, recombinant INSP220 may be used in a wide variety of screening assays. Whenever it is appropriate, heterologous DNA sequences encoding protein sequences that assist with the expression, the secretion, and/or the detection of these recombinant sequences may be added in frame to the DNA encoding INSP220 or INSP220-derived proteins.

Example 2: Cloning of INSP220

2.1 Preparation of human cDNA templates

First strand cDNA can be prepared from a variety of human tissue total RNA samples (Clontech, Stratagene, Ambion, Biochain Institute and in-house preparations) using Superscript II or Superscript III RNase H Reverse Transcriptase (Invitrogen) according to the manufacturer’s protocol.

For Superscript II: Oligo (dT)$_{15}$ primer (1 µl at 500 µg/ml) (Promega), 2 µg human total RNA, 1 µl 10 mM dNTP mix (10 mM each of dATP, dGTP, dCTP and dTTP at neutral pH) and sterile distilled water to a final volume of 12 µl is combined in a 1.5 ml Eppendorf tube, heated to 65 °C for 5 min and chilled on ice. The contents are collected by brief centrifugation and 4 µl of 5X First-Strand Buffer, 2 µl 0.1 M DTT, and 1 µl RnaseOUT™ Recombinant Ribonuclease Inhibitor (40 units/µl, Invitrogen) are added. The contents of the tube are mixed gently and incubated at 42 °C for 2 min, then 1 µl (200 units) of Superscript II™ enzyme is added and mixed gently by pipetting. The mixture is incubated at 42 °C for 50 min and then inactivated by heating at 70 °C for 15
min. To remove RNA complementary to the cDNA, 1 µl (2 units) of *E. coli* RNase H (Invitrogen) is added and the reaction mixture incubated at 37 °C for 20 min.

For Superscript III: 1 µl Oligo(dT)$_{20}$ primer (50 µM, Invitrogen), 2 µg human total RNA, 1 µl 10 mM dNTP mix (10 nM each of dATP, dGTP, dCTP and dTTP at neutral pH) and sterile distilled water to a final volume of 10 µl are combined in a 1.5 ml Eppendorf tube, heated to 65 °C for 5 min and then chilled on ice. For each RT reaction a cDNA synthesis mix is prepared as follows: 2 µl 10X RT buffer, 4 µl 25 mM MgCl$_2$, 2 µl 0.1 M DTT, 1 µl RNaseOUT™ (40 U/µl) and 1 µl Superscript III™ RT enzyme are combined in a separate tube and then 10 µl of this mix is added to the tube containing the RNA/primer mixture. The contents of the tube are mixed gently, collected by brief centrifugation, and incubated at 50 °C for 50 min. The reaction is terminated by incubating at 80 °C for 5 min and the reaction mixture is then chilled on ice and collected by brief centrifugation.

To remove RNA complementary to the cDNA, 1 µl (2 units) of *E. coli* RNase H (Invitrogen) is added and the reaction mixture is incubated at 37 °C for 20 min.

The final 21 µl reaction mix is diluted by adding 179 µl sterile water to give a total volume of 200 µl. The RNA samples are combined into pools such that each pool contained up to five different cDNA samples. 5 µl of each cDNA pool is used as a template for PCR in a 50 µl final reaction volume and this should consist of 1 µl of each cDNA sample in that pool. This represents approximately 20 ng of each individual cDNA template.

2.2 cDNA Libraries

Human cDNA libraries (in bacteriophage lambda (λ) vectors) can be purchased from Stratagene, Clontech or Invitrogen, or prepared at the Serono Pharmaceutical Research Institute in λ ZAP, λ GTIO, λ GTI 1, or TriplEx2 vectors according to the manufacturer's protocol (Stratagene and Clontech). Bacteriophage λ DNA is 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).

2.3 Gene specific cloning primers for PCR
Two pairs of PCR primers having a length of between 18 and 30 bases may be designed for amplifying the INSP220 predicted cds using Primer Designer Software (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, USA). PCR primers are optimized to have a Tm close to 55 ± 10 °C and a GC content of 40-60%. Primers are selected which have high selectivity for the target sequence (INSP220) with little or no none specific priming. The primers are designed to form two nested pairs such that INSP220-CP3/INSP220-CP4 primers are positioned internally to primers INSP220-CP1/INSP220-CP2.

2.4 PCR amplification of INSP220 from human cDNA templates

Gene-specific cloning primers INSP220-CP1 and INSP220-CP2 can be designed to amplify a cDNA fragment containing the INSP220 cds. The primer pair is used with the panel of cDNA libraries and pools of human cDNA samples as PCR templates. This PCR1 is performed in a final volume of 50 µl containing 1X AmpliTaq™ buffer, 200 µM dNTPs, 50 pmoles of each cloning primer, 2.5 units of AmpliTaq™ (Applied Biosystems) and approximately 20 ng of cDNA library template or 100 ng cDNA pool template using an MJ Research DNA Engine, programmed as follows: 94 °C, 2 min; 40 cycles of 94 °C, 1 min, 55 °C, 1 min, and 72 °C, 1 min; followed by 1 cycle at 72 °C for 7 min and a holding cycle at 4 °C.

Each PCR1 product is then used as the template for PCR2 using amplification primers INSP220-CP3 and INSP220-CP4 designed to amplify a cDNA fragment within the INSP220-CP1/INSP220-CP2 product. PCR2 is performed in a final volume of 50 µl containing 1X AmpliTaq™ buffer, 200 µM dNTPs, 50 pmoles of each cloning primer, 2.5 units of AmpliTaq™ (Applied Biosystems), and 1 µl of PCR1 product using an MJ Research DNA Engine, programmed as follows: 94 °C, 2 min; 40 cycles of 94 °C, 1 min, 59 °C, 1 min, and 72 °C, 1 min; followed by 1 cycle at 72 °C for 7 min and a holding cycle at 4 °C.

30 µl of each PCR1 and PCR2 amplification product is visualized on a 0.8 % agarose gel in 1 X TAE buffer (Invitrogen). Products of the approximately the expected molecular weight are purified from the gel using the Wizard PCR Preps DNA Purification System (Promega), eluted in 50 µl of water and subcloned directly.
2.5 **Subcloning of PCR Products**

The PCR products were subcloned into the topoisomerase I modified cloning vector (pCR4-TOPO) using the TA cloning kit purchased from the Invitrogen Corporation using the conditions specified by the manufacturer. Briefly, 4 µl of gel purified PCR product is incubated for 15 min at room temperature with 1 µl of TOPO vector and 1 µl salt solution. The reaction mixture is then transformed into *E. coli* strain TOPIO (Invitrogen) as follows: a 50 µl aliquot of One Shot TOPIO cells is thawed on ice and 2 µl of TOPO reaction is added. The mixture is incubated for 15 min on ice and then heat shocked by incubation at 42 °C for exactly 30 s. Samples are returned to ice and 250 µl of warm (room temperature) SOC media was added. Samples are incubated with shaking (220 rpm) for 1 h at 37 °C. The transformation mixture is then plated on L-broth (LB) plates containing ampicillin (100 µg/ml) and incubated overnight at 37 °C.

2.6 ** colonies PCR**

Colonies are 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 containing 1X AmpliTaq™ buffer, 200 µM dNTPs, 20 pmoles of T7 primer, 20 pmoles of T3 primer, 1 unit of AmpliTaq™ (Applied Biosystems) using an MJ Research DNA Engine. The cycling conditions are as follows: 94 °C, 2 min; 30 cycles of 94 °C, 30 sec, 48 °C, 30 sec and 72 °C for 1 min. Samples are maintained at 4 °C (holding cycle) before further analysis.

PCR products are analyzed on 1 % agarose gels in 1 X TAE buffer. Colonies which give PCR products of approximately the expected molecular weight are grown up overnight at 37 °C in 5 ml L-Broth (LB) containing ampicillin (100 µg/ml), with shaking at 220 rpm.

2.7 **Plasmid DNA preparation and sequencing**

Miniprep plasmid DNA is prepared from the 5 ml cultures using a Biorobot 8000 robotic system (Qiagen) or Wizard Plus SV Minipreps kit (Promega cat. no. 1460) according to the manufacturer's instructions. Plasmid DNA is eluted in 80 µl of sterile water. The DNA concentration is measured using an Eppendorf BO photometer or Spectramax 190 photometer (Molecular Devices). Plasmid DNA (200-500 ng) is subjected to DNA
sequencing with the sequencing primers T7 and T3 using the BigDye Terminator system (Applied Biosystems cat. no. 4390246) according to the manufacturer's instructions. Sequencing reactions are purified using Dye-Ex columns (Qiagen) or Montage SEQ 96 cleanup plates (Millipore cat. no. LSKS09624) then analyzed on an Applied Biosystems 3700 sequencer.

Example: 3 Generation of Gateway compatible INSP220 ORF fused to an in frame 6HIS tag sequence.

INSP220 is cloned by nested PCR and therefore the cDNA insert in the pCR4-TOPO clone (plasmid pCR4-TOPO-INSP220) may be missing several nucleotides at the 5' end and at the 3' end of the coding sequence. Incorporation of missing nucleotides, 6HIS tag and stop codon can all be accomplished by including the appropriate nucleotides in the primers used for PCR amplification.

Plasmid pCR4-TOPO-INSP220 is used as a PCR template to generate the full-length ORF containing a C-terminal 6HIS tag and a stop codon. The first stage of this Gateway cloning process involves a two step PCR reaction which generates the full-length ORF of INSP220 flanked at the 5' end by an attB1 recombination site and Kozak sequence, and flanked at the 3' end by a sequence encoding an in-frame 6 histidine (6HIS) tag, a stop codon and the attB2 recombination site (Gateway compatible cDNA). The first PCR reaction PCRI, (in a final volume of 50 µl) contains respectively: 1 µl (25 ng) of plasmid pCR4-TOPO-INSP220, 4.0 µl dNTPs (10 mM), 5 µl of 10X Pfx polymerase buffer, 1 µl MgSO4 (50 mM), 1.0 µl each of gene specific primer (to give a final concentration of 100 pmoles), and 0.5 µl Platinum Pfx DNA polymerase (Invitrogen). The PCR reaction is performed using an initial denaturing step of 95 °C for 2 min, followed by 30 cycles of 94 °C for 30 s; 64°C for 30 s and 68 °C for 1 min; and a final extension cycle of 68 °C for 5 minutes and a holding cycle of 4 °C. The amplification product is directly purified using the Perfectprep Gel cleanup kit (Eppendorf) and recovered in 50 µl sterile water according to the manufacturer's instructions. A 2 µl aliquot is visualized on 1.6 % agarose gel in 1 X TAE buffer in order to verify that the product is of the expected molecular weight.
The second PCR reaction (in a final volume of 50 µl) contains 1 µl of diluted purified PCR1 product (to a final concentration of 10 ng), 4.0 µl dNTPs (10 mM), 5 µl of 10X Pfx polymerase buffer, 1 µl MgSO₄ (50 mM), 1.0 µl of each Gateway conversion primer (to give a final concentration of 100 pmoles) and 0.5 µl of Platinum Pfx DNA polymerase. The conditions for the 2nd PCR reaction are: 95 °C for 2 min, followed by 30 cycles of 94 °C for 30 s; 60°C for 30 s and 68 °C for 1 min; and a final extension cycle of 68 °C for 5 minutes and a holding cycle of 4 °C. The PCR product is gel purified using the Perfectprep Gel cleanup kit (Eppendorf) and recovered in 50 µl sterile water according to the manufacturer’s instructions. A 2 µl aliquot is visualized on 1.6 % agarose gel in 1X TAE buffer in order to verify that the product was of the expected molecular weight.

3.1 Subcloning of Gateway compatible INSP220-6HIS ORF into Gateway entry vector pDONR221

The second stage of the Gateway cloning process involved subcloning of the Gateway modified PCR product into the Gateway entry vector pDONR221 (Invitrogen) as follows: 5 µl of gel extracted product from PCR2 is incubated with 1.5 µl pDONR221 vector (0.1 µg/µl), 2 µl BP buffer and 1.5 µl of BP clonase enzyme mix (Invitrogen) in a final volume of 10 µl at RT for 1h. The reaction is stopped by addition of 1 µl proteinase K (2 µg/µl) and incubated at 37 °C for a further 10 min. An aliquot of this reaction (2 µl) is used to transform DH5α strain (Invitrogen) as follows: a 50 µl aliquot of DH5α cells is thawed on ice and 2 µl of reaction mixture added. The mixture is incubated for 30 min on ice and then heat shocked by incubation at 42 °C for exactly 30 s. Samples are returned to ice and 250 µl of warm SOC media (room temperature) is added. Samples are incubated with shaking (250 rpm) for 1 h at 37 °C. The transformation mixture is then plated on L-broth (LB) plates containing kanamycin (40 µg/ml) and incubated overnight at 37 °C.

Five transformants are picked and patched on LB agar plates containing kanamycin (40 µg/ml) and incubated overnight at 37 °C. A scoop of the grown culture from the patched plate is resuspended in 50 µl of water and boiled for 5 minutes to lyse the cells. The cell lysate was centrifuged to remove the cell debris and the supernatant obtained is used as a template for colony PCR screening.
The PCR mixture (in a final volume of 25 µl) contains 10 µl of the centrifuged cell lysate, 2.0 µl dNTPs (10 mM), 2.5 µl of Taq polymerase buffer, 0.5 µl of screening primers (to give a final concentration of 100 picomoles) and 0.5 µl of Taq DNA polymerase.

The conditions for the screening PCR reaction are: 95 °C for 2 min, followed by 30 cycles of 94 0C for 30 s; 60°C for 30 s and 72 °C for 1 min; and a final extension cycle of 72 °C for 5 minutes and a holding cycle of 4 °C. The PCR products are loaded onto a 1.6 % agarose gel to verify the fragment size.

One positive clone is selected and plasmid mini-prep DNA is prepared from 5 ml cultures using QIAprep Spin Miniprep kit (Qiagen). Plasmid DNA (150-200 ng) is subjected to DNA sequencing with 21M13 and M13Rev primers using the CEQ Dye Terminator Cycle sequencing Quick Start Kit (Beckman Coulter P/N 608120) according to the manufacturer's instructions. Sequencing reactions are analyzed on CEQ 2000 XL DNA analysis system (Beckman Coulter P/N 608450). After sequence confirmation of the insert, pDONR221_INSP220-6HIS is then used for creating the expression clones.

3.2 Subcloning of Gateway compatible INSP2201 ORF into expression vectors pEAKUd and pDEST12.2

Plasmid DNA (2 µl or approx. 150 ng) of pDONR221_INSP2201-6HIS is then used in a recombination reaction containing 1.5 µl of either pEAK12d vector or pDEST12.2 vector (0.1 µg/µl), 2 µl LR buffer and 1.5 µl of LR clonase (Invitrogen) in a final volume of 10 µl.

The reaction is stopped by addition of 1 µl proteinase K (2 µg/µl) and incubated at 37 °C for a further 10 min. An aliquot of this reaction (2 µl) is used to transform DH5α strain (Invitrogen) as follows: a 50 µl aliquot of DH5α cells was thawed on ice and 2 µl of reaction mixture added. The mixture is incubated for 30 min on ice and then heat shocked by incubation at 42 °C for exactly 30 s. Samples are returned to ice and 250 µl of warm SOC media (room temperature) is added. Samples are incubated with shaking (250 rpm) for 1 h at 37 °C. The transformation mixture is then plated on L-broth (LB) plates containing Ampicillin (100 µg/ml) and incubated overnight at 37 °C.

Transformants are then picked and patched on LB agar plates containing Ampicillin (100 µg/ml) and incubated overnight at 37 °C. A scoop of the grown culture from the patched
plate is resuspended in 50 µl of water and boiled for 5 minutes to lyse the cells. The cell lysate is centrifuged to remove the cell debris and the supernatant obtained is used as a template for colony PCR screening.

The PCR mixture (in a final volume of 25 µl) contained 10 µl of the centrifuged cell lysate, 2.0 µl dNTPs (10 mM), 2.5 µl of Taq polymerase buffer, 0.5 µl of screening primers (to give a final concentration of 100 picomoles and 0.5 µl of Taq DNA polymerase. The conditions for the screening PCR reaction are: 95 °C for 2.min, followed by 30 cycles of 94 °C for 30 s; 60°C for 30 s and 72 °C for 1 min; and a final extension cycle of 72 °C for 5 minutes and a holding cycle of 4 °C. The PCR products are loaded onto a 1.6 % agarose gel to verify the fragment size.

One positive clone can be selected and plasmid mini-prep DNA was prepared from 5 ml cultures using QIAprep Spin Miniprep kit (Qiagen).

Plasmid DNA (150 - 200 ng) in the pEAK12d vector is subjected to DNA sequencing with the sequencing primers pEAK12 FP and pEAK12 RP as described above. Plasmid DNA (150 - 200 ng) in the pDEST12.2 vector is subjected to DNA sequencing with the sequencing primers 21M13 FP and M13Rev RP as described above.

Sequence confirmed clones are designated as pEAK12d_INSP220-6HIS and pDEST1.2JNSP220-6HIS.

Maxi-prep DNA is prepared from a 500 ml culture of the sequence verified clone pEAK12d_INSP220-6HIS using a Qiagen Maxi prep kit according to the manufacturer's instructions. Plasmid DNA is resuspended at a concentration of 1 µg/µl in TE buffer and stored at -20 °C.

Endotoxin-free maxi-prep DNA is prepared from a 500 ml culture of the sequence verified clone (pDEST12.2_INSP220-6HIS) using the EndoFree Plasmid Mega kit (Qiagen) according to the manufacturer's instructions. Purified plasmid DNA is resuspended in endotoxin free TE buffer at a final concentration of at least 3 µg/µl and stored at -20 °C.

Example 4: Assays Suitable for Exploration of the Biological Relevance of INSP220 Function
Sequences of this nature are predicted to be extracellular matrix proteins and to bind/interact with other extracellular matrix proteins.

The protein nidogen, which has a nidogen domain, has been characterized. Nidogens (entactins) are ubiquitous components of basement membranes. Vertebrate Nidogen-1 and Nidogen-2 consist of three globular regions, G1-G3, linked by a flexible linker (Gl-G2) and a rod-like tandem of cbEGF and other cysteine-rich domains (G2-G3), respectively. Gl encodes the nidogen domain which is where the similarity between INSP220 and NIDO1 and NIDO-2 lies. The NIDO domain is not related to any established domain. G2 consists of an EGF domain and a domain of approximately 230 residues that is not related to any established domain. G3 contains a six-bladed β-propeller related to domains in the LDL receptor and several other proteins. Binding sites for major basement membrane proteins have been mapped to Nidogens: perlecan and collagen IV bind to G2, whereas the laminin γ1 chain binds to G3. There are not any proteins which have been shown to interact with the nidogen domain (Gl) of NIDO-I or NIDO-2.

Other potential domains that INSP220 may interact with are the laminin-type epidermal growth factor-like (EGF-Lam or LE) domain, the laminin G-like (LamG or LG) domain, the fibronectin type I (FN1 or Fl), type II (FN2 or F2), and type III (FN3 or F3) domains, fibrillin domains, follistatin-like and EF hand domains, and von Willebrand factor type A (VWA) domains are found in a number of ECM proteins to name a few.

A screen for compounds which inhibit the interaction of INSP220 and any of these domains could be configured by producing INSP220 in vitro, produce the aforementioned domains in vitro and assay for compounds with agonise or antagonise the interaction. An example of assaying these interactions is found in Ries A, Gohring W, Fox JW, Timpl R, Sasaki T. (Eur J Biochem. 2001 Oct;268(19):5119-28. Recombinant domains of mouse Nidogen-1 and their binding to basement membrane proteins and monoclonal antibodies).

A paper outlining the cloning of the mouse orthologue of INSP220 claims that mouse INSP220 is a novel marker of kidney stroma cells (Leimeister C, Schumacher N, Diez H, Gessler M., Dev Dyn. 2004 Jun;230(2):371-7.). They call this protein Snep (stromal Nidogen extracellular matrix protein). They show that it is additionally expressed in
mesenchymal cells of other embryonic tissues and within the nervous system. Of interest, Snep transcripts are also found at sites of embryonic apoptosis. Furthermore, comparative expression analysis of kidney stroma markers suggests that Snep is expressed in a specific subpopulation of stromal cells and may provide environmental cues to support regular development. Hence we can hypothesise that human SNEP is involved in regulation of development of tissues where it is expressed and may play a role in the apoptosis of those cells.

Example 5: Microarray studies

Custom microarrays have been manufactured using Agilent Technologies' (Agilent Technologies Inc, Palo Alto, CA) non-contact in situ synthesis process of printing 60-mer length oligonucleotide probes, base-by-base, from digital sequence files. This is achieved with an inkjet process which delivers extremely small, accurate volumes (picoliters) of the chemicals to be spotted. Standard phosphoramidite chemistry used in the reactions allows for very high coupling efficiencies to be maintained at each step in the synthesis of the full-length oligonucleotide. Precise quantities are reproducibly deposited "on the fly." This engineering feat is achieved without stopping to make contact with the slide surface and without introducing surface-contact feature anomalies, resulting in consistent spot uniformity and traceability. (Hughes et al. (2001) Nat. Biotech. Apr; 19(4): 342-7. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer).

Probe Synthesis

Methodologies were carried out according to Agilent instructions. Essentially, cDNA synthesis and subsequent T7 polymerase amplification of Cyanine 3(5)-CTP labeled cRNA probe was carried out using Agilent's low RNA input fluorescent linear amplification kit from a template of 5µg of total RNA according to the kit protocol (version 2 August 2003, Agilent, Palo Alto, CA). cRNA is then fragmented using Agilent's In Situ hybridization kit-plus and hybridized both according to Agilent's protocol (Agilent 60-mer oligo microarray processing protocol version 4.1 April 2004, Agilent, Palo Alto, CA).

Microarray Chip Design
10,536 probes are on the array
• 5557 of the probes designed specifically to detect secreted sequences of primary interest
• 1000 probes designed as negative controls
• 500 probes designed as positive controls
• Remainder of the probes were designed to public domain sequences which are known to be either secreted soluble extracellular proteins or membrane bound proteins with an extracellular domain in contact with the extracellular milieu.

Studies specific for INSP220

INSP220 is formed from separate component exons. We intend to profile the chips using probe synthesized from 10 normal tissues, bone marrow, brain, lung, ovary, PBMCs, placenta, prostate, spleen and testis. Expression reports are obtainable on an exon by exon basis.

Averaging is performed for the data, using the One-step Tukey Bi-Weight Algorithm (Data Analysis and Regression: A Second Course in Statistics", Mosteller and Tukey, Addison-Wesley, 1977, pp. 203-209; see also Affymetrix MAS5.0 algorithm). The purpose of this is to define a robust estimate of the average value of a dataset. In this case our datasets will comprise multiple probe expression values for a single exon.

This custom array is useful for a number of reasons. First, it allows the existence and sequence of the transcript to be confirmed. Second, the tissue distribution of the INSP220 polypeptide sequence can be evaluated and thus the role of this polypeptide in disease can be clarified. The array can also be used as a diagnostic tool, to diagnose disease incidence in patients with disease conditions with which this polypeptide is correlated. The use of exon-specific probes allows any variance in expression of splice variants of this polypeptide sequence to be evaluated, in general, in specific tissues and in specific disease states.
CLAIMS

1. A polypeptide, which polypeptide:

   (i) comprises or consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68;

   (ii) is a fragment thereof which comprises at least a fragment of the amino acid sequence as recited in SEQ ID NO:4 or which has an antigenic determinant in common with the polypeptides of SEQ ID NO:2 or SEQ ID NO:4; or

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

2. A polypeptide which is a functional equivalent according to claim 1 (iii), characterised in that it is homologous to the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68.

3. A polypeptide which is a fragment or a functional equivalent as recited in claim 1 or 2, which has greater than 80% sequence identity with the amino acid sequence recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68 or with an active fragment thereof, preferably greater than 85%, 90%, 95%, 98%, 98.5%, 99% or 99.5% sequence identity.

4. A polypeptide which is a functional equivalent as recited in any one of claims 1 to 3, which exhibits significant structural homology with a polypeptide having the amino acid sequence recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66 or SEQ ID NO:68.

5. A polypeptide which is a fragment as recited in claims 1 or 3 having an antigenic determinant in common with the polypeptide of part (i) of claim 1 which consists of or comprises 7 or more amino acid residues from the amino acid sequence recited in SEQ ID NO:2, or SEQ ID NO:4.

6. A fusion protein comprising a polypeptide according to any previous claim.

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

8. A purified nucleic acid molecule according to claim 7, which comprises or consists of the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, or is a redundant equivalent or fragment thereof.

9. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claim 7 or 8.

10. A vector comprising a nucleic acid molecule as recited in any one of claims 7 to 9.

11. A host cell transformed with a vector according to claim 10.

12. A ligand which binds specifically to, and preferably modulates the activity of, a polypeptide according to any one of claims 1 to 6.

13. A ligand according to claim 12, which is an antibody.

14. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1 to 6.

15. A compound according to claim 14 that binds to a polypeptide according to any one of claims 1 to 6 without inducing any of the biological effects of the polypeptide.

16. A compound according to claim 15, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.

17. A polypeptide according to any one of claims 1 to 6, a nucleic acid molecule according to any one of claims 7 to 9, a vector according to claim 10, a host cell according to claim 11, a ligand according to claim 12 or claim 13, or a compound according to any one of claims 14 to 16, for use in therapy or diagnosis of disease.

18. 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 claims 1 to 6, or assessing the activity of a polypeptide according to any one of claims 1 to 6, 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.
19. A method according to claim 18 that is carried out \textit{in vitro}.

20. A method according to claim 18 or claim 19, which comprises the steps of:
   a) contacting a ligand according to claim 12 or claim 13 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and
   b) detecting said complex.

21. A method according to claim 18 or claim 19, 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 7 to 9 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.

22. A method according to claim 18 or claim 19, 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 7 to 9 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.

23. A method according to claim 18 or claim 19 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 7 to 9 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.

24. The method of claim 23, 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.

25. The method of either claim 23 or 24, 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.

26. A method according to any one of claims 18-25, wherein said disease includes, but is not limited to, hereditary deafness, disorders of the development of the kidney, the lungs, the pancreas, salivary gland, upper jaw, lower jaw, the brain and the cerebellum, and diseases involving unregulated complement activation for example, auto-immune diseases including psoriasis, adult/acute respiratory distress syndrome, bullous pemphigoid, rheumatoid arthritis, systemic lupus erythematosus, ischemia-reperfusion injury; renal and glomerular diseases; neurodegenerative, cerebrovascular and neuroinflammatory diseases including cerebral ischemia and trauma, Alzheimer's disease, multiple sclerosis; cardiovascular diseases and other pathological conditions.

27. A pharmaceutical composition comprising a polypeptide according to any one of claims 1 to 6, a nucleic acid molecule according to any one of claims 7 to 9, a vector...
according to claim 10, a host cell according to claim 11, a ligand according to claim 12 or claim 13, or a compound according to any one of claims 14 to 16.

28. A vaccine composition comprising a polypeptide according to any one of claims 1 to 6 or a nucleic acid molecule according to any one of claims 7 to 9.

29. A polypeptide according to any one of claims 1 to 6, a nucleic acid molecule according to any one of claims 7 to 9, a vector according to claim 10, a host cell according to claim 11, a ligand according to claim 12 or claim 13, a compound according to any one of claims 14 to 16, or a pharmaceutical composition according to claim 27, for use in the manufacture of a medicament for the treatment of a certain disease including, but not limited to, hereditary deafness, disorders of the development of the kidney, the lungs, the pancreas, salivary gland, upper jaw, lower jaw, the brain and the cerebellum, and diseases involving unregulated complement activation for example, auto-immune diseases including psoriasis, adult/acute respiratory distress syndrome, bullous pemphigoid, rheumatoid arthritis, systemic lupus erythematosus, ischemia-reperfusion injury; renal and glomular diseases; neurodegenerative, cerebrovascular and neuroinflammatory diseases including cerebral ischemia and trauma, Alzheimer's disease, multiple sclerosis; cardiovascular diseases and other pathological conditions.

30. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1 to 6, a nucleic acid molecule according to any one of claims 7 to 9, a vector according to claim 10, a host cell according to claim 11, a ligand according to claim 12 or claim 13, a compound according to any one of claims 14 to 16, or a pharmaceutical composition according to claim 27.

31. A method according to claim 30, 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, host cell, ligand, compound or composition administered to the patient is an agonist.

32. A method according to claim 30, 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, host cell, ligand, compound or composition administered to the patient is an antagonist.

33. 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 to 6, or the level of expression of a nucleic acid molecule according to any one of claims 7 to 9 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.

34. 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 to 6, or a nucleic acid molecule according to any one of claims 7 to 9 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.

35. 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 7 to 9; 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.

36. The kit of claim 35, further comprising a third container holding an agent for digesting unhybridised RNA.

37. 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 7 to 9.

38. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1 to 6; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

39. 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 to 6.
40. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 39 with a candidate compound and determining the effect of the compound on the disease of the animal.

41. The use of an INSP220 polypeptide as a target for screening candidate drugs for treating or preventing a nidogen/sneg-like molecule related disorder.

42. Method of selecting biologically active compounds comprising:

   (i) contacting a candidate compound with recombinant host cells expressing an INSP220 polypeptide;

   (ii) selecting compounds that bind said INSP220 polypeptide at the surface of said cells and/or that modulate the activity of the INSP220 polypeptide.
Figure 1

Key:
Boxed  Signal peptide
Bold  Nidogen-like (NIDO) domain
Grey  Calcium-binding EGF-like (EGF_CA) domain
Italic & Bold  Complement control protein (CCP) domain
Underlined  Fibronectin type 3 (FN3) domain

[Sequence of amino acids]
**Figure 2**

Query: INSPI220.pep  
(1352 letters)

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>ID</th>
<th>Accession</th>
<th>Description</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ref</td>
<td>XP_059482.4</td>
<td>PREDICTED: FLJ00133 protein [Homo sapiens]</td>
<td>2748</td>
<td>0.0</td>
</tr>
<tr>
<td>dbj</td>
<td>BAB84901.1</td>
<td>FLJ00133 protein [Homo sapiens]</td>
<td>2627</td>
<td>0.0</td>
</tr>
<tr>
<td>ref</td>
<td>NP_766051.3</td>
<td>secreted Nidogen domain protein [Mus musculus]</td>
<td>2505</td>
<td>0.0</td>
</tr>
<tr>
<td>gb</td>
<td>AAO41836.1</td>
<td>secreted protein SST3 [Mus musculus]</td>
<td>2044</td>
<td>0.0</td>
</tr>
<tr>
<td>ref</td>
<td>XP_237415.2</td>
<td>similar to secreted protein SST3 [Rattus n...</td>
<td>1942</td>
<td>0.0</td>
</tr>
<tr>
<td>gb</td>
<td>AAQ04556.1</td>
<td>insulin responsive sequence DNA binding prot...</td>
<td>1734</td>
<td>0.0</td>
</tr>
<tr>
<td>ref</td>
<td>XP_516203.1</td>
<td>PREDICTED: similar to FLJ00133 protein [Pa...</td>
<td>1610</td>
<td>0.0</td>
</tr>
<tr>
<td>gb</td>
<td>AAM27393.1</td>
<td>SNED1 protein [Homo sapiens]</td>
<td>1159</td>
<td>0.0</td>
</tr>
<tr>
<td>sp</td>
<td>P10079</td>
<td>FBPI STFPU fibropellin I precursor (Epidermal gro...</td>
<td>1008</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>514 e-144</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 3

>ref|NP 766051.3| secreted Nidogen domain protein [Mus musculus]
amb|CA848492.1| secreted Nidogen domain protein precursor [Mus musculus]

Length = 1403

Score = 2505 bits (6492), Expect = 0.0
Identities = 1143/1373 (83%), Positives = 1225/1373 (90%), Gaps = 32/1373 (2%)

Query: 1
MRHGVAWALLVAAALGLGARGVGAVALADFYFGAGERGAVTDQDDGSSGLLRLSVPEF 60
MR G AWALLA*AAALGL RGVR AVALADFYPFG +RGD VPQKDDGSGSL+PLSVPEF

Subject: 1
PFFCGAESGLVNNNIGISFLKEVSQPTFAPFIAFDKCVVAAFWAVDNNRAGDYVYRE 120
PFFCGAESGLVNNNIGISFLKEVSQPTFAPFIAFDKCVVAAFWAVDNNRAGDYVYRE

Query: 121
ATDPAMLRRATEDVRHYPFEDLFNNNATWVFATWYRVTFFGSSSSSFVNTQTVILTDGK 180
ATDPAML RRATED+R YFPEL D+ATWVFATWYRVTFFGSSSSSFVNTQTVILTDGK+

Subject: 121
LSFTNPYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE 240
LSFTNPYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE

Query: 181
SGFTGRCHLVDNEASQPCQNGTGCTCTHGINISFRCSQCPAGPFGPTCETASQPSFCDTEKOCQH 360
SGFTGRCHLVDNEASQPCQNGTGCTCTHGINISFRCSQCPAGPFGPTCETASQPSFCDTEKOCQH

Subject: 181
ASFTPNYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE 240
ASFTPNYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE

Query: 241
TTTNVVGQFWAFRIDDADQVTQVGCQCTTSCVAALRFPLCNGLKICDDCVTGNPSYTSCTSL 300
TTTNVVGQFWAFRIDDADQVTQVGCQCTTSCVAALRFPLCNGLKICDDCVTGNPSYTSCTSL

Subject: 241
SFTINYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE 240
SFTINYESIYWTGTHASSGGNANGLSGIQIAAQAFGNAQDGGGRYPFPASPSGRTADMAEVE

Query: 301
GGCQCGQGGGAGAVGQAGYTGACEMDVMDDCSDFCPLNNGGCSVCDLVGYTCLEAFEPKGSL 420
GGCQCGQGGGAGAVGQAGYTGACEMDVMDDCSDFCPLNNGGCSVCDLVGYTCLEAFEPKGSL

Subject: 301
GGCQCGQGGGAGAVGQAGYTGACEMDVMDDCSDFCPLNNGGCSVCDLVGYTCLEAFEPKGSL 420
GGCQCGQGGGAGAVGQAGYTGACEMDVMDDCSDFCPLNNGGCSVCDLVGYTCLEAFEPKGSL
SEQWIN 99, version 1.02

DNA Homo sapiens

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>atgcggcagc</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
</tr>
<tr>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
</tr>
<tr>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
</tr>
<tr>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
<td>gcgtcgcctg</td>
</tr>
</tbody>
</table>

Protein Homo sapiens

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Arg</td>
<td>His</td>
<td>Gly</td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>Leu</td>
<td>Gln</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Ala</td>
<td>Asp</td>
<td>Phe</td>
<td>Tyr</td>
</tr>
<tr>
<td>Pro</td>
<td>Phe</td>
<td>Gln</td>
<td>Val</td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>Asp</td>
<td>Ala</td>
</tr>
<tr>
<td>Val</td>
<td>Thr</td>
<td>Pro</td>
<td>Lys</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>Arg</td>
<td>Pro</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td>Val</td>
<td>Pro</td>
<td>Phe</td>
<td>Pro</td>
</tr>
<tr>
<td>Pro</td>
<td>Phe</td>
<td>Phe</td>
<td>Gln</td>
</tr>
<tr>
<td>Ala</td>
<td>Gly</td>
<td>His</td>
<td>Ser</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>Tyr</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Phe</td>
<td>Tyr</td>
<td>Pro</td>
<td>Phe</td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>Gly</td>
<td>Gln</td>
<td>Asp</td>
<td>Gln</td>
</tr>
</tbody>
</table>

Ala VaI Ala Leu Ala Asp Phe Tyr Pro Phe Gln Ala Gln Arg Gln Asp
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>VaI</td>
<td>Thr</td>
<td>Pro</td>
</tr>
<tr>
<td>Lys</td>
<td>GIn</td>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td>GIy</td>
<td>GIy</td>
<td>Ser</td>
<td>GIy</td>
</tr>
<tr>
<td>Leu</td>
<td>Arg</td>
<td>Pro</td>
<td>Leu</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>VaI</td>
<td>Pro</td>
<td>Phe</td>
</tr>
<tr>
<td>Phe</td>
<td>Pro</td>
<td>Phe</td>
<td>GIy</td>
</tr>
<tr>
<td>Ala</td>
<td>GIu</td>
<td>His</td>
<td>Ser</td>
</tr>
<tr>
<td>GIu</td>
<td>GIu</td>
<td>GIe</td>
<td>Leu</td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DNA**

```
<210>  5
<211>  288
<212>  DNA
<213>  Homo sapiens
```

```
>210
gtgacaaca cccgcatct ctcctctctg aaggggttt tcaqtcac cccagtggcc
60
tccccattg ccagaggcc ctgcgtgttg gcaqctcctt ggcqagatgt ggacacccq
120
cgcagccgc acctgtacta cccgagccag ccagqccag ccatgtgcgg cgcagccag
180
gacgctca gcgtactttt cccgagctc atgcacctgt ggttttttgtt
240
gccacctcg acctgtgga cttctttgga ggcagttcct catccctt
288
```

```
<210>  6
<211>  96
<212>  PRT
<213>  Homo sapiens
```

```
>212
<210>
<213>
```

```
<212>
PRT
Homo sapiens
<400> 8
```

```
Val Asn Asn Gly H e H e Ser Phe Leu Lys GIu Val Ser GIe Phe
```

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr</td>
<td>Pro</td>
<td>VaI</td>
<td>Ala</td>
</tr>
<tr>
<td>Phe</td>
<td>Pro</td>
<td>H e</td>
<td>Ala</td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Arg</td>
<td>Cys</td>
</tr>
<tr>
<td>Val</td>
<td>VaI</td>
<td>VaI</td>
<td>Ala</td>
</tr>
<tr>
<td>Ala</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

```
<210>  6
<211>  47
<212>  DNA
<213>  Homo sapiens
```

```
<210>  7
<211>  141
<212>  DNA
<213>  Homo sapiens
```

```
<212>
PRT
Homo sapiens
```

```
<210>  8
<211>  47
<212>  PRT
<213>  Homo sapiens
```

```
<210>  8
<211>  47
<212>  PRT
<213>  Homo sapiens
```

```
Val Asn Thr Phe GIIn Thr Val Leu H e Thr Asp Gly Lys Leu Ser Phe
```

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Val</td>
<td>Asn</td>
<td>Thr</td>
<td>Phe</td>
</tr>
<tr>
<td>GIIn</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>H e</td>
<td>Thr</td>
<td>Asp</td>
<td>Gly</td>
</tr>
<tr>
<td>Lys</td>
<td>Leu</td>
<td>Ser</td>
<td>Phe</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>
Thr  H  e  Phe  Asn  Tyr  Glu  Ser  H  e  Val  Trp  Thr  Thr  Glu  Thr  His  Ala
20     25     30
Ser  Ser  Gly  Gly  Asn  Ala  Thr  Glu  Leu  Glu  Glu  H  e  Ala  Ala  Gln
35     40     45

DNA
Homo sapiens

<210> 9
<211> 163
<213> Homo sapiens
<400> 9
gctggcttca agcgaggcga tgggcagcgt tacttcagta tccccggctc gcgcacagca  60
gacatggccg aggtgagac caccaccaac gtgggtgtgc cccggcgtgc gcgcttcaga 120
atcgatgtag cccagtgccg cgtggggggc tgcggccata caa  163

<210> 10
<211> 55
<212> PRT
<213> Homo sapiens
<400> 10
Ala  Gly  Phe  Asn  Ala  Gly  Asp  Gly  Gln  Arg  Tyr  Phe  Ser  H  e  Pro  Gly
1       5       10       15
Ser  Arg  Thr  Ala  Asp  Met  Ala  Glu  Glu  Thr  Thr  Thr  Asn  Val  Glu
20     25     30
VaI  Pro  Glu  Arg  Trp  Ala  Phe  Arg  H  e  Asp  Ala  Gln  VaI  Arg  VaI
35     40     45
Gly  Glu  Cys  Gly  His  Thr  Thr
50     55

<210> 11
<211> 126
<212> DNA
<213> Homo sapiens
<400> 11
cg tgtccgttg cctgcccttg ccgccctgcc tcaacggcgg caagtgcacg gacgactgcg  60
tcacggcga ccctctctac acctgtctct gctctccgg ctctcagggc cggaggtgcc 120
acctgg  

<210> 12
<211> 42
<212> PRT
<213> Homo sapiens

<400> 12
Ser  Val  Cys  Leu  Ala  Leu  Arg  Pro  Cys  Leu  Asn  Gly  Gly  Lys  Cys  H  e
1       5       10       15
Asp  Asp  Cys  Val  Thr  Glu  Asn  Pro  Ser  Tyr  Thr  Cys  Ser  Cys  Leu  Ser
20     25     30
Gly  Phe  Thr  Glu  Arg  Arg  Cys  His  Leu  Asp
35     40

<210> 13
<211> 114
<212> DNA
<213> Homo sapiens
<400> 13
acgtgaacga atgtgcctcc cagccctgtc agaatggtgg gacctgtact cacggcatca 60
acagtttccg ctgccagcgg cggcagcgcag atgggggcc cacctgtgag acag 114
<210> 14
<211> DNA
<212> PRT
<213> Homo sapiens

VaI Asn GIu Cys Ala Ser GIy Pro Cys GIIn Asn GIy GIy Thr Cys Thr
1 5 10 15
His GIy lie Asn Ser Phe Arg Cys GIy Cys Pro Ala GIy Phe GIy GIy
20 25 30
Pro Thr Cys GIu Thr Ala
35
<210> 15
<211> DNA
<212> PRT
<213> Homo sapiens

GIn Ser Pro Cys Thr Lys GIu Cys GIIn His GIy GIy GIy GIy Cys GIy
1 5 10 15
VaI GIu Asn GIy Ser Ala VaI Cys VaI Cys GIu Ala GIy Tyr Thr GIy
20 25 30
Ala Ala Cys GIu Met Asp
35
<210> 17
<211> DNA
<212> PRT
<213> Homo sapiens

VaI Asp Asp Cys Ser Pro Asp Pro Cys Leu Asn GIy GIy Ser Cys VaI
1 5 10 15
Asp Leu Val Gly Asn Tyr Thr Cys Leu Cys Ala Glu Pro Phe Lys Gly
20 25 30

Leu Arg Cys Glu Thr Gly
35

<210> 19
<211> 126
<212> DNA
<213> Homo sapiens

<400> 19
gagacatcc agtgccagac gcctgcctct ccacaatggg gcacactgtg 60
ctgtgcaggt gcctgcaagg ctctcatgggc ctggactgca 120
gggaga 126

<210> 20
<211> 42
<212> PRT
<213> Homo sapiens

<400> 20
Asp His Pro Val Pro Asp Ala Cys Leu Ser Ala Pro Cys His Asn Gly
1 5 10 15
Gly Thr Cys Val Asp Ala Asp Gin Gly Tyr Val Cys Gly Cys Pro Gly
20 25 30
Gly Phe Met Gly Leu Asp Cys Arg Glu Arg
35 40

<210> 21
<211> 105
<212> DNA
<213> Homo sapiens

<400> 21
gagtccccga tgtgctgag tcgcgcaacg gagccagatg cctgggcccc aacaccaccc 60
tctgccagtg cccccctgga ttcctctggc ttcctgtgca atttg 105

<210> 22
<211> 35
<212> PRT
<213> Homo sapiens

<400> 22
VaI Pro Asp Asp Cys Glu Cys Arg Asn Gly GIy Arg Cys Leu GIy Ala
1 5 10 15
Asn Thr Thr Leu Cys Gin Cys Pro Leu GIy Phe Phe GIy Leu Leu Cys
20 25 30
Gly Phe GIu
35

<210> 23
<211> 114
<212> DNA
<213> Homo sapiens

<400> 23
aaatcacacg ctgcctgc aacatgaaca cacatgcggg gatgagggg tacgcaatgg 60
agcaggccc gagctacctc tcggtctggc aacccgaca caatgccagc cact 114
<210>  29
<211>  117
<212> DNA
<213> Homo sapiens

<400>  29
ggaagccaga ctcgtgtgcc tctggcccct gtcacaacgg cggcacctgc ttccactaca  60
ttggcaaata caagtgtgac tgtccccccag gcttctcggc gcggcacgct gagatag  117

<210>  30
<211>  39
<212> PRT
<213> Homo sapiens

<400>  30
Lys Pro Asp Ser Cys Ala Ser GIy Pro Cys His Asn GIy GIy Thr Cys  1  5  10  15
Phe His Tyr lle GIy Lys Tyr Lys Cys Asp Cys Pro Pro GIy Phe Ser  20  25  30
GIy Arg His Cys GIu lie Ala  35

<210>  31
<211>  114
<212> DNA
<213> Homo sapiens

<400>  31
ccccctcccc ctcgtcccg gacccggttg tgaatggggg cacctgcaag gacccggaca  60
cggatttctt ctgccactgc caagcagggt acatgggacg cccgtgccag gcag  114

<210>  32
<211>  38
<212> PRT
<213> Homo sapiens

<400>  32
Pro Ser Pro Cys Phe Arg Ser Pro Cys Val Asn GIy GIy Thr Cys GIu  1  5  10  15
Asp Arg Asp Thr Asp Phe Phe Cys His Cys GIIn Ala GIy Tyr Met GIy  20  25  30
Arg Arg Cys GIIn Ala GIu  35

<210>  33
<211>  174
<212> DNA
<213> Homo sapiens

<400>  33
aggtgactg ogggcccccg gaggaggtga agcaaqgncac actgacgttc aacggcaacgc  60
ggctgggcgc ggtggccctg taticatgttg acctgctgca cacgctgacg gccccgacgc  120
gcatccggtg otgcccccca cacggtgtct gaggagccgc tccccagctgc ctgtg  174

<210>  34
<211>  58
<212> PRT
<213> Homo sapiens
Val Asp Cys GIy Pro Pro GIu GIu Val Lys His Ala Thr Leu Arg Phe
1 5 10 15
Asn GIy Thr Arg Leu GIy Ala Val Ala Leu Tyr Ala Cys Asp Arg GIy
20 25 30
Tyr Ser Leu Ser Ala Pro Ser Arg lie Arg VaI Cys Gin Pro His GIy
35 40 45
Val Trp Ser GIu Pro Pro Gin Cys Leu GIu
50 55

<210> 35
<211> 114
<212> DNA
<213> Homo sapiens

aaatcgtga gtgcqgtct cagccgtgcc tgcatgqggq ctcttgctag gaccqcgqtg 60
cgtggtacct gtgcctctgc agcacagqct atgaggqgc ccactgtgag otgg 114

<210> 36
<211> 38
<212> PRT
<213> Homo sapiens

lie Asp GIu Cys Arg Ser Gin Pro Cys Leu His GIy GIy Ser Cys Gin
1 5 10 15
Asp Arg VaI Ala GIy Tyr Leu Cys Leu Cys Ser Thr GIy Tyr GIu GIy
20 25 30
Ala His Cys GIu Leu GIu
35

<210> 37
<211> 114
<212> DNA
<213> Homo sapiens

agagggatga gtgcqgagct cacccgtqca gaaatgqagg gtcctgqagg aacctcccag 60
gggctatgt ctgcqggtgc cctgcagqct tcggtggagt ccactgtgag acag 114

<210> 38
<211> 38
<212> PRT
<213> Homo sapiens

<400> 38
Arg Asp GIu Cys Arg Ala His Pro Cys Arg Asn GIy GIy Ser Cys Arg
1 5 10 15
Asn Leu Pro GIy Ala Tyr VaI Cys Arg Cys Pro Ala GIy Phe VaI GIy
20 25 30
VaI His Cys GIu Thr GIu
35

<210> 39
<211> 114
<212> DNA
<213> Homo sapiens

<400> 39
aggtggacgc ctgcgactcc agcccctgcc agcatggagg ccggtgtgag agcggcggcg 60
gggctcactt gtgcgtctgc ccagagagtct ttctgggctt ccactgcgag acag 114

<210> 40
<211> 38
<212> PRT
<213> Homo sapiens

<400> 40
VaI Asp Ala Cys Asp Ser Ser Pro Cys Gln His GIy GIy Arg Cys GIu 1 5 10 15
Ser GIy GIy GIy Ala Tyr Leu Cys VaI Cys Pro GIu Ser Phe Phe GIy 20 25 30
Tyr His Cys GIu Thr VaI 35

<210> 41
<211> 9 9
<212> DNA
<213> Homo sapiens

<400> 41
tgagtgaccc ctgcttctcc agccccctgtg ggggccgttg ctattgcttg gccagcaacg 60
gctccccacag ctgcacctgc aagtgccagtg aaagggccga ggaactgcgccc aaag 114

<210> 42
<211> 38
<212> PRT
<213> Homo sapiens

<400> 42
Ser Asp Pro Cys Phe Ser Ser Ser Pro Cys GIy GIy Arg GIy Tyr Cys Leu 1 5 10 15
Ala Ser Asn GIy Ser His Ser Cys Thr Cys Lys VaI GIy Tyr Thr GIy 20 25 30
GIu Asp Cys Ala Lys GIu 35

<210> 43
<211> 297
<212> DNA
<213> Homo sapiens

<400> 43
agctctttccc accgacggcc cttcaagatgg agagagtggg ggagagtggg gtctctatct 60
cctggaacc gcacatgtgt ccagccgcca ggcagatgct tgctggctac gcggtcacct 120
acgtctttcc cagccggtccc taccgccgca cagacattgt ggacaggacc gctctctgctc 180
accagcttcc gggctggcca ggggcaaggg cctacaacct cttccgcttcc tcagtgaagc 240
gaaacagtta caacaagaat gacatcagca ggctgctgctt gctgtgctgccc cgcagcc 297

<210> 44
<211> 99
<212> PRT
<213> Homo sapiens
<400>  44
Leu Phe Pro Pro Thr Ala Leu Lys Met GIu Arg VaI GIu GIu Ser GIy
 1    5    10    15
VaI Ser H e Ser Trp Asn Pro Pro Asn GIy Pro Ala Ala Arg GIn Met
 20   25   30
Leu Asp GIy Tyr Ala VaI Thr Tyr VaI Ser Ser Asp GIy Ser Tyr Arg
 35   40   45
Arg Thr Asp Phe VaI Asp Arg Thr Arg Ser Ser His GIn Leu GIn Ala
 50   55   60
Leu Ala Ala GIy Arg Ala Tyr Asn lie Ser VaI Phe Ser VaI Lys Arg
 65   70   75   80
Asn Ser Asn Asn Lys Asp H e Ser Arg Pro Ala VaI Leu Leu Ala
 85   90   95
Arg Thr Arg

<210>  45
<211>  184
<212>  DNA
<213>  Homo sapiens

<400>  45
gacccgcccc tgtgaagagc ttcgaggtca ccaatgtgac ggctagcacc atctcagtgc
360
agttggccct gcacaggatc cgccatgcca ccgtcagtgg ggtccgtgtg tccatccgcc
120
acccctggag cctcagggac caggccaccg atgtggacag gatggacag aggttacacct
180
ttag
<210>  46
<211>  61
<212>  PRT
<213>  Homo sapiens

<400>  46
Pro Arg Pro VaI GIu GIy Phe GIu VaI Thr Asn VaI Thr Ala Lea Ser Thr
 1    5    10    15
H e Ser VaI GIn Trp Ala Leu His Arg H e Arg His Ala Thr VaI Ser
 20   25   30
GIy VaI Arg VaI Ser H e Arg His Pro GIn Ala Leu Arg Asp GIn Ala
 35   40   45
Thr Asp VaI Asp Arg Ser VaI Asp Arg Phe Thr Phe Arg
 50   55   60

<210>  47
<211>  113
<212>  DNA
<213>  Homo sapiens

<400>  47
ggccctgctg cctggaaga ggtacacccat ccagctgacg acccctcagt ggtcagggg
60
agaggacac cccacagaga gcctggccac gcgccgcacg cagtggtgga ccc
113
<210>  48
<211>  38
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;212&gt;</td>
<td></td>
<td>PRT</td>
</tr>
<tr>
<td>&lt;213&gt;</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>&lt;400&gt;</td>
<td>48</td>
<td>Ala Leu Leu Pro Gly Lys Arg Tyr Thr He GIn Leu Thr Thr Leu Ser</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>GIn Leu Arg GIn GIn His Pro Thr GIn Ser Leu Ala Thr Ala Pro</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>Thr His VaI Trp Thr Arg</td>
</tr>
<tr>
<td>&lt;210&gt;</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>&lt;211&gt;</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>&lt;212&gt;</td>
<td></td>
<td>DNA</td>
</tr>
<tr>
<td>&lt;213&gt;</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>&lt;400&gt;</td>
<td>49</td>
<td>ggcggccccgg ccagcagaaa cttgacgccg ccagtccttg gccccaggtgg</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>tctggaggtgc ccagaatcctg tggagcctta tctcatcaat gtcgaccacca</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>ggccagacca ccagagagcc gcccaccctct acatcatcct ct</td>
</tr>
<tr>
<td>&lt;210&gt;</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>&lt;211&gt;</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>&lt;212&gt;</td>
<td></td>
<td>PRT</td>
</tr>
<tr>
<td>&lt;213&gt;</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>&lt;400&gt;</td>
<td>50</td>
<td>Pro Leu Pro Pro Ala Asn Leu Thr Ala Ala Arg VaI Thr Ala Thr Ser</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>Ala Ala VaI VaI Trp Asp Ala Pro Thr Pro GIn Ser Leu Leu GIn Ala</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>&lt;210&gt;</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>&lt;211&gt;</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>&lt;212&gt;</td>
<td></td>
<td>DNA</td>
</tr>
<tr>
<td>&lt;213&gt;</td>
<td></td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>&lt;400&gt;</td>
<td>51</td>
<td>cccccagaaa tgtgcgtgac agagctggc accaggagag acacccactt cgggtgcctca</td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>agacaagaccc gccccgccgc ccgctgcgcgg agctgcgtcag gtcaatgac cacagcggcc</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>cccgagaccc caccagccgc ccag</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;212&gt;</td>
<td></td>
<td>PRT</td>
</tr>
<tr>
<td>&lt;213&gt;</td>
<td></td>
<td>Homo sapiens</td>
</tr>
</tbody>
</table>

```plaintext
Homo sapiens
```

```plaintext
DNA
```
Pro Arg Asp GIy Ala Asp Arg Arg Trp His GIy GIy His His Pro
1 5 10 15
Arg VaI Leu Lys Asn Arg Pro Pro Pro Ala Arg Leu Pro GIu Leu Arg
20 25 30
Leu Leu Asn His Ser Ala Pro GIu Thr Pro Thr GIu Pro Pro Arg
35 40 45

DNA Homo sapiens
 gtctcggag ctgtggacg gcagaggaag agttcgcgg gctcacc cagacgcc accgtgagat cac
 60 83

Phe Ser GIu Leu VaI Asp GIy Arg GIy Arg VaI Ser Ala Arg Phe GIy
1 5 10 15
GIy Ser Pro Ser Lys Ala Ala Thr VaI Arg Ser His
20 25

DNA Homo sapiens
 acgtccctgg caactgttca gaaaccctt gtccagacgg aggcttgtt gtgcgggccg
 60

cagcgccttg caagtggtac tgcgggccag gttcaaggg cagcgcgtgc gacgtc
gttcgggag
 117

VaI Pro GIy Asn Cys Ser GIu Asn Pro Cys GIu Asn GIy GIy Thr Cys
1 5 10 15
VaI Pro GIy Ala Asp Ala His Ser Cys Asp Cys GIy Pro GIy Phe Lys
20 25 30
GIy Arg Arg Cys GIu Leu Ala
35

DNA Homo sapiens
 57
cctgtataaa ggtgcgcagc cctgcacaa ggctgttctc cagacaaag gcctttccag 60
tctgggaggg aggcgtctgt caccacgt 88

<210> 58
<211> 29
<212> PRT
<213> Homo sapiens

<400> 58
Cys Lys Val Ser Arg Pro Cys Thr Arg Leu Phe Ser GIu Thr Lys
1  5  10  15
Ala Phe Pro Val Trp GIu GIy GIy Val Cys His His Val
20  25

<210> 59
<211> 37
<212> DNA
<213> Homo sapiens

<400> 59
gaaacaaagt aagagtcaga cactggagaa atctttaa 37

<210> 60
<211> 11
<212> DNA
<213> Homo sapiens

<400> 60
Lys GIin Ser Lys Ser Gin Thr Leu GIu Lys Ser
1  5  10

<210> 61
<211> 4059
<212> DNA
<213> Homo sapiens

<400> 61
atgcgcacag gctgcgctgt gcgcgtcgcct gcggcgcctg cagccctgctg 40

<210> 62
<211> 120
<212> DNA
<213> Homo sapiens

<400> 62
gggtgcgcag gcggcgccag tcctgggaggg aggcgtctgt caccacgt 120

gcccgtcaggt tctgcgacaa ccggcgccag tcggcgcacaa 300

gggggggcct gcggcgcctg gcggcgcctg gcggcgcctg 1080

gtggggcggcg ccgctgactg cagcaggggag actgggcttc gcggccgaca 1320

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 1560

tggggcggcg ccgctgactg cagcaggggag actgggcttc gcggccgaca 1800

ggctggactg tgaagcctct cggcgggctg gccggcggcc gcggcgggga 2160

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 2520

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 2880

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 3240

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 3600

tgctggtgag ggcggcggcc cggcgggctg gccggcggcc gcggcgggga 4059
tttgaaatca cagccatgcc ctgcaacatg aacacacagt gcccagatgg gggctactgc
atggagcacg gcgggagcta cctctgcgtc ... Arg Pro Leu Ser VaI Pro Phe Pro Phe GIy
Ala GIu His Ser GIy Leu Tyr VaI Asn Asn Asn GIy lie lie Ser Phe
Leu Lys Glu Val Ser Gin Phe Thr Pro Val Ala Phe Pro He Ala Lys
65  70  75  80
Asp Arg Cys Val Val Ala Ala Phe Trp Ala Asp Val Asp Asn Arg Arg
100
Ala Glu Asp Val Tyr Tyr Arg Glu Ala Thr Asp Pro Ala Met Leu Arg
115
Arg Ala Thr Glu Asp Val Arg His Tyr Phe Pro Glu Leu Leu Asp Phe
130  135  140
Asn Ala Thr Trp Val Phe Val Ala Thr Trp Tyr Arg Val Thr Phe Phe
145  150  155  160
Gly Gly Ser Ser Ser Pro Val Asn Thr Phe Gin Thr Val Leu He
165  170  175
Thr Asp Gly Lys Leu Ser Phe Thr He Phe Asn Tyr Glu Ser He Val
180  185  190
Trp Thr Thr Gly Thr His Ala Ser Ser Gly Gin Ala Thr Gly Leu
195  200  205
Gly Gly He Ala Ala Gin Ala Gin Phe Asn Ala Gly Asp Gly Gin Arg
210  215  220
Tyr Phe Ser He Pro Gin Ser Arg Thr Ala Asp Met Ala Gin Val Gin
225  230  235  240
Thr Thr Thr Asn Val Gin Val Pro Gin Arg Trp Ala Phe Arg He Asp
245  250  255
Asp Ala Gin Val Arg Val Gin Gly Cys Gly His Thr Thr Ser Val Cys
260  265  270
Leu Ala Leu Arg Pro Cys Leu Asn Gin Gly Lys Cys He Asp Gin Cys
275  280  285
Val Gin Asn Gin Ser Tyr Thr Cys Ser Cys Leu Ser Gin Phe Thr
290  295  300
Gly Gin Arg Gin His Leu Gin Gin Gin Cys Leu Gin Ser Gin Gin Gin
305  310  315  320
Gin Gin Gin Gly Thr Cys Thr Gin He Gin Gin Asn Gin Gin Gin Gin
325  330  335
Cys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Ser Gin Gin Gin Gin
340  345  350
Asp Thr Lys Gin His Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
355  360  365
Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
370  375  380
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
385  390  395  400
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
410  415  420
Ser Cys Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
435  440  445

405  Phe Lys GIy Leu Arg Cys GIu Thr GIy Asp His Pro VaI Pro Asp Ala
410  420
415  Cys Leu Ser Ala Pro Cys His Asn GIy GIy Thr Cys VaI Asp Ala Asp
420  435
430  GIn GIy Tyr VaI Cys GIu Cys Pro GIu GIy Phe Met GIy Leu Asp Cys
440  455
450  Arg GIu Arg VaI Pro Asp Asp Cys GIu Cys Arg Asn GIy GIy Arg Cys
460  475
470  Leu GIy Ala Asn Thr Thr Leu Cys GIn Pro Leu GIy Phe Phe GIy
480  495
490  Leu Leu Cys GIu Phe GIu He Thr Ala Met Pro Cys Asn Met Asn Thr
500  510
505  GIn Cys Pro Asp GIy GIy Tyr Cys Met GIu His GIy GIy Ser Tyr Leu
510  525
520  Cys VaI Cys His Thr Asp His Asn Ala Ser His Ser Leu Pro Ser Pro
530  540
535  Cys Asp Ser Asp Pro Cys Phe Asn GIy GIy Ser Cys Asp Ala His Asp
540  555
550  Asp Ser Tyr Thr Cys GIu Cys Pro Arg GIy Phe His GIy Lys His Cys
560  575
570  GIu Lys Ala Arg Pro His Leu Cys Ser Ser GIy Pro Cys Arg Asn GIy
580  590
585  GIy Thr Cys Lys GIu Ala GIy GIy GIu Tyr His Cys Ser Cys Pro Tyr
590  605
600  Arg Phe Thr GIy Arg His Cys GIu le GIy Lys Pro Asp Ser Cys Ala
610  620
615  Ser GIy Pro Cys His Asn GIy GIy Thr Cys Phe His Tyr le GIy Lys
620  635
630  Tyr Lys Cys Asp Cys Pro Pro GIy Phe Ser GIy Arg His Cys GIu He
640  655
650  Ala Pro Ser Pro Cys Phe Arg Ser Pro Cys VaI Asn GIy GIy Thr Cys
660  670
665  GIu Asp Arg Asp Thr Asp Phe Phe Cys His Cys GIu Ala GIy Tyr Met
670  685
675  GIy Arg Arg Cys GIu Ala GIu VaI Asp Cys GIy Pro Pro GIu GIu VaI
690  705
695  Lys His Ala Thr Leu Arg Phe Asn GIy Thr Arg Leu GIy Ala VaI Ala
705  720
710  Leu Tyr Ala Cys Asp Arg GIy Tyr Ser Leu Ser Ala Pro Ser Arg He
720  735
730  Arg VaI Cys GIu Pro His GIy VaI Trp Ser GIu Pro Pro GIu Cys Leu
<table>
<thead>
<tr>
<th>740</th>
<th>745</th>
<th>750</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIu lie Asp GIu Cys Arg Ser GIn Pro Cys Leu His GIy GIy Ser Cys</td>
<td>755</td>
<td>760</td>
</tr>
<tr>
<td>GIin Asp Arg VaI Ala GIy Tyr Leu Cys Leu Cys Ser Thr GIy Tyr GIu</td>
<td>770</td>
<td>775</td>
</tr>
<tr>
<td>GIy Ala His Cys GIu Leu GIu Arg Asp GIu Cys Arg Ala His Pro Cys</td>
<td>785</td>
<td>790</td>
</tr>
<tr>
<td>Arg Asn GIy GIy Ser Cys Arg Asn Leu Pro GIy Ala Tyr VaI Cys Arg</td>
<td>805</td>
<td>810</td>
</tr>
<tr>
<td>Cys Pro Ala GIy Phe VaI GIy VaI His Cys GIu Thr GIu VaI Asp Ala</td>
<td>820</td>
<td>825</td>
</tr>
<tr>
<td>Cys Asp Ser Ser Pro Cys GIin His GIy GIy Arg Cys GIu Ser GIy GIy</td>
<td>835</td>
<td>840</td>
</tr>
<tr>
<td>GIy Ala Tyr Leu Cys VaI Cys Pro GIu Ser Phe Phe GIy Tyr His Cys</td>
<td>850</td>
<td>855</td>
</tr>
<tr>
<td>GIu Thr VaI Ser Asp Pro Cys Phe Ser Ser Pro Cys GIy GIy Arg GIy</td>
<td>865</td>
<td>870</td>
</tr>
<tr>
<td>Tyr Cys Leu Ala Ser Asn GIy Ser His Ser Cys Thr Cys Lys VaI GIy</td>
<td>885</td>
<td>890</td>
</tr>
<tr>
<td>Tyr Thr GIu GIy Asp Cys Ala Lys GIu Leu Phe Pro Pro Thr Ala Leu</td>
<td>900</td>
<td>905</td>
</tr>
<tr>
<td>Lys Met GIu Arg VaI GIu GIu Ser GIy VaI Ser lie Ser Trp Asn Pro</td>
<td>915</td>
<td>920</td>
</tr>
<tr>
<td>Pro Asn GIy Pro Ala Ala Arg GIn Met Leu Asp GIy Tyr Ala VaI Thr</td>
<td>930</td>
<td>935</td>
</tr>
<tr>
<td>Tyr VaI Ser Ser Asp GIy Ser Tyr Arg Arg Thr Asp Phe VaI Asp Arg</td>
<td>945</td>
<td>950</td>
</tr>
<tr>
<td>Thr Arg Ser Ser His GIn Leu GIn Ala Leu Ala Ala GIy Arg Ala Tyr</td>
<td>965</td>
<td>970</td>
</tr>
<tr>
<td>Asn H e Ser VaI Phe Ser VaI Lys Arg Asn Ser Asn Asn Lys Asn Asp</td>
<td>890</td>
<td>985</td>
</tr>
<tr>
<td>lie Ser Arg Pro Ala VaI Leu Leu Ala Arg Thr Arg Pro Arg Pro VaI</td>
<td>995</td>
<td>1000</td>
</tr>
<tr>
<td>GIu GIy Phe GIu VaI Thr Asn VaI Thr Thr He e Ser VaI GIn</td>
<td>1015</td>
<td>1020</td>
</tr>
<tr>
<td>Trp Ala Leu His Arg H e Arg His Ala Thr VaI Ser GIy VaI Arg VaI</td>
<td>1025</td>
<td>1030</td>
</tr>
<tr>
<td>Ser H e Arg His Pro GIu Ala Leu Arg Asp GIn Ala Thr Asp VaI Asp</td>
<td>1045</td>
<td>1050</td>
</tr>
<tr>
<td>Arg Ser VaI Asp Arg Phe Thr Phe Arg Ala Leu Leu Pro GIy Lys Arg</td>
<td>1060</td>
<td>1065</td>
</tr>
<tr>
<td>Tyr Thr H e GIn Leu Thr Thr Leu Ser GIy Leu Arg GIy GIu GIu His</td>
<td>1070</td>
<td>1075</td>
</tr>
</tbody>
</table>
Pro Thr GIu Ser Leu Ala Thr Ala Pro Thr His VaI Trp Thr Arg Pro
Leu Pro Pro Ala Asn Leu Thr Ala Ala Arg VaI Thr Ala Thr Ser Ala
His VaI VaI Trp Asp Ala Pro Thr Pro GIy Ser Leu Leu GIu Ala Tyr
VaI lie Asn VaI Thr Thr Ser GI In Ser Thr Tyr VaI Pro
Asn GIy Lys Leu Ala Ser Tyr Thr VaI Arg Asp Leu Leu Pro GIy Arg
Arg Tyr GI In Leu Ser VaI lie Ala GI In Ser Thr GIu' Leu GIy Pro
GI In Ser GIu Pro Ala His Leu Tyr H e H e Thr Ser Pro Arg Asp
GIy Ala Asp Arg Arg Trp His GI In GIy GIy His Pro Arg VaI Leu
Lys Asn Arg Pro Pro Pro Ala Arg Leu Pro GIu Leu Arg Leu Leu Asn
Asp His Ser Ala Pro GIu Thr Pro Thr GI In Pro Pro Arg Phe Ser GIu
Leu VaI Asp GIy Arg GIy Arg VaI Ser Ala Arg Phe GIy GIy Ser Pro
Ser Lys Ala Ala Thr VaI Arg Ser His VaI Pro GIy Asn Cys Ser GIu
Asn Pro Cys GI In Asn GIy GIy Thr Cys VaI Pro GIy Ala Asp Ala His
Ser Cys Asp Cys GIy Pro GIy Phe Lys GIy Arg Cys GIu Leu Ala
Cys H e Lys VaI Ser Arg Pro Cys Thr Arg Leu Phe Ser GIu Thr Lys
Ala Phe Pro VaI Trp GIu GIy GIy VaI Cys His His VaI Lys GI n Ser
Lys Ser GI In Thr Leu GIu Lys Ser

<210> 63
<211> 3987
<212> DNA
<213> Homo sapiens
<400> 63
gcgtggtgcc ttgccgactt ctaccgtttc ggcgccgagc ggcgccgacg cgtcaccccc 60
aagcaggacg acggcggctc ggggctgcgg ccgctctcgg tgtctttccc gttctcgg 120
gccgagcct cggacgtctta cgtaacgacc aacggtatca ttctctctctt gaaggsagtt 180
tcctagtctta ccccagttgc tttccccatt gccaagggcc gctgcgtggt ggagcccttc 240
Homo sapiens

Ala Val Ala Leu Ala Asp Phe Tyr Pro Phe Gly Ala Glu Arg Gly Asp
1 5 10 15

Ala Val Thr Pro Lys Gin Asp Asp Gly Gly Ser Gly Leu Arg Pro Leu
20 25 30

Ser Val Pro Phe Pro Phe Gly Val His Ser Gly Leu Tyr Val
35 40 45

Asn Asn Gin lie He Ser Phe Leu Lys Gin Val Ser Gin Phe Thr
50 55 60

Pro Val Ala Phe Pro His Ala Lys Gin Cys Val Val Val Ala Phe
65 70 75 80

Trp Ala Asp Val Asp Arg Arg Ala Gly Asp Val Tyr Tyr Arg Gin
85 90 95

Ala Thr Asp Pro Ala Met Leu Arg Arg Ala Gin Asp Val Arg His
100 105 110

Tyr Phe Pro Gin Leu Leu Asp Phe Asn Ala Thr Trp Val Phe Val Ala
115 120 125

Thr Trp Tyr Gin Val Thr Phe Phe Gly Gin Ser Ser Ser Ser Pro Val
130 135 140

Asn Thr Phe Gin Thr Val Leu He Thr Gin Gin Gin Leu Gin Gin Gin Gin
145 150 155 160

He Phe Asn Tyr Gin Gin Gin Gin Val Gin Thr Gin Gin Gin Gin Gin Gin
165 170 175

Ser Gin Gin Ala Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin
180 185 190

Phe Asn Ala Gin Gin Gin Arg Tyr Phe Ser He Gin Gin Gin Gin Gin Gin
195 200 205

Thr Ala Gin Met Gin Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
210 215 220

Gly Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
225 230 235 240

Cys Gin Thr Gin Gin Gin Ser Val Cys Leu Leu Arg Gin Ser Gin Gin Gin
245 250 255

Gly Gin Lys Cys He Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
260 265 270

Cys Gin Gin Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
275 280 285

Asn Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
290 295 300
GIy lie Asn Ser Phe Arg Cys Gin Cys Pro Ala GIy Phe GIy GIy Pro
305 310 315 320
Thr Cys GIu Thr Ala Gin Ser Pro Cys Asp Thr Lys GIu Cys Gin His
325 330 335
GIy GIy Gin Gin Gin Gin Gin Gin Gin Ser Ala VaI Cys VaI Cys Gin
340 345 350
Ala GIy Tyr Thr GIy Ala VaI Cys GIu Met Asp VaI Asp Asp Cys Ser
355 360 365
Pro Asp Pro Cys Leu Asn GIy GIy Ser Cys VaI Asp Leu VaI GIy Gin
370 375 380
Tyr Thr Cys Leu Cys Ala GIu Pro Phe Lys GIu Leu Arg Cys GIu Thr
385 390 395 400
GIy Asp His Pro VaI Pro Asp Ala Cys Leu Ser Ala Pro Cys His Asn
405 410 415
GIy GIy Thr Cys VaI Asp Ala Asp GIy GIy Tyr VaI Cys GIu Cys Pro
420 425 430
GIu GIy Phe Met GIy Leu Asp Cys Arg GIu Arg VaI Pro Asp Asp Cys
435 440 445
GIu Cys Arg Asn GIy GIy Arg Cys Leu GIy Ala Asn Thr Thr Leu Cys
450 455 460
GIu Cys Pro Leu GIy Phe Phe GIu Leu Leu Cys GIu Phe GIu lie Thr
465 470 475 480
Ala Met Pro Cys Asn Met Asn Thr GIu Cys Pro Asp GIy GIy Tyr Cys
485 490 495
Met GIu His GIy GIy Ser Tyr Leu Cys VaI Cys His Thr Asp His Asn
500 505 510
Ala Ser His Ser Leu Pro Ser Pro Cys Asp Ser Asp Pro Cys Phe Asn
515 520 525
GIy GIy Ser Cys Asp Ala His Asp Asp Ser Tyr Thr Cys GIu Cys Pro
530 535 540
Arg GIy Phe His GIy Lys His Cys GIu Lys Ala Arg Pro His Leu Cys
545 550 555 560
Ser Ser GIy Pro Cys Arg Asn GIy GIy Thr Cys Lys GIu Ala GIy GIy
565 570 575
GIu Tyr His Cys Ser Cys Pro Tyr Arg Phe Thr GIy Arg His Cys GIu
580 585 590
lie GIy Lys Pro Asp Ser Cys Ala Ser GIy Pro Cys His Asn GIy GIy
595 600 605
Thr Cys Phe His Tyr lie GIy Lys Tyr Lys Cys Asp Cys Pro Pro GIy
610 615 620
Phe Ser GIy Arg His Cys GIu lie Ala Pro Ser Pro Cys Phe Arg Ser
625 630 635 640
Ala Arg Thr Arg Pro Arg Pro VaI GIu GIy Phe GIu VaI Thr Asn VaI

980 985 990

Thr Ala Ser Thr lie Ser VaI GIn Trp Ala Leu His Arg lie Arg His

995 1000 1005

Ala Thr VaI Ser GIy VaI Arg VaI Ser lie Arg His Pro GIu Ala Leu

1010 1015 1020

Arg Asp GIn Ala Thr Asp VaI Asp Arg Ser VaI Asp Arg Phe Thr Phe

1025 1030 1035 1040

Arg Ala Leu Leu Pro GIy Lys Arg Tyr Thr lie GIn Leu Thr Thr Leu

1045 1050 1055

Ser GIy Leu Arg GIy GIu GIu His Pro Thr GIu Ser Leu Ala Thr Ala

1060 1065 1070

Pro Thr His VaI Trp Thr Arg Pro Leu Pro Pro Ala Asn Leu Thr Ala

1075 1080 1085

Ala Arg VaI Thr Ala Thr Ser Ala His VaI VaI Trp Asp Ala Pro Thr

1090 1095 1100

Pro GIy Ser Leu Leu GIu Ala Tyr VaI lie Asn VaI Thr Thr Ser GIn

1105 1110 1115 1120

Ser Thr Lys Ser Arg Tyr VaI Pro Asn GIy Lys Leu Ala Ser Tyr Thr

1125 1130 1135

VaI Arg Asp Leu Leu Pro GIy Arg Arg Tyr GIn Leu Ser VaI lie Ala

1140 1145 1150

VaI GIn Ser Thr GIu Leu GIy Pro GIu His Ser GIu Pro Ala His Leu

1155 1160 1165

Tyr lie lie Thr Ser Pro Arg Asp GIy Ala Asp Arg Arg Trp His GIn

1170 1175 1180

GIy GIy His His Pro Arg VaI Leu Lys Asn Arg Pro Pro Pro Ala Arg

1185 1190 1195 1200

Leu Pro GIu Leu Arg Leu Leu Asn His Ser Ala Pro GIu Thr Pro

1205 1210 1215

Thr GIn Pro Pro Arg Phe Ser GIu Leu VaI Asp GIy Arg GIy Arg VaI

1220 1225 1230

Ser Ala Arg Phe GIy GIy Ser Pro Ser Lys Ala Ala Thr VaI Arg Ser

1235 1240 1245

His VaI Pro GIy Asn Cys Ser GIu Asn Pro Cys GIn Asn GIy GIy Thr

1250 1255 1260

Cys VaI Pro GIy Ala Asp Ala His Ser Cys Asp Cys GIu Pro GIy Phe

1265 1270 1275 1280

Lys GIy Arg Arg Cys GIu Leu Ala Cys H e Lys VaI Ser Arg Pro Cys

1285 1290 1295

Thr Arg Leu Phe Ser GIu Thr Lys Ala Phe Pro VaI Trp GIu GIy GIy

1300 1305 1310
ttctcagtga agcgaaacag taacaacaag aatgacatca gcaggcctgc cgtgctgctg 3000
gcccgcacgc gaccccgccc tgtggaaggc ttcgaggtca ... Ser Pro VaI Asn Thr Phe GIn Thr VaI Leu H e
165 170 175
Thr Asp GIy Lys Leu Ser Phe Thr H e Phe Asn Tyr GIu Ser H e VaI
<210> 66
<211> 1358
<212> PRT
<213> Artificial Sequence

<220> INSP220 full protein sequence .and HIS tag

<400> 66
Met Arg His GIy VaI Ala Trp Ala Leu Leu VaI Ala Ala Ala Leu GIy
1 5 10 15
Leu GIy Ala Arg GIy VaI Arg GIy Ala VaI Ala Leu Ala Asp Phe Tyr
20 25 30
Pro Phe GIy Ala GIu Arg GIy Asp Ala VaI Thr Pro Lys GIu Asp Asp
35 40 45
GIy GIy Ser GIy Leu Arg Pro Leu Ser VaI Pro Phe Pro Phe GIy
50 55 60
Ala GIu His Ser GIy Leu Tyr VaI Asn Asn GIy H e H e Ser Phe
65 70 75 80
Leu Lys GIu VaI Ser GIu Phe Thr Pro VaI Ala Phe Pro H e Ala Lys
85 90 95
Asp Arg Cys VaI VaI Ala Ala Phe Trp Ala Asp VaI Asp Asn Arg Arg
100 105
Ala GIy Asp VaI Tyr Tyr Arg GIu Ala Thr Asp Pro Ala Met Leu Arg
115 120 125
Arg Ala Thr GIu Asp VaI Arg His Tyr Phe Pro GIu Leu Leu Asp Phe
130 135
140
Asn Ala Thr Trp VaI Phe VaI Ala Thr Trp Tyr Arg VaI Thr Phe Phe
145 150 155 160
GIy GIy Ser Ser Ser Ser Pro VaI Asn Thr Phe GIn Thr VaI Leu H e
165 170 175
Thr Asp GIu Lys Ser Leu Phe Thr H e Phe Asn Tyr GIu Ser H e VaI
<table>
<thead>
<tr>
<th>180</th>
<th>185</th>
<th>190</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp Thr Thr GIy Thr His Ala Ser Ser GIy GIy Asn Ala Thr GIy Leu 195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy GIy lie Ala Ala GIy GIy Phe Asn Ala GIy Asp GIy GIy Gin Arg 210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr Phe Ser lie Pro GIy Ser Arg Thr Ala Asp Met Ala GIu VaI GIu 225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Thr Thr Asn VaI GIy VaI Pro GIy Arg Trp Ala Phe Arg lie Asp 245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Ala GIy Val Arg VaI GIy GIy Cys GIy His Thr Thr Ser VaI Cys 260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ala Leu Arg Pro Cys Leu Asn GIy GIy Lys Cys H e Asp Asp Cys 275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VaI Thr GIy Asn Pro Ser Tyr Thr Cys Ser Cys Leu Ser GIy Phe Thr 290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy Arg Arg Cys His Leu Asp VaI Asn GIy Cys Ala Ser GIy Gin Pro Cys 305</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy Gin Asn GIy Thr Cys Thr His GIy H e Asn Ser Phe Arg Cys GIy Gin 325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys Pro Ala GIy Phe GIy GIy Pro Thr Cys GIu Thr Ala GIu Ser Pro 340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys Asp Thr Lys GIy Cys Gin His GIy GIy Gin Cys GIu GIu Gin Gin 355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy Ser Ala VaI Cys VaI Cys Gin Ala GIy Tyr Thr GIy Ala Ala Cys 370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIu Met Asp VaI Asp Gin Cys Ser Pro Asp Pro Cys Leu Asn GIy GIy Gin 385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Cys VaI Asp Leu VaI GIy Asn Tyr Thr Cys Leu Cys Ala GIu Pro 405</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe Lys GIy Leu Arg Cys GIu Thr GIy Asp His Pro VaI Pro Asp Ala 420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys Leu Ser Ala Pro Cys His Asn GIy GIy Thr Cys VaI Asp Ala Asp 435</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy Gin Tyr VaI Cys GIu Cys Pro GIu GIy Phe Met GIu Leu Asp Cys 450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg GIu Arg VaI Pro Asp Asp Cys GIu Cys Arg Asn GIy GIy Gin Arg Cys 465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu GIy Ala Asn Thr Thr Leu Cys Gin Cys Pro Leu GIy Phe Phe GIy Gin 485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Leu Cys GIu Phe GIu H e Thr Ala Met Pro Cys Asn Met Gin Thr 500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GIy Gin Pro Asp GIy GIy Tyr Cys Met GIu His GIy GIy Ser Tyr Leu 510</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
W2
Cys VaI Cys His Thr Asp His Asn Ala Ser His Ser Leu Pro Ser Pro 515
Cys Asp Ser Asp Pro Cys Phe Asn GIy GIy Ser Cys Asp Ala His Asp 520
Asp Ser Tyr Thr Cys GIu Cys Pro Arg GIy Phe His GIy Lys His Cys 525
GIu Lys Ala Arg Pro His Leu Cys Ser Ser GIy Pro Cys Arg Asn GIy 530
GIy Thr Cys Lys GIu Ala GIy GIy GIu Tyr His Cys Ser Cys Pro Tyr 535
Arg Phe Thr GIy Arg His Cys GIu H e GIy Lys Pro Asp Ser Cys Ala 540
Ser GIy Pro Cys His Asn GIy GIy Thr Cys Phe His Tyr H e GIy Lys 545
Tyr Lys Cys Asp Cys Pro Pro GIy Phe Ser GIy Arg His Cys GIu H e 550
Ala Pro Ser Pro Cys Phe Arg Ser Pro Cys VaI Asn GIy GIy Thr Cys 555
GIu Asp Arg Asp Thr Asp Phe Cys His Cys GIu Ala GIy Tyr Met 560
GIy Arg Arg Cys GIu Ala GIu VaI Asp Cys GIy Pro Pro GIu GIu VaI 565
Lys His Ala Thr Leu Arg Phe Asn GIy Thr Arg Leu GIy Ala VaI Ala 570
Leu Tyr Ala Cys Asp Arg GIy Tyr Ser Leu Ser Ala Pro Ser Arg H e 575
Arg VaI Cys GIu Pro His GIy VaI Trp Ser GIu Pro Pro GIu GIu Cys 580
GIu H e Asp GIu Cys Arg Ser GIu Pro Cys Leu His GIy GIy Ser Cys 585
GIu Asp Arg VaI Ala GIy Tyr Leu Cys Leu Cys Ser Thr GIy Tyr GIu 590
GIu Ala His Cys GIu Leu GIu Arg Asp GIu Cys Arg Ala His Pro Cys 595
Arg Asn GIy GIy Ser Cys Arg Asn Leu Pro GIy Ala Tyr VaI Cys Arg 600
Cys Pro Ala GIy Phe VaI GIy VaI His Cys GIu Thr GIu VaI Asp Ala 605
Cys Asp Ser Ser Pro Cys GIu His GIy GIy Arg Cys GIu Ser GIy GIy 610
GIy Ala Tyr Leu Cys VaI Cys Pro GIu Ser Phe Phe GIy Tyr His Cys
GIy Ala Asp Arg Arg Trp His GIy GIy His His Pro Arg VaI Leu
1205 1210 1215
Lys Asn Arg Pro Pro Pro Ala Arg Leu Pro GIu Leu Arg Leu Leu Asn
1220 1225 1230
Asp His Ser Ala Pro GIu Thr Pro Thr GIu Pro Arg Phe Ser GIu
1235 1240 1245
Leu VaI Asp GIy GIy Arg VaI Ser Ala Arg Phe GIy GIy Ser Pro
1250 1255 1260
Ser Lys Ala Ala Thr VaI Arg Ser His VaI Pro GIu Asn Cys Ser GIu
1265 1270 1275 1280
Asn Pro Cys GIu Asn GIy GIy Thr Cys VaI Pro GIu Ala Asp Ala His
1285 1290 1295
Ser Cys Asp Cys GIy Pro GIy Phe Lys GIy Arg Cys GIu Leu Ala
1300 1305 1310
Cys He Lys VaI Ser Arg Pro Cys Thr Arg Leu Phe Ser GIu Thr Lys
1315 1320 1325
Ala Phe Pro VaI Trp GIu GIy GIy VaI Cys His His VaI Lys GIu Ser
1330 1335 1340
Lys Ser GIu Thr Leu GIu Lys Ser His His His His His
1345 1350 1355

<210> 67
<211> 4002
<212> DNA
<213> Artificial Sequence

<220> INSP220 mature full nucleotide sequence and HIS tag

<400> 67
gcggtgccccc ttgccgacct ctaccctttc ggcgcgcagc ggcgcgcagc cgtcaccccc 60
aacagcagctt acggcagtctgc acggcagtctgc gacgtccgag gctccctccc ctgctctcgt 120
gccgagcaacct ccggacacatc ccgaaacctc gcagaagact tcctctctc tcctctctct 180
tttcatcctt catccacctt cgtcatcagc tctcaatcact caccagcagc atggagctctc 240
tgggccgatg tgaacaccatg gaggctgagg gcagcagcctg ggagctgcctg tctctctctc 300
gccgtgtgctg cgggagcatg gagggtgagg gcagcagcctg ggagctgcctg tctctctctc 360
aatcccaacctg gatgccgagt gcgcagcctg cggcggccacctgccgcgagt gcgcagcctg 420
ttcctcctgt cagacgtgtc cgtcctctct caccagcagc atggagctctc 480
atatcttcaact atgtgagcac ccgacagcagc acaggcagcag cgggggcaac 540
gccactggcct tcgccgagcct gcagcagcagc gctgatctgca gcgcagcagc gggcgagct 600
tacttcagtt taccgcaagc caccgtccag gccagcagcag cgggctggagc 660
tggtgtggtgc cggggtggtgc ggctgtcgct atcttgcttg gggggggggg 720
tgcggcgacat caaccgccctct ctggccgagc cggccgagc cggccgagc 780
tccagcgcct gcggcgagc cggccgagc cggccgagc cggccgagc 840
gccgagcggt gcccctccag ccgaggcagc tcttcctcct gcggccgagc 900
acctgtgcct gcgcgacgcc cggagcagc cggagcagc cggagcagc 960
acctgtgcct gcgcgacgcc cggagcagc cggagcagc cggagcagc 1020
cagtgagagg atggctgctg ggtggtggct gcggccgagc gcggccgagc 1080
gatagtgatg tccggtggtgc cggcagcagc cggcagcagc cggcagcagc 1140
catagtggg ttctcgcttg cgccgagcagc cgccgagcagc cgccgagcagc 1200
ggagctgcct gcgcgagc cggcagcagc cggcagcagc cggcagcagc 1260
gtgctgagat cagcgtgagat cagcgtgagat cagcgtgagat cagcgtgagat 1320
Agggagagag tccccgatga ctgtgagtgc cgcaacggag gcagatgcct gggcgccaac 1380
accaccctct gccagtgccc cctgggattc tttgggcttc ...
Ser GIy Ser GIy Leu Arg Pro Leu 202530
Ser VaI Pro Phe Pro Phe GIy Ala GIu His Ser GIy Leu Tyr VaI 354045

Another sequence is provided, including translational reading frames and amino acid translations. The text contains a mix of nucleotide sequences, amino acid translations, and alignments, indicating the context of protein coding and modification.
Tyr Thr Cys Leu Cys Ala GIu Pro Phe Lys GIy Leu Arg Cys GIu Thr

GIy Asp His Pro VaI Pro Asp Ala Cys Leu Ser Ala Pro Cys His Asn

GIy GIy Thr Cys VaI Asp Ala Asp GIn GIy Tyr VaI Cys GIu Cys Pro

GIu GIy Phe Met GIy Leu Asp Cys Arg GIu Arg VaI Pro Asp Asp Cys

GIu Cys Arg Asn GIy GIy Arg Cys Leu GIy Ala Asn Thr Thr Leu Cys

GIu Cys Pro Leu GIy Phe Phe GIy Leu Leu Cys GIu Phe GIu lie Thr

Ala Met Pro Cys Asn Met Asn Thr GIn Cys Pro Asp GIy GIy Tyr Cys

Met GIu His GIy GIy Ser Tyr Leu Cys VaI Cys His Thr Asp His Asn

Ala Ser His Ser Leu Pro Ser Pro Cys Asp Ser Asp Pro Cys Phe Asn

GIy GIy Ser Cys Asp Ala His Asp Ser Tyr Thr Cys GIu Cys Pro

Arg GIy Phe His GIy Lys His Cys GIu Lys Ala Arg Pro His Leu Cys

Ser Ser GIy Pro Cys Arg Asn GIy GIy Thr Cys Lys GIu Ala GIy GIy

GIu Tyr His Cys Ser Cys Pro Tyr Arg Phe Thr GIy Arg His Cys GIu

H e GIy Lys Pro Asp Ser Cys Ala Ser GIy Pro Cys His Asn GIy GIy

Thr Cys Phe His Tyr H e GIy Lys Tyr Lys Cys Asp Cys Pro Pro Gly

Phe Ser GIy Arg His Cys GIu H e Ala Pro Ser Pro Cys Phe Arg Ser

Pro Cys VaI Asn GIy GIy Thr Cys GIu Asp Arg Asp Thr Asp Phe Phe

Cys His Cys GIn Ala GIy Tyr Met GIy Arg Arg Cys GIa GIu VaI

Asp Cys GIy Pro Pro GIu GIu VaI Lys His Ala Thr Leu Arg Phe Asn

GIy Thr Arg Leu GIy Ala VaI Ala Leu Tyr Ala Cys Asp Arg GIy Tyr

Ser Leu Ser Ala Pro Ser Arg H e Arg VaI Cys GIu Pro His GIy VaI
Trp Ser GIu Pro Pro GIn Cys Leu GIu lie Asp GIu Cys Arg Ser GIn
725 730 735
Pro Cys Leu His GIy GIy Ser Cys GIn Asp Arg VaI Ala GIy Tyr Leu
740 745 750
Cys Leu Cys Ser Thr GIy Tyr GIu GIy Ala His Cys GIu Leu GIu Arg
755 760 765
Asp GIu Cys Arg Ala His Pro Cys Arg Asn GIy GIy Ser Cys Arg Asn
770 775 780
Leu Pro GIy Ala Tyr VaI Cys Arg Cys Pro Ala GIy Phe VaI GIy VaI
785 790 795 800
His Cys GIu Thr GIu VaI Asp Ala Cys Asp Ser Ser Pro Cys GIn His
805 810 815
GIy GIy Arg Cys GIu Ser GIy GIy GIy Ala Tyr Leu Cys VaI Cys Pro
820 825 830
GIu Ser Phe Phe GIy Tyr His Cys GIu Thr VaI Ser Asp Pro Cys Phe
835 840 845
Ser Ser Pro Cys GIy GIy Arg GIy Tyr Cys Leu Ala Asn GIy Ser
850 855 860
His Ser Cys Thr Cys Lys VaI GIy Tyr Thr GIu GIu Asp Cys Ala Lys
865 870 875 880
GIu Leu Phe Pro Pro Thr Ala Leu Lys Met GIu Arg VaI GIu GIu Ser
885 890 895
GIy VaI Ser lie Ser Trp Asn Pro Pro Asn GIy Pro Ala Ala Arg GIn
900 905 910
Met Leu Asp GIy Tyr Ala VaI Thr Tyr VaI Ser Ser Asp GIy Ser Tyr
915 920 925
Arg Arg Thr Asp Phe VaI Asp Arg Thr Arg Ser Ser His GIn Leu GIu
930 935 940
Ala Leu Ala AlA GIy Arg Ala Tyr Asn lie Ser VaI Phe Ser VaI Lys
945 950 955 960
Arg Asn Ser Asn Lys Asn Asp lie Ser Arg Pro Ala VaI Leu Leu
965 970 975
Ala Arg Thr Arg Pro Arg Pro VaI GIu GIy Phe GIu VaI Thr Asn VaI
980 985 990 995
Thr Ala Ser Thr lie Ser VaI GIn Thr Ala Leu His Arg lie Arg His
995 1000 1005
Ala Thr VaI Ser GIy VaI Arg VaI Ser H e Arg His Pro GIu Ala Leu
1010 1015 1020
Arg Asp GIu Ala Thr Asp VaI Asp Ser VaI Asp Arg Phe Thr Phe
1025 1030 1035 1040
Arg Ala Leu Leu Pro GIy Lys Arg Tyr Thr lie GIn Leu Thr Thr Leu
1045 1050 1055 1060
Ser GIy Leu Arg GIy GIu GIu His Pro Thr GIu Ser Leu Ala Thr Ala
1060 1065 1070
Pro Thr His VaI Trp Thr Arg Pro Leu Pro Pro Ala Asn Leu Thr Ala
1075 1080 1085
Ala Arg VaI Thr Ala Thr Ser Ala His VaI VaI Trp Asp Ala Pro Thr
1090 1095 1100
Pro GIy Ser Leu Leu GIu Ala Tyr VaI lie Asn VaI Thr Thr Ser Gin
1105 1110 1115 1120
Ser Thr Lys Ser Arg Tyr VaI Pro Asn GIy Lys Leu Ala Ser Tyr Thr
1125 1130 1135
VaI Arg Asp Leu Leu Pro GIy Arg Arg Tyr Gin Leu Ser VaI lie Ala
1140 1145 1150
VaI Gin Ser Thr Gin Leu Gin Pro Gin His Ser GIu Pro AlA His Leu
1155 1160 1165
Tyr He He He Ser Pro Arg Asp GIy Ala Asp Arg Arg Trp His Gin
1170 1175 1180
GIy GIy His His Pro Arg VaI Leu Lys Asn Arg Pro Pro Pro Ala Arg
1185 1190 1195 1200
Leu Pro GIu Leu Arg Leu Leu Asp His Ser Ala Pro GIu Thr Pro
1205 1210 1215
Thr Gin Pro Pro Pro Pro Pro Gin Gin Leu Gin Gin Gin Gin Gin Gin
1220 1225 1230
Ser Ala Arg Phe GIy GIy Ser Pro Ser Lys Ala Ala Thr VaI Arg Ser
1235 1240 1245
His VaI Pro GIy Asn Cys Ser GIu Asn Pro Cys Gin Asn GIy Gin Gin
1250 1255 1260
Cys VaI Pro GIy Ala Asp Ala His Ser Cys Asp Cys GIy Pro GIy Phe
1265 1270 1275 1280
Lys GIy Arg Arg Cys GIu Leu Ala Cys He Lys VaI Ser Arg Pro Cys
1285 1290 1295
Thr Arg Leu Phe Ser GIu Thr Lys Ala Phe Pro VaI Trp GIu Gin Gin
1300 . 1305 1310
VaI Cys His His VaI Lys Gin Ser Lys Ser Gin Thr Leu Gin Lys Gin
1315 1320 1325
His His His His His
1330
<210> 69
<211> 522
<212> PRT
<213> Homo sapiens
<400> 69
Asp Ala Cys Asp Ser Ser Pro Cys Gin His GIy GIy Arg Cys Gin Ser
1 5 10 15
GIy Pro GIn His Ser GIu Pro Ala His Leu Tyr lie He Thr Ser Pro
355 360 365
Arg Asp GIy Ala Asp Arg Arg Trp His Gin GIy GIy His His Pro Arg
370 375 380
Val Leu Lys Asn Arg Pro Pro Pro Ala Arg Leu Pro GIu Leu Arg Leu
385 390 395 400
Leu Asn Asp His Ser Ala Pro GIu Thr Pro Thr Gin Pro Pro Arg Phe
405 410 415
Ser GIu Leu VaI Asp GIy Arg GIy Arg VaI Ser AlA Arg Phe GIy GIy
420 425 430
Ser Pro Ser Lys Ala Ala Thr VaI Arg Ser His VaI Pro GIy Asn Cys
435 440 445
Ser GIu Asn Pro Cys Gin Asn GIy GIy Thr Cys VaI Pro GIy Ala Asp
450 455 460
Ala His Ser Cys Asp Cys GIy Pro GIy Phe Lys GIy Arg Arg Cys GIu
465 470 475 480
Leu Ala Cys lie Lys VaI Ser Arg Pro Cys Thr Arg Leu Phe Ser GIu
485 490 495
Thr Lys Ala Phe Pro VaI Trp GIu GIy GIy VaI Cys His His VaI Lys
500 505 510
Gin Ser Lys Ser GIn Thr Leu GIu Lys Ser
515 520

<210> 70
<211> 641
<212> PRT
<213> Homo sapiens

<Glu Asn Gin Ser Ala VaI Cys VaI Cys Gin Ala Gin Tyr Thr Gin Ala
1 5 10 15
Ala Cys GIu Met Asp VaI Asp Asp Ser Pro Asp Pro Cys Leu Asn
20 25 30
GIy GIy Ser Cys VaI Asp Leu VaI GIy Asn Tyr Thr Cys Leu Cys Ala
35 40 45
GIu Pro Phe Lys GIu Leu Arg Cys GIu Thr GIy Asp His Pro VaI Pro
50 55 60
Asp Ala Cys Leu Ser Ala Pro Cys His Asn GIy GIy Thr Cys VaI Asp
65 70 75 80
Ala Asp GIn GIy Tyr VaI Cys GIu Cys Pro GIu GIy Phe Met GIy Leu
85 90 95
Asp Cys Arg GIu Arg VaI Pro Asp Asp Cys GIu Cys Arg Asn GIy GIy
100 105 110
Arg Cys Leu GIy Ala Asn Thr Thr Leu Cys Gin Cys Pro Leu GIy Phe
115 120 125
Phe GIy Leu Leu Cys GIu Phe GIu lie Thr Ala Met Pro Cys Asn Met
130 135 140
Asn Thr GIu Cys Pro Asp GIy GIy Tyr Cys Met GIu His GIy GIy Ser
145 150 155 160
Tyr Leu Cys VaI Cys His Thr Asp His Asn Ala Ser His Ser Leu Pro
165 170 175
Ser Pro Cys Asp Ser Asp Pro Cys Phe Asn GIy GIy Ser Cys Asp Ala
180 185 190
His Asp Asp Ser Tyr Thr Cys GIu Cys Pro Arg GIy Phe His GIy Lys
195 200 205
His Cys GIu Lys Ala Arg Pro His Leu Cys Ser Ser GIy Pro Cys Arg
210 215 220
Asn GIy GIy Thr Cys Lys GIu Ala GIy GIy GIu Tyr Cys Ser Cys
225 230 235 240
Pro Tyr Arg Phe Thr GIy Arg His Cys GIu lie GIy Lys Pro Asp Ser
245 250 255
Cys Ala Ser GIy Pro Cys His Asn GIy GIy Thr Cys Phe His Tyr He e
260 265 270
GIy Lys Tyr Lys Cys Asp Cys Pro Pro GIy Phe Ser GIy Arg His Cys
275 280 285
GIu He e Ala Pro Ser Pro Cys Phe Arg Ser Pro Cys VaI Asn GIy GIy
290 295 300
Thr Cys GIu Asp Arg Asp Thr Asp Phe Phe Cys His Cys GIu Ala GIy
305 310 315 320
Tyr Met GIy Arg Arg Cys GIu Ala GIu VaI Asp Cys GIu Pro Pro GIu
325 330 335
GIu VaI Lys His Ala Thr Leu Arg Phe Asn GIy Thr Arg Leu GIy Ala
340 345 350
VaI Ala Leu Tyr Ala Cys Asp Arg GIy Tyr Ser Leu Ser Ala Pro Ser
355 360 365
Arg He e Arg VaI Cys GIu Pro His GIy VaI Trp Ser GIu Pro Pro GIu
370 375 380
Cys Leu GIu He e Asp GIu Cys Arg Ser GIu Pro Cys Leu His GIy GIy
385 390 395 400
Ser Cys GIu Asp Arg VaI Ala GIy Tyr Leu Cys Leu Cys Ser Thr GIy
405 410 415
Tyr GIu GIy Ala His Cys GIu Leu GIu Arg Asp GIu Cys Arg Ala His
420 425 430
Pro Cys Arg Asn GIy GIy Ser Cys Arg Asn Leu Pro GIy Ala Tyr VaI
435 440 445
Cys Arg Cys Pro Ala GIy Phe VaI GIy VaI His Cys GIu Thr GIu VaI
450 455 460
Asp Ala Cys Asp Ser Ser Pro Cys Gln His G1y G1y Arg Cys Glu Ser
465 470 475 480
G1y G1y G1y Ala Tyr Leu Cys Val Cys Pro Glu Ser Phe Phe G1y Tyr
485 490 495
His Cys Glu Thr Val Ser Asp Pro Cys Phe Ser Ser Pro Cys Glu G1y
500 505 510
Arg G1y Tyr Cys Leu Ala Ser Asn G1y Ser His Ser Cys Thr Cys Lys
515 520 525
Val G1y Tyr Thr G1y G1u Asp Cys Ala Lys G1u Leu Phe Pro Pro Thr
530 535 540
Ala Leu Lys Met Glu Arg Val G1u G1u Ser G1y Val G1y Ser lie Ser Trp
545 550 555 560
Asn Pro Pro Asn G1y Pro Ala Ala Arg Gln Met Leu Asp G1y Tyr Ala
565 570 575
Val Thr Tyr Val Ser Ser Asp G1y Ser Tyr Arg Arg Thr Asp Phe Val
580 585 590
Asp Arg Thr Arg Ser Ser His Gln Leu Gln Ala Leu Ala Ala G1y Arg
595 600 605
Ala Tyr Asn lie Ser Val Phe Ser Val Lys Arg Asn Ser Asn Asn Lys
610 615 620
Asn Asp H e Ser Arg Pro Ala Val Leu Leu Ala Arg Thr Arg Pro Arg
625 630 635 640
Pro

<210>  71
<211>  13
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein linker sequence

<400>  71
G1u Phe G1y Ala G1y Leu Val Leu G1y G1y G1n Phe Met
1  5  10
A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/47 C12N15/12

According to International Patent Classification (IPC) or both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database consulted during the international search (name of database and where practical, search terms used)

EPO-Internal, Sequence Search, EMBASE, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
</table>

Further documents are listed in the continuation of Box C

* Special categories of cited documents

* document defining the general state of the art which is not considered to be of particular relevance
*E* earlier document published on or after the international filing date
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
*O* document referring to an oral disclosure, use, exhibition or other means
*P* document published prior to the international filing date but later than the priority date claimed

*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*X* document of particular relevance the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*Y* document of particular relevance the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*S* document member of the same patent family

Date of the actual completion of the international search: 22 January 2007

Date of mailing of the international search report: 07/02/2007

Name and mailing address of the ISA/

European Patent Office, P B 581 B Patentlaan 2 NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040, Tx 31 651 epo nl, Fax (+31-70) 340-3016

Authorized officer:

Wiame, Use
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>DATABASE EMBL [Onl ine] 6 April 1 2003 (2003-04-06), &quot;Mus muscul us secreted protein SST3 mRNA, complete cds.&quot; XP002415842 retrieved from EBI access on no. EM PRO:AY169783 Database accession no. AY169783 the whole document</td>
<td>1-7,9, 27</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication where appropriate, of the relevant passages</td>
<td>Relevant to claim</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>page 172; example 20; sequences 79-82. page 275 - page 290</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Database accession no. AK074062 the whole document</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Database accession no. AC104809 the whole document</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the whole document</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>DATABASE UniProt [Online] 1 October 2003 (2003-10-01), &quot;FLJ00146 protein (Fragment).&quot; XP002415845 Database accession no. Q8TEP7 cited in the application</td>
<td>1-5</td>
</tr>
<tr>
<td></td>
<td>the whole document</td>
<td></td>
</tr>
</tbody>
</table>

Form PCT/ISA/210 (continuation of second sheet) (April 2005)
Continuation of Box II.1

Although claim 18, and claims 19-26 insofar as they are dependent on claim 18, are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition. Although claims 30-32 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box II.2

Claims Nos.: 14-16

Present claims 14-16 relate to compounds defined by reference to a desirable characteristic or property, namely increasing or decreasing the level of expression or activity of a polypeptide of the invention. The claims cover all compounds having this characteristic or property, whereas the application does not provide support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search for said claims is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search for said claims impossible. The same applies to claims 17, 20, 27 and 29-32 insofar as they are dependent on claims 14-16.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.
INTERNATIONAL SEARCH REPORT

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos 1-10 because they relate to subject matter not required to be searched by this Authority, namely

   see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos 14-16 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically

   see FURTHER INFORMATION sheet PCT/ISA/210

3. [J] Claims Nos because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [X] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims

2. [j] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee

3. [X] As only some of the required additional search fees were timely paid by the applicant, this International Search Report

4. [J] No required additional search fees were timely paid by the applicant Consequently, this International Search Report is restricted to the invention first mentioned in the claims it is covered by claims Nos

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest

☐ No protest accompanied the payment of additional search fees

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2004)
# INTERNATIONAL SEARCH REPORT

Information on patent family members

<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>US 2004214189 A1</td>
<td>28-10-2004</td>
</tr>
<tr>
<td>US 2003125296 A1</td>
<td>03-07-2003</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td>WO 2004087874 A1</td>
<td>14-10-2004</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td>WO 02083841 A</td>
<td>24-10-2002</td>
<td>CA 2442729 A1</td>
<td>24-10-2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1399537 A2</td>
<td>24-03-2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OP 2004533235 T</td>
<td>04-11-2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2004006205 A1</td>
<td>08-01-2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1578900 A2</td>
<td>28-09-2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2005095589 A1</td>
<td>05-05-2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2003040616 A1</td>
<td>27-02-2003</td>
</tr>
<tr>
<td>JP 2006158244 A</td>
<td>22-06-2006</td>
<td>NONE</td>
<td></td>
</tr>
</tbody>
</table>

Form PCT/ISA/210 (patent family annex) (April 2005)