(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 June 2007 (28.06.2007)

PCT

(10) International Publication Number
WO 2007/072012 A1

(51) International Patent Classification:
C07K 14/81 (2006.01)

(21) International Application Number:
PCT/GB2006/004846

(22) International Filing Date:
21 December 2006 (21.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0525999.9 21 December 2005 (21.12.2005) GB

(71) Applicant *(for all designated States except US)*: **ARES TRADING S.A.** [CH/CH]; Zone Industrielle de L'Ouriettaz, CH-1170 Aubonne (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FAGAN, Richard, Joseph** [US/GB]; Inpharmatica Ltd, 1 New Oxford St, London WC1A 1NU (GB). **MICHALOVICH, David** [GB/GB]; Inpharmatica Ltd, 1 New Oxford St, London WC1A 1NU (GB). **WHITE, Simon** [GB/GB]; Inpharmatica Ltd, 1 New Oxford St, London WC1A 1NU (GB). **MITTER, Richard** [GB/GB]; Inpharmatica Ltd, 1 New Oxford St, London WC1A 1NU (GB).

(74) Agents: **GOODFELLOW, Hugh, Robin** et al.; Carpmaels & Ransford, 43-45 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, 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, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

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.

(54) **Title:** NOVEL MEMBERS OF THE KAZAL FAMILY OF SERINE PROTEASE INHIBITORS

(57) **Abstract:** This invention relates to novel proteins, termed INSP 170, INSP222 and INSP223, herein identified as novel human members of the Kazal family of serine protease inhibitors, and to the use of these and nucleic acid sequences from the encoding genes in the diagnosis, prevention and treatment of disease.

WO 2007/072012 A1

Novel members of the Kazal family of serine protease inhibitors

This invention relates to novel proteins, termed INSP170, INSP222 and INSP223, herein identified as novel human members of the Kazal family of serine protease inhibitors, and to the use of these proteins and nucleic acid sequences from the encoding genes 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.

Introduction

I. Introduction to Serine Protease Inhibitors

Serine protease inhibitors are found in animals, plants and microorganisms and are involved in many different biological processes, from blood coagulation in vertebrates to response to mechanical damage in plants. Serine protease inhibitors are widespread throughout the plant kingdom where they play an important role in protection against pests and pathogens. Serine protease inhibitors can be divided into approximately 20 different families on the basis of sequence and structural similarity (Otlewski, J et al 1999 Acta

Biochim Pol; 46 (3): 531-365 and Bode, W and Huber, R 2000 Biochim Biophys Acta 1477(1-2): 241-252, see Table 1).

The majority of serine protease inhibitors, as their name suggests, inhibit only serine proteases, however there are examples of cross-class inhibition of non-serine proteases (Laskowski Jr. M, Encyclopedia of Molecular Biology, vol 4). In addition, there are examples of proteins which are classified as members of a serine protease inhibitor family but lack the necessary residues to perform this function. For example, Ovalbumin, while classified as a serpin, is not an active protease inhibitor. Some members of the BPTI (Kunitz) family, such as dentrotoxins and bungarotoxins also lack protease inhibitor function and instead have an entirely different activity as K^+ and Ca^{2+} channel blockers. (Laskowski Jr. M, Encyclopedia of Molecular Biology, vol 4).

Active serine protease inhibitors operate by strongly binding the active site of their target protease in place of the preferred substrate. About 12 residues of the inhibitor make close contacts with the protease. However, the strength of interaction and specificity of inhibitors depends, in particular, on the primary specificity site "P1". (Laskowski Jr. M, Encyclopedia of Molecular Biology, vol 4). For example, inhibitors of trypsin and trypsin-like enzymes tend to have Lys or Arg as the P1 residue, whilst inhibitors of chymotrypsin and chymotrypsin-like enzymes tend to have either Pro, Tyr, Phe, Leu or Met as their P1 residue and inhibitors of elastase-like enzymes have Ala or Ser as their P1 residue (Laskowski, M. & Kato, I. (1980) Protein inhibitors of proteinases. Ann. Rev. Biochem., 49, 593-626). Thr at the P1 position is common in proteinase inhibitors present in secretory proteins.

Three different types of Serine protease inhibitor can be distinguished on the basis of mechanism: (i) canonical (standard mechanism) inhibitors, (ii) non-canonical inhibitors, and (iii) serpins.

(i) Canonical inhibitors are relatively small proteins. They all possess an exposed and convex binding loop of a characteristic canonical conformation, but are otherwise unrelated in structure. The mechanism of inhibition in this group is consistently very similar and resembles that of an ideal substrate.

(ii) Non-canonical inhibitors, originating from blood sucking organisms, specifically block enzymes of the blood-clotting cascade.

(iii) Serpins interact with their target proteases in a substrate-like manner. However, cleavage of a single peptide bond in a flexible and exposed binding loop leads to dramatic structural changes.

There is a great need for the identification of novel serine proteases, particularly human members of Kazal family of serine protease inhibitors, as these proteins may play a role in a number of disease conditions.

The Invention

The invention is based on the surprising finding that the human INSP170, INSP222 and INSP223 polypeptides are members of the Kazal family of serine protease inhibitors. INSP170 and INSP223 are predicted to contain 2 Kazal domains and to show homology to known mammalian carnivore double-headed protease inhibitors ("bikazals"). Thus the Kazal family of serine protease inhibitors includes both proteins having only one Kazal domain and proteins having more than one Kazal domain.

In one embodiment of the first aspect of the invention, there is provided a polypeptide which:

(i) comprises the amino acid sequence as recited in SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26 and/or SEQ ID NO:28;

(ii) is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or has an antigenic determinant in common with the polypeptides of (i); or

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

Preferably, the polypeptide according to the first embodiment of the first aspect of the invention comprises the amino acid sequence as recited in SEQ ID NO:28.

According to a second embodiment of this first aspect of the invention, there is provided a polypeptide which consists of the amino acid sequence as recited in SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24 , SEQ ID NO:26 and/or SEQ ID NO:28.

The polypeptide having the sequence recited in SEQ ID NO:18 is referred to hereafter as the "INSP222 protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:20 is referred to hereafter as the "INSP222 protein sequence exon 2". The polypeptide having the sequence recited in SEQ ID NO:22 is referred to hereafter as the

"INSP222 protein sequence exon 3". The polypeptide having the sequence recited in SEQ ID NO:24 is referred to hereafter as the "INSP222 protein sequence exon 4". The polypeptide having the sequence recited in SEQ ID NO:26 is referred to hereafter as the "INSP222 protein sequence exon 5". The polypeptide having the sequence recited in SEQ ID NO:28 is referred to hereafter as the "INSP222 full protein sequence".

The term "INSP222 polypeptides" as used herein includes polypeptides comprising the INSP222 protein sequence exon 1, the INSP222 protein sequence exon 2, the INSP222 protein sequence exon 3, the INSP222 protein sequence exon 4, the INSP222 protein sequence exon 5 and the INSP222 full protein sequence.

In a third embodiment of the first aspect of the invention, there is provided a polypeptide which:

(i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and/or SEQ ID NO:16;

(ii) is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or has an antigenic determinant in common with the polypeptides of (i); or

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

Preferably, the polypeptide according to this first embodiment of the first aspect of the invention comprises the amino acid sequence as recited in SEQ ID NO:12 or SEQ ID NO 16.

According to a fourth embodiment of this first aspect of the invention, there is provided a polypeptide which consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and/or SEQ ID NO:16.

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as the "INSP170 protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as the "INSP170 protein sequence exon 2". The polypeptide having the sequence recited in SEQ ID NO:6 is referred to hereafter as the "INSP170 protein sequence exon 3 ". The polypeptide having the sequence recited in SEQ ID NO:8 is referred to hereafter as the "INSP170 protein sequence exon 4". The

polypeptide having the sequence recited in SEQ ID NO:10 is referred to hereafter as the "INSP170 protein sequence exon 5". The polypeptide having the sequence recited in SEQ ID NO:12 is referred to hereafter as the "INSP170 full protein sequence". The polypeptide having the sequence recited in SEQ ID NO:14 is referred to hereafter as the "INSP170 mature protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:16 is referred to hereafter as the "INSP170 full mature protein sequence".

The term "INSP170 polypeptides" as used herein includes polypeptides comprising the INSP170 protein sequence exon 1, the INSP170 protein sequence exon 2, the INSP170 protein sequence exon 3, the INSP170 protein sequence exon 4, the INSP170 protein sequence exon 5, the INSP170 full protein sequence, the INSP170 mature protein sequence exon 1 and the INSP170 full mature protein sequence.

Although the Applicant does not wish to be bound by this theory, it is postulated that amino acid residues 1-18 of the INSP170 full length polypeptide (SEQ ID NO:16) form a signal peptide, as shown in the schematic representation below:

MKGFTAFAIL GLAVTAWATS PYGKSGFLSN SHAFMLTVNC SRYIKGLKTA

CPRDWRPICG TDQKTYSNEC MFCMQNQEVD QEFQLRKLHE GKCIQCTRYS

EACTMDYTLH CGSDGNVYSN RCTFCNTVVK SQGAIWLKNY GLC

The INSP170 exon 1 and INSP170 full length polypeptide sequences without the postulated signal sequence are recited in SEQ ID NO:14 and SEQ ID NO:16, respectively. The polypeptide having the sequence recited in SEQ ID NO:14 is referred to herein as the "INSP170 mature protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:16 is referred to hereafter as the "INSP170 full mature protein sequence".

Further, although the Applicant does not wish to be bound by this theory, it is possible that the INSP170 protein sequence exon 1 or the INSP170 mature protein sequence exon 1 may consist of only part of the sequence recited in SEQ ID NO:2 or SEQ ID NO:14 respectively. Where this is the case, the INSP170 protein sequence exon 1 or INSP170 mature protein sequence exon 1 will differ from that recited in SEQ ID NO:2 or SEQ ID NO:14 respectively by way of deletion of one or more amino acid residues from the 3' terminal end of the INSP170 protein sequence exon 1 or INSP170 mature protein sequence exon 1. Thus although the INSP170 protein sequence exon 1 preferably consists of residues 1-37 as recited in SEQ ID NO:2, it may alternatively consist of residues 1-36, 1-

35, 1-34, 1-33, 1-32, 1-31, 1-30, 1-29 or a polypeptide in which further residues have been deleted from the 3' end of SEQ ID NO:2. Polypeptides comprising the sequence of SEQ ID NO:2 with one or more residues deleted from its 3' end in combination with the sequences recited in one or more of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 and SEQ ID NO:10 are thus specifically included within the scope of the invention. Likewise, although the INSP170 mature protein sequence exon 1 preferably consists of residues 1-19 as recited in SEQ ID NO:14, it may alternatively consist of residues 1-18, 1-17, 1-16, 1-15, 1-14, 1-13, 1-12, 1-11 or a polypeptide in which further residues have been deleted from the 3' end of SEQ ID NO:14. Polypeptides comprising the sequence of SEQ ID NO:14 with one or more residues deleted from its 3' end in combination with the sequences recited in one or more of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8 and SEQ ID NO:10 are thus specifically included within the scope of the invention.

In the fifth embodiment of the first aspect of the invention, there is provided a polypeptide which:

(i) comprises the amino acid sequence as recited in 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 and/or SEQ ID NO:46;

(ii) is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or has an antigenic determinant in common with the polypeptides of (i); or

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

Preferably, the polypeptide according to this first embodiment of the first aspect of the invention comprises the amino acid sequence as recited in SEQ ID NO:42 or SEQ ID NO 46.

According to a sixth embodiment of this first aspect of the invention, there is provided a polypeptide which consists of the amino acid sequence as recited in 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 and/or SEQ ID NO:46.

The polypeptide having the sequence recited in SEQ ID NO:30 is referred to hereafter as the "INSP223 protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:32 is referred to hereafter as the "INSP223 protein sequence exon 2". The

polypeptide having the sequence recited in SEQ ID NO:34 is referred to hereafter as the "INSP223 protein sequence exon 3 ". The polypeptide having the sequence recited in SEQ ID NO:36 is referred to hereafter as the "INSP223 protein sequence exon 4". The polypeptide having the sequence recited in SEQ ID NO:38 is referred to hereafter as the "INSP223 protein sequence exon 5". The polypeptide having the sequence recited in SEQ ID NO:40 is referred to hereafter as "INSP223 protein sequence exon 6". The polypeptide having the sequence recited in SEQ ID NO:42 is referred to hereafter as the "INSP223 full protein sequence". The polypeptide having the sequence recited in SEQ ID NO:44 is referred to hereafter as the "INSP223 mature protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:46 is referred to hereafter as the "INSP223 full mature protein sequence".

The term "INSP223 polypeptides" as used herein includes polypeptides comprising the INSP223 protein sequence exon 1, the INSP223 protein sequence exon 2, the INSP223 protein sequence exon 3, the INSP223 protein sequence exon 4, the INSP223 protein sequence exon 5, the INSP223 protein sequence exon 6, the INSP223 full protein sequence, the INSP223 mature protein sequence exon 1 and the INSP223 full mature protein sequence.

Although the Applicant does not wish to be bound by this theory, it is postulated that amino acid residues 1-18 of the INSP223 full length polypeptide (SEQ ID NO:42) form a signal peptide, as shown in the schematic representation below:

MKGFTAFAIL GLAVTAWATS PYGTKVNCSR YIKGLKTACP RDWRPICGTD QKTYSNECMF CMQNQEVDQE FQLRKLHEGK CIQCTRYSEA CTMDYTLHCG SDGNVYSNRC TFCNTVVKSQ GAIWLKNYGL C

The INSP223 exon 1 and INSP223 full length polypeptide sequences without the postulated signal sequence are recited in SEQ ID NO:44 and SEQ ID NO:46, respectively. The polypeptide having the sequence recited in SEQ ID NO:44 is referred to herein as the "INSP223 mature protein sequence exon 1". The polypeptide having the sequence recited in SEQ ID NO:46 is referred to hereafter as the "INSP223 full mature protein sequence".

The term "Kazal family of serine protease inhibitors" refers to a molecule containing at least one Kazal domain. Kazal domains are found in many proteins with diverse roles. In

particular, Kazal domains are found in many members of the canonical (standard mechanism) family of serine protease inhibitors such as pancreatic secretory trypsin inhibitor, avian ovomucoid, acrosin inhibitor and elastase inhibitor. Between 1 and 7 Kazal domains are found in each of these proteins. Although the presence of a Kazal domain in a protein is usually indicative of serine protease inhibitor activity, Kazal domains are also seen in other proteins such as agrin and follistatin which are not known to function as serine protease inhibitors. Further examples of Kazal domain-containing proteins are SPINK1, SPINK2, SPINK5, recombinant TgPI-1, LDTI, RECK, Agrin and Follistatin. Each of these is discussed briefly below:

SPINK1 is a pancreatic secretory trypsin inhibitor. It is believed to prevent the premature activation of pancreatic zymogens by trypsin, otherwise resulting in pancreatitis. It is also expressed by the mucosa of the GI tract where it's thought to prevent the breakdown of the mucosal cells (Witt *et al*, 2000, Nature: Genetics, 25, 213-216; OMIM: 167790).

SPINK2 is an acrosin-trypsin inhibitor expressed in testis, epididymis and seminal vesicle (OMIM: 605753). Acrosin is the major protease of mammalian spermatozoa. SPINK2 is a strong inhibitor of acrosin in both the male and female genital tracts.

SPINK5 has been shown to have an in-vitro anti-trypsin activity and is expressed in epithelia and in the thymus (OMIM: 605010).

The intracellular parasite *Toxoplasma gondii* secretes protease inhibitors that protect the parasite from proteolytic enzymes found in the lower intestine, (e.g. Recombinant TgPI-1 inhibits trypsin, chymotrypsin, pancreatic elastase and neutrophil elastase) (Morris *et al*, 2002, Biological Chemistry, 277, 45259-45266).

LDTI, a secreted Kazal domain-containing protein isolated from the medicinal leech has been shown to inhibit human mast cell tryptase. The physiological role of tryptase are not clear though the central role of mast cells in allergic and inflammatory reactions suggests that tryptase may be involved in the pathogenesis of disorders such as asthma, interstitial lung disease, psoriasis, arthritis, gingivitis and peridontitis (Stubbs *et al*, 1997, JBC, 272, 19931-19937).

RECK inhibits MMP9, MMP2 and MMP14. The precise mechanism of inhibition is not known and may not be due to direct interaction between the MMPs and RECK's Kazal

domains. RECK down-regulation by oncogenic signals may facilitate tumor invasion and metastasis.

Agrin is a neuronal aggregating factor that induces the aggregation of acetylcholine receptors and other postsynaptic proteins on muscle fibers and is crucial for the formation of the neuromuscular junction. Despite containing Kazal domains, Agrin is not known to function as a serine protease inhibitor.

Follistatin regulates several signaling pathways in a cell and tissue specific manner by inactivating TGFβ-like growth factors. Follistatin is an antagonist of activin and is also a specific inhibitor of the secretion and biosynthesis of follicle-stimulating hormone (FSH) (Innis *et al, 2003,* JBC, 278, 39969-39977).

In Pacific herring, Kazal-type trypsin inhibitors are synthesized in follicle cells, secreted, accumulated in the egg chorion during oocyte development and released at spawning to activate spermatozoa motility (Oda, S. et al., (1998), Dev. Biol., 1:204(1): 55-63). Thus a role in fertility of Kazal-type proteins may not be restricted to males.

Known human members of the Kazal family of serine protease inhibitors, such as SPINK1 and SPINK4, are typically short (~80 amino acids) and are composed of an N-terminal signal peptide followed by a C-terminal Kazal domain. However, double-headed mammalian protease inhibitors having two Kazal domains are also known ("bikazals") (for example, see Reisinger et al., Biol. Chem. Hoppe Seyler, 1987, 368(6):717-726; Reisinger et al., Protein Seq. Data Anal., 1988, 1(4):259-261; Hochstrasser et al., Protein Seq. Data Anal., 1989, 2(6):453-456; Hochstrasser et al., Hoppe Seylers Z Physiol. Chem., 1975, 356(12):1865-1877; Hochstrasser and Fritz, Hoppe Seylers Z Phyiol. Chem., 1975, 356(10):1659-1662; Hochstrasser et al., Comp. Biochem. Physiol. B., 1993, 106(1): 103-108; Hochstrasser et al., Comp. Biochem. Physiol. B., 106(1):103-108).

Kazal domains generally consist of 50 to 60 amino acid residues, although a protein having a single domain of only 36 residues has recently been described (Nirmala, X. *et al.* Eur. J. Biochem. (2001) 268, 2064-2073.

The structure of a Kazal domain includes 2 short α-helices and a 3-stranded anti-parallel β-sheet. These structures form 3 loops (A, B and C) that are stabilised and closed into rings by disulfide bonds between cysteine residues (Laskowski, M. & Kato, I. (1980) Protein inhibitors of proteinases. Ann. Rev. Biochem. 49, 593-626). Typically, Kazal domains

contain six cysteine residues. The A-ring is closed by a disulfide bond between CysI-CysV and the A- and B-rings are separated by the CysII-CysIV bridge. The CysI-CysV bridge and CysII-CysIV bridge also serve to link the N-terminus and the reactive site region, respectively, to the central α-helix. The C-ring is closed by a disulfide bridge between CysIII-CysVI, which links the C-terminus to the central β-sheet. The number of amino acid residues in the A-loop between CysI and CysII varies from 18 in avian ovomucoids to just one in the insect inhibitors. Generally, the C-loop is between 31 and 34 residues. However, bdellin-B3 from leech contains a C-loop of just 27 residues (Fritz, H. & Krejcu, K. (1976) Trypsin-plasmid inhibitors (bdellins) from leeches. In Methods in Enzymology (L. Lorand, ed.), 45, pp. 797-806. Academic Press, New York, USA).

An Arg at P_1 has been shown to confer specificity for inhibition of trypsin whilst preventing interaction with trypsinogen. The presence of a Met or Tyr at P_1 has been shown to confer specificity for chymotrypsin-subilisin-elastase (Laskowski & Kato, Ann. Rev. Biochem., (1980) 49:593-626).

Many serine protease inhibitors play a role in disease processes.

Manipulation of serine protease inhibitor function is a means to alter the disease phenotype. As such, the identification of novel serine protease inhibitors is highly relevant for the treatment and diagnosis of disease, particularly those listed in Table 1.

Inhibitors can be engineered to be specific for some serine proteases only - similar enzymes involved in different contexts can remain unaffected. Unique structural features have been identified in some naturally occurring serine protease inhibitors and a range of pentapeptide analogues that incorporated these features have been synthesized. These analogues have a range of selective actions against trypsin, chymotrypsin and kallikrein.

Over the past five years, many groups have reported pre-clinical results with direct-acting thrombin inhibitors and several of these are now moving into clinical trials. In addition, many groups have disclosed the discovery of potent, orally bioavailable factor Xa inhibitors. Several of these compounds are now in early clinical trials and the results are forthcoming.

Alteration of the activity of serine protease inhibitors is thus a means to alter the phenotype of diseases in which these inhibitors are implicated. As such, identification of novel serine

protease inhibitors is highly relevant for the treatment and diagnosis of disease, particularly those identified in Table 1 below.

Table 1. Serine Protease Inhibitors and Disease

Serine Protease Inhibitor	**Disease**
Neuroserpin	Familial dementia (encephalopathy) (Davis RL et al 1999, Nature. 23;401(6751):376-9).
	Alzheimer's disease and transmissible prion encephalopaties (Carrell RW and Gooptu B 1998 Curr Opin Struct Biol 8:799-809)
	Cerebral ischemia/reperfusion, stroke (Yepes M et al. 2000 Blood 96(2):569-76.)
Alpha-1-antitrypsin	Congenital Alpha 1-antitrypsin (alpha 1AT) deficiency - (Familial) Emphysema (Stein, PE and Carell, RW 1995 Nat Struct Biol 2:96-113). Lung disease including cystic fibrosis (Doring G. 1999 Pediatr Pulmonol 28(5): 363-75), adult respiratory distress syndrome (Gadek JE *et al* 1996 Chest 110(6 Suppl):273S-277S), COPD (Eur Respir J 1997 10(6):1380-91) and asthma (Hum Mutat 2001 17(2):156).
	Blood clotting disorder: various disorders affecting medium sized arteries-intracranial aneurysms, cervicocephalic arterial dissections and fibromuscular dysplasia (SchievinkWI *et al* 1998 Neurosurgery 43(2):229-33).
	Liver disease: hepatitis (Stein PE and Carell RW 1995 Nat Struct Biol 2:96-113), juvenile cirrhosis (Parmar JS and Lomas DA 2000 J R Coll Physisicians Lond 34(3):295-300) hepatocellular carcinoma (Perlmutter DH 2000 Clin Liver Dis 4(2):387-408 vi).
Antithrombin	Cardiovascular disease (thrombosis), hereditary thrombophilia, bleeding (Stein, PE and Carell, RW 1995 Nat Struct Biol 2:96-113).

C1-inhibitor	Heriditary angioedema (Stein, PE and Carell, RW 1995 Nat Struct Biol 2(2):96-113) and acquired angiodema (Valsecchi R et al 1997 Dermatology 195(2):169-72).
	Hereditary angioneurotic edema (Farkas H et al 2000 Orv Hetil 141(47):2541-7).
	Capillary leak syndrome (CLS) (Stiller B *et al* 2001 Intensive Care Med, 27(1):193-200).
Angiotensinogen	Essential hypertension (Brand E *et al* 2000 Herz 25(1):15-25) pregnancy-induced hypertension (preeclampsia), coronary atherosclerosis, increased risk of coronary heart disease (CHD), myocardial infarction (MI) (Chen D *et al* 1998 Zhonghua Yi Xue Yi Chuan Xue Za Zhi 10;15(3): 133-5).
Antiplasmin	Bleeding (Stein PE and Carell, RW 1995 Nat Struct Biol 2(2):96-113).
Alpha-1-antichymotrypsin	Allergenic hypersensitivity (Carrell RW and Coughlin PB MB Encyclopedia) (Malerba G, *et al* J Allergy Clin Immunol 2001 107(4):654-8).
Maspin	Breast cancer (Hendrix, MJC 2000 Nature Medecine 6(4):374-376).
Nexin-1	Male (in)fertility (Murer V *et al* 2001 Proc Natl Acad Sci USA 98(6):3029-33)
Serpin proteinase inhibitor 9	Chronic inflammatory diseases (Young JL et al 2000 J Exp Med 191 (9):1535-44).

Further, a recent review by Baker et al (Baker, A.H. et al., J. Cell. Sci., (2002) Oct 1; 115 (Pt 19): 3719-3727) discusses the use of protease inhibitors as potential cancer therapeutics due to their ability to inhibit metalloproteinases. One of the inhibitors discussed is a Kazal domain-containing protein called RECK (reversion inducing cysteine rich protein with Kazal motifs) that is downregulated in tumours. Thus Kazal domain-containing proteins are potentially useful agents in the treatment and diagnosis of cancer.

Preferably the "Kazal family of serine protease inhibitors" may be a molecule containing a

Kazal domain detected with an e-value lower than 0.1, 0.01, 0.001, 0.0001, 0.0002, 0.000001 or 0.0000001.

Preferably the "Kazal family of serine protease inhibitors" 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, 0.000001 or 0.0000001.

Preferably, a polypeptide according to any of the above-described aspects of the invention acts as a member of the Kazal family of serine protease inhibitors. By "acts as a member of the Kazal family of serine protease inhibitors" we refer to polypeptides that comprise a Kazal domain and exhibit serine protease inhibitor activity. We also refer to polypeptides that comprise amino acid sequence or structural features that can be identified as conserved features within the polypeptides of this family. In particular, we refer to the presence of cysteine residues in specific positions within the polypeptide that allow the formation of intra-domain disulphide bonds.

Although the Applicant does not wish to be bound by this theory, it is postulated that the N-terminal Kazal domain of the INSP170 and INSP223 polypeptides of the present invention show specificity for trypsin since the P1 residue is arginine. Again, although the Applicant does not wish to be bound by this theory, it is postulated that the C-terminal Kazal domain of the INSP170, INSP222 and INSP223 polypeptides of the present invention show specificity for chymotrypsin or chymotrypsin-like enzymes since the C-terminal Kazal domain contains a methionine at the P1 position (Nirmala, X. *et al.*, Eur. J. Biochem., (2001) Apr, 268(7):2064-73).

An example of a method in which the activity of a polypeptide according to the present invention can be determined is by measuring inhibition of a target serine protease. For example, if trypsin is used as the target enzyme of the polypeptide according to the present invention, its inhibition can be measured with the azocoll assay (Chavira, R.J. et al., (1984) Assaying proteinases with azocoll. Anal. Biochem., 136, 446-450). This assay can be modified for microplate ELISA photometric readers (see Wedde, M. et al. (1998), Eur. J. Biochem., 255, 535-543). This involves evaporating HPLC fractions to dryness, dissolving the residues and serially diluting them in 50 mM tris/HCl buffer, pH 7.5. Azocoll is then washed (75 mg per 10 mL) in 50 mM tris/HCl buffer, pH 7.5, for 2 hours, followed by centrifugation and resuspension in fresh buffer (15 mg/mL). The analysed protein samples are then pipetted into 96-well microplates (100 µL per well). After

addition of 100 µL bovine trypsin solution (2.5 µg/mL) and 100 µL homogenous azocoll suspension to each well, the plate is incubated at 37°C for 2 hours, with shaking. Azocoll is allowed to settle for 15 minutes and 100 µL aliquots of clear supernatant from the reaction mixture are then transferred onto another plate for optical density measurements at 490 nm. Appropriate negative controls without trypsin and positive controls without the inhibitor should also be assayed simultaneously and the difference in absorbance between the positive and negative control is then taken as 100% proteinase activity. A decrease of this activity would indicate that the respective fraction contained a polypeptide according to the present invention having serine proteinase inhibitor activity.

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 tags 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 INSP170, INSP222 or INSP223 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 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.

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 molecules are specifically excluded from the scope of the invention. Preferably, genomic DNA larger than 10 kbp (kilo baise 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 nucleic acid molecule comprises comprises the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP170 protein sequence exon 1), SEQ ID NO:3 (encoding the INSP170 protein sequence exon 2), SEQ ID NO:5 (encoding the INSP170 protein sequence exon 3), SEQ ID NO:7 (encoding the INSP170 protein sequence exon 4), SEQ ID NO:9 (encoding the INSP170 protein sequence exon 5), SEQ ID NO:11 (encoding the INSP170 full protein sequence), SEQ ID NO:13 (encoding the INSP170 mature protein sequence exon 1), SEQ ID NO:15 (encoding the INSP170 full mature protein sequence), SEQ ID NO:17 (encoding the INSP222 protein sequence exon 1), SEQ ID NO:19 (encoding the INSP222 protein sequence exon 2), SEQ ID NO:21 (encoding the INSP222 protein sequence exon 3), SEQ ID NO:23 (encoding the INSP222 protein sequence exon 4), SEQ ID NO:25 (encoding the INSP222 protein sequence exon 5), SEQ ID NO:27 (encoding the INSP222 full protein sequence), SEQ ID NO:29 (encoding the INSP223 protein sequence exon 1), SEQ ID NO:31 (encoding the INSP223 protein sequence exon 2), SEQ ID NO:33 (encoding the INSP223 protein sequence exon 3), SEQ ID NO:35 (encoding the INSP223 protein sequence exon 4), SEQ ID NO:37 (encoding the INSP223 protein sequence exon 5), SEQ ID NO:39 (encoding the INSP223 protein sequence exon

6), SEQ ID NO:41 (encoding the INSP223 full protein sequence), SEQ ID NO:43 (encoding the INSP223 mature protein sequence exon 1) and/or SEQ ID NO:45 (encoding the INSP223 full mature protein sequence) or is a redundant equivalent or fragment of any one of these sequences.

The invention further provides that the purified nucleic acid molecule consists of the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the INSP170 protein sequence exon 1), SEQ ID NO:3 (encoding the INSP170 protein sequence exon 2), SEQ ID NO:5 (encoding the INSP170 protein sequence exon 3), SEQ ID NO:7 (encoding the INSP170 protein sequence exon 4), SEQ ID NO:9 (encoding the INSP170 protein sequence exon 5), SEQ ID NO:11 (encoding the INSP170 full protein sequence), SEQ ID NO:13 (encoding the INSP170 mature protein sequence exon 1), SEQ ID NO:15 (encoding the INSP170 full mature protein sequence), SEQ ID NO:17 (encoding the INSP222 protein sequence exon 1), SEQ ID NO:19 (encoding the INSP222 protein sequence exon 2), SEQ ID NO:21 (encoding the INSP222 protein sequence exon 3), SEQ ID NO:23 (encoding the INSP222 protein sequence exon 4), SEQ ID NO:25 (encoding the INSP222 protein sequence exon 5), SEQ ID NO:27 (encoding the INSP222 full protein sequence), SEQ ID NO:29 (encoding the INSP223 protein sequence exon 1), SEQ ID NO:31 (encoding the INSP223 protein sequence exon 2), SEQ ID NO:33 (encoding the INSP223 protein sequence exon 3), SEQ ID NO:35 (encoding the INSP223 protein sequence exon 4), SEQ ID NO:37 (encoding the INSP223 protein sequence exon 5), SEQ ID NO:39 (encoding the INSP223 protein sequence exon 6), SEQ ID NO:41 (encoding the INSP223 full protein sequence), SEQ ID NO:43 (encoding the INSP223 mature protein sequence exon 1) and/or SEQ ID NO:45 (encoding the INSP223 full mature protein sequence) or is a redundant equivalent or fragment of any one of these sequences.

Preferably, the purified nucleic acid according to the above-described aspect of the invention is a nucleic acid having the sequence of SEQ ID NO:11 (encoding the INSP170 full length protein sequence), SEQ ID NO:15 (encoding the INSP170 mature full length protein sequence), SEQ ID NO:27 (encoding the INSP222 full protein sequence), SEQ ID NO:41 (encoding the INSP223 full protein sequence) or SEQ ID NO:45 (encoding the INSP223 full mature protein sequence).

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 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.

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 a member of the Kazal family of serine protease inhibitors of the first aspect of the invention. Preferably, the ligand inhibits the function of a polypeptide of the first aspect of the invention which is a member of the Kazal family of serine protease inhibitors. 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.

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.

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

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 INSP170, INSP222 and INSP223 polypeptides allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease. As used herein, the term "disease" also includes disorders. 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.

Another aspect of this invention resides in the use of an INSP170, INSP222 or INSP223 gene or polypeptide as a target for the screening of candidate drug modulators, particularly candidate drugs active against Kazal domain related disorders.

A further aspect of this invention resides in methods of screening of compounds for therapy of Kazal domain related disorders, comprising determining the ability of a compound to bind an INSP170, INSP222 or INSP223 gene or polypeptide, or a fragment thereof.

A further aspect of this invention resides in methods of screening of compounds for therapy of Kazal domain related disorders, comprising testing for modulation of the activity of an INSP170, INSP222 or INSP223 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 of diseases in which members of the Kazal family of serine protease inhibitors are implicated "Kazal domain related disorder". Such diseases may include cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma; autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation and organ transplant rejection; cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, and ischemia; neurological disorders including central nervous system disease, brain injury, amyotrophic lateral sclerosis, and pain; developmental disorders; metabolic disorders including diabetes mellitus, osteoporosis, and obesity, AIDS and renal disease; infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions. Preferably, such diseases include chronic inflammatory diseases such as chronic pancreatitis, including idiopathic, hereditary and alcoholic chronic pancreatitis, fibrocalculous pancreatic diabetes, asthma, interstitial lung disease, psoriasis, arthritis gingivitis, peridontitis, cancers such as genitourinary tumours including ovarian cancer, prostate cancer and also breast cancer and esophageal

cancer, Netherton Syndrome, fertility disorders including male (in)fertility and female (in)fertility, arthritis, Alzheimer's disease and transmissible prion encephalopathies, liver disease including hepatitis, juvenile cirrhosis and hepatocellular carcinoma, cardiovascular disease, hereditary thrombophilia, bleeding, coronary atherosclerosis, increased risk or coronary heart disease, myocardial infarction, hereditary angioedema, acquired angioedema, hereditary angioneurotic edema, capillary leak syndrome, essential hypertension, pregnancy-induced hypertension (preeclampsia), allergenic hypersensitivity, cerebral ischemia/reperfusion, hemorrhagic stroke, Parkinson's disease, chronic obstructive pulmonary disease, congenital alpha 1-antitrypsin deficiency, familial emphysema, lung disease including cystic fibrosis, adult respiratory distress syndrome, chronic obstructive pulmonary disease, blood clotting disorders, various disorders affecting medium sized arteries including intracranial aneurysms and cervicocephalic arterial dissections, fibromuscular dysplasia, Hepatitis C, familial dementia, and other disorders of coagulation and fibrinolysis. Preferably, the diseases are those in which members of the Kazal family of serine protease inhibitors are implicated. According to the eighth aspect of the invention, there is also provided 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 contraception. These moieties of the first, second, third, fourth, fifth, sixth or seventh aspect of the invention may also be used in the manufacture of a medicament for the treatment of such diseases or for contraception.

Thus, embodiments of this invention provide for the use of the polypeptides in inflammatory diseases, autoimmue diseases, encephalopathies, cerebrovascular/cardiovascular diseases, blood clotting disorders, cancer, lung diseases, liver diseases or male infertility.

In one embodiment, the encephalopathy is selected from dementia, Alzheimer's disease or transmissible prion encephalopaty

In one embodiment, the cerebrovascular disease is selected from cerebral ischemia/reperfusion or stroke

In one embodiment, the cardiovascular disease is selected from thrombosis, hereditary thrombophilia, bleeding, essential hypertension, pregnancy-induced hypertension

(preeclampsia), coronary atherosclerosis, coronary heart disease (CHD) or myocardial infarction (MI).

In one embodiment, the lung disease is selected from cystic fibrosis, adult respiratory distress syndrome, COPD, asthma or emphysema.

In one embodiment, the blood clotting disorder is selected from disorders affecting medium sized arteries, intracranial aneurysms, cervicocephalic arterial dissections or fibromuscular dysplasia.

In one embodiment, the liver disease is selected from hepatitis, juvenile cirrhosis or hepatocellular carcinoma.

In one embodiment, the inflammatory disease is selected from angioedema, allergenic hypersensitivity, inflammatory skin diseases or chronic inflammatory diseases. Preferably, the angioedema is selected from heriditary angioedema, acquired angioedema or hereditary angioneurotic edema. Preferably, the inflammatory skin disease is selected from atopic dermatitis or psoriasis

In one embodiment, cancer is selected from breast cancer or prostate cancer.

In one embodiment, the autoimmune disease is selected from rheumatoid arthritis, Crohn's disease, Multiple Sclerosis, inflammatory bowel disease (IBD), systemic lupus erithematosus (SLE), Sjogren's syndrome or myasthenia gravis.

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) of prostate cancer and male (in)fertility disorders.

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

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

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

Preferably, the disease diagnosed by a method of the ninth aspect of the invention is a disease in which members of the Kazal family of serine protease inhibitors are implicated, as described above.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as a Kazal family of serine protease inhibitors. Suitable uses of the polypeptides of the invention as Kazal family of serine protease inhibitors include screening methods to identify ligands and other agonist or antagonist molecules that are useful in therapy and diagnosis of diseases and conditions in which this category of protein is implicated.

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 cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma; autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation and organ transplant rejection; cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, and ischemia; neurological disorders including central nervous system disease, brain injury, amyotrophic lateral sclerosis, and pain; developmental disorders; metabolic disorders including diabetes mellitus, osteoporosis, and obesity, AIDS and renal disease; infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions.

Preferably, the disease is a disease in which members of the Kazal family of serine protease inhibitors are implicated, as described above.

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.

Preferably, the disease is a disease in which members of the Kazal family of serine protease inhibitors are implicated, as described above.

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 INSP170, INSP222 and INSP223 polypeptides are Kazal domain containing proteins and thus have roles in many disease states. Antagonists of the INSP170, INSP222 and INSP223 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 used 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, analogues, 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 INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 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 polypeptides that reflect their identity as novel members of the Kazal family of serine protease inhibitors.

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

Standard abbreviations for nucleotides and amino acids are used in this specification.

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

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985);

Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

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 INSP170, INSP222 or INSP223 is a fusion protein.

These fusion proteins can be obtained by cloning a polynucleotide encoding a polypeptide comprising a sequence having at least 85% of homology with INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 polypeptide. Examples of heterologous sequences, that can be comprised in the fusion proteins either at the N- or C-terminus, include: extracellular domains of membrane-bound protein, immunoglobulin constant regions (Fc regions), multimerization domains, domains of extracellular proteins, signal sequences, export sequences, and sequences allowing purification by affinity chromatography.

Many of these heterologous sequences are commercially available in expression plasmids since these sequences are commonly included in fusion proteins in order to provide additional properties without significantly impairing the specific biological activity of the protein fused to them (Terpe K, 2003, Appl Microbiol Biotechnol, 60:523-33). 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.* 1989, Proc Natl Acad Sci USA, 86:821-4) or by the "HA" tag, an epitope derived from the influenza hemagglutinin protein (Wilson *et al.* 1994, Cell, 37:767-78). 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 INSP170, INSP222 or INSP223 polypeptides may be purified by means of a hexa-histidine peptide fused at the C-terminus of INSP170, INSP222 or INSP223. 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:47) 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 (*i.e.* an increased 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 INSP170, INSP222 and INSP223 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 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the INSP170, INSP222 and INSP223 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 Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic

residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. Such mutants also include polypeptides in which one or more of the amino acid residues includes a substituent group.

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 2.

Specific, non-conservative mutations can be also introduced in the polypeptides of the invention with different purposes. Mutations reducing the affinity of the PMP-22/EMP/MP20/Claudin-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 3. A non-exhaustive list of amino acid derivatives also

include aminoisobutyric acid (Aib), hydroxyproline (Hyp), 1,2,3,4-tetrahydro-isoquinoline-3-COOH, indoline-2carboxylic 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 INSP167 or INSP170 polypeptides, or with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, 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 Threader technology that forms one aspect of the search tools used to generate the Biopendium™ search database may be used (see PCT application WO 01/69507) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the INSP170, INSP222 and INSP223 polypeptides, are predicted to be members of the Kazal family of serine protease inhibitors, by virtue of sharing significant structural homology with the INSP170, INSP222 and INSP223 polypeptide sequences. By "significant structural homology" is

meant that the Inpharmatica Genome Threader predicts two proteins to share structural homology with a certainty of 10% and above.

The polypeptides of the first aspect of the invention also include fragments of the INSP170, INSP222 and INSP223 polypeptides and fragments of the functional equivalents of the INSP170, INSP222 and INSP223 polypeptides, provided that those fragments are members of the Kazal family of serine protease inhibitors or have an antigenic determinant in common with the INSP170, INSP222 or INSP223 polypeptides.

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 INSP170, INSP222 and INSP223 polypeptides or one of their 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. Fragments according to the invention may be 1-100 amino acids in length, preferably, 5-50, preferably 7-20, more preferably 10-15 amino acids in length.

Nucleic acids according to the invention are preferably 5-400 nucleotides in length, preferably 10-300 nucleotides, preferably 15-250, preferably 20-200, preferably 30-150, preferably 50-10 nucleotides in length. Polypeptides according to the invention are preferably 5-150 amino acids in length, preferably 10-100, preferably 15-75, preferably 25-50 amino acids in length.

Fragments of the full length INSP170, INSP222 and INSP2223 polypeptides may consist of combinations of 2, 3, 4 or all 5 neighbouring exon sequences in the INSP170 polypeptide sequence, or 2, 3, 4 or all 5 neighbouring exon sequences in the INSP222 polypeptide sequence, or 2, 3, 4, 5 or all 6 neighbouring exon sequences in the INSP223 polypeptide sequence.

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 known members of the Kazal family of serine protease inhibitors.

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 known members of the Kazal family of serine protease inhibitors.

Preferably, there is a measurable increase in the affinity for a polypeptide of the invention as compared with known members of the Kazal family of serine protease inhibitors.

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

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

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

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

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

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

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554;

Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

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

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode a polypeptide sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30 SEQ ID NO:32 SEQ ID NO:34 SEQ ID NO:36 SEQ ID NO:38 SEQ ID NO:40 SEQ ID NO:42 SEQ ID NO:44 and/or SEQ ID NO:46 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 a polypeptide of this invention may be identical to the coding sequence of one or more of the nucleic acid molecules disclosed herein.

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 and/or SEQ ID NO:46. Such nucleic acid molecules may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or prepro-polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include additional sequences which encode additional amino acids, such as those which provide additional functionalities.

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 molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

"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 the INSP170, INSP222 or INSP223 polypeptides and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to such coding sequences, or is 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%, 99% or more 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 INSP170, INSP222 or INSP223 polypeptides.

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 INSP167 or INSP170 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the genes 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 INSP170, INSP222 or INSP223 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:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43 or SEQ ID NO:45) 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 nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the

polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

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, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

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

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

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. The vectors pCR4-TOPO-INSP187-EC-SV2, pENTR_INSP187ECSV2-6HIS, pEAK12d_INSP187ECSV2-HIS and pDEAT12.2_INSP187ECSV2-6HIS are preferred examples of suitable vectors for use in accordance with the aspects of this invention relating to INSP222.

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

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

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for

secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

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

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

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

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

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

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

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

In particular, all plants from which protoplasts can be isolated and cultured to give whole

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

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

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

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

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

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

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence

activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) J. Exp. Med, 158, 1211-1216).

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

Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

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

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

employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

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

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

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 INSP170, INSP222 or INSP223 polypeptide as a target for screening candidate drugs for treating or preventing Kazal domain related disorders.

Another object of this invention resides in methods of selecting biologically active compounds, said methods comprising contacting a candidate compound with an INSP170, INSP222 or INSP223 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 an INSP170, INSP222 or INSP223 polypeptide with a candidate compound, and selecting compounds that bind said INSP170, INSP222 or INSP223 polypeptide at the surface of said cells and/or that modulate the activity of the INSP170, INSP222 or INSP223 polypeptide.

A "biologically active" compound denotes any compound having biological activity in a subject, preferably therapeutic activity, more preferably a compound having Kazal domain activity, and further preferably a compound that can be used for treating INSP170, INSP222 or INSP223 related disorders, or as a lead to develop drugs for treating Kazal domain related disorders. A "biologically active" compound preferably is a compound that modulates the activity of INSP170, INSP222 or INSP223.

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 Kazal domain related disorder, such as an animal model.

Preferred selected compounds are agonists of INSP170, INSP222 or INSP2230, i.e., compounds that can bind to INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 polypeptide according to the present invention and determining the ability of said test compound to modulate the activity of said INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 gene according to the present invention and determining the ability

of said test compound to modulate the expression of said INSP170, INSP222 or INSP223 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 INSP170, INSP222 or INSP223 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 al., Current Protocols in Immunology 1(2):Chapter 5 (1991).

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 Kazal domain 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, immobilised 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 INSP170, INSP222 or INSP223 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 an INSP170, INSP222 or INSP223 polypeptide) or any cell that expresses an endogenous INSP170, INSP222 or INSP223 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, Kluyveromyces* 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 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 a 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 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 an agonist or antagonist of a polypeptide of the present invention comprises:

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

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

(a) incubating a labelled ligand with a 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.

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.

The INSP170, INSP222 and INSP223 polypeptides of the present invention may modulate cellular growth and differentiation. Thus, the biological activity of the INSP170, INSP222 and INSP223 exon polypeptides and the INSP170, INSP222 and INSP223 polypeptides can be examined in systems that allow the study of cellular growth and differentiation such

as *in vitro* tissue culture. Stimulation or inhibition of cellular proliferation may be measured by a variety of assays.

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

The effect of the polypeptide upon cellular differentiation can be measured by contacting stem cells or embryonic cells with various amounts of the INSP170, INSP222 or INSP223 exon polypeptides and the INSP170, INSP222 or INSP223 polypeptides and observing the effect upon differentiation of the stem cells or embryonic cells. Tissue-specific antibodies and microscopy may be used to identify the resulting cells.

The INSP170, INSP222 or INSP223 exon polypeptides and the INSP170, INSP222 or INSP223 polypeptides may also be found to modulate immune and/or nervous system cell proliferation and differentiation in a dose-dependent manner in the above-described assays. Thus, the "functional equivalents" of the INSP170, INSP222 or INSP223 exon polypeptides and the INSP170, INSP222 and INSP223 polypeptides include polypeptides that exhibit any of the same growth and differentiation regulating 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 INSP170, INSP222 or INSP223 exon polypeptides and the INSP170, INSP222 or INSP223 polypeptides, preferably the "functional equivalents" will exhibit substantially similar dose-dependence in a given

activity assay compared to the INSP170, INSP222 or INSP223 exon polypeptides and the INSP170, INSP222 or INSP223 polypeptides.

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.

Screening assays known in the art that are suitable for the identification of modulators of the polypeptides of this invention include the GAL4 assay (for example, Jurgen M. Lehmann *et al.*, J. Biol. Chem., Jun 1995; 270: 12953- 12956; Raivio T. *et al.*, J. Clin. Endocrinol. Metab. 86 (2001) 1539- 44), biophysical techniques such as surface plasmon resonance (supplied by Biacore AB, Uppsala, Sweden) and assays based on fluorescence polarisation or FRET for the ligand-induced binding of the co-activator peptide (Mukherjee R. *et al.*, J Steroid Biochem. Mol. Biol. 2002 Jul;81(3):217-25).

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 INSP170, INSP222 or INSP223 polypeptide or fragment thereof, whereby the fragment is preferably a INSP170, INSP222 or INSP223 gene-specific fragment, for isolating or generating an agonist or stimulator of the INSP170, INSP222 or INSP223 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% or even 99% by weight.

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

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

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

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

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

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

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

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a

second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

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

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

RNA molecules may be modified to increase intracellular stability and half-life. Possible

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

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

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

Gene therapy of the present invention can occur *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 Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest.

These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

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

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

Vaccines according to the invention may either be prophylactic (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 a disease;

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

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

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

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

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

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

through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

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

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

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

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee *et al*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) Proc. Natl. Acad. Sci. 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/25116 (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 in which members of the Kazal family of serine protease inhibitors are implicated, as described

above.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the INSP170, INSP222 and INSP223 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: Top ten results from BLAST against NCBI non-redundant database using SEQ ID NO:26 (INSP170 polypeptide sequence).

Figure 2: Alignment generated by BLAST between SEQ ID NO:26 and the top hit, (P08480) double-headed protease inhibitor, submandibular gland, from cat.

Figure 3: Alignment of the INSP170 polypeptide sequence (SEQ ID NO:26) to secreted double-headed ("Bikazal") Kazal domain-containing protein family members.

Figure 4: NCBI-CDD analysis of INSP170; and;

Figure 5: Signal P analysis of INSP170.

Figure 6: INSP222 DNA sequence. The position and sense of the PCR primers are indicated by arrows.

Figure 7: The nucleotide sequence with translation of the INSP222 PCR product cloned using INSP170fwd and INSP170rev PCR primers.

Figure 8: ClustalW alignment of the nucleotide sequence of the INSP222 cloned sequence versus INSP222 predicted sequence.

Figure 9: Multiple sequence alignment of INSP170, INSP221 and INSP222 full length polypeptides. Sequence features are highlighted.

Figure 10: Exon Structure of INSP170, INSP221 and INSP222

Figure 11: Expression reports for Exon 1 of INSP170.

Figure 12: Expression reports for Exon 2 of INSP170.

Figure13: Expression reports for Exon 3 of INSP170.

Figure 14: Expression reports for Exon 4 of INSP170.

Figure 15: Expression reports for Exon 5 of INSP170.

Table 2

Amino Acid	Synonymous Groups	More Preferred Synonymous Groups
Ser	Gly, Ala, Ser, Thr, Pro	Thr, Ser
Arg	Asn, Lys, Gln, Arg, His	Arg, Lys, His
Leu	Phe, Ile, Val, Leu, Met	Ile, Val, Leu, Met
Pro	Gly, Ala, Ser, Thr, Pro	Pro
Thr	Gly, Ala, Ser, Thr, Pro	Thr, Ser
Ala	Gly, Thr, Pro, Ala, Ser	Gly, Ala
Val	Met, Phe, Ile, Leu, Val	Met, Ile, Val, Leu
Gly	Ala, Thr, Pro, Ser, Gly	Gly, Ala
Ile	Phe, Ile, Val, Leu, Met	Ile, Val, Leu, Met
Phe	Trp, Phe,Tyr	Tyr, Phe
Tyr	Trp, Phe,Tyr	Phe, Tyr
Cys	Ser, Thr, Cys	Cys
His	Asn, Lys, Gln, Arg, His	Arg, Lys, His
Gln	Glu, Asn, Asp, Gln	Asn, Gln
Asn	Glu, Asn, Asp, Gln	Asn, Gln
Lys	Asn, Lys, Gln, Arg, His	Arg, Lys, His
Asp	Glu, Asn, Asp, Gln	Asp, Glu
Glu	Glu, Asn, Asp, Gln	Asp, Glu
Met	Phe, Ile, Val, Leu, Met	Ile, Val, Leu, Met
Trp	Trp, Phe,Tyr	Trp

Table 3

Amino Acid	Synonymous Groups
Ser	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Arg	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-.Met, D-Ile, Orn, D-Orn
Leu	D-Leu, Val, D-Val, AdaA, AdaG, Leu, D-Leu, Met, D-Met
Pro	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Thr	D-Thr, Ser, D-Ser, allo-Thr, Met,D-Met, Met(O), D-Met(O), Val, D-Val
Ala	D-Ala, Gly, Aib, B-Ala, Acp, L-Cys, D-Cys
Val	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met, AdaA, AdaG
Gly	Ala, D-Ala, Pro, D-Pro, Aib, .beta.-Ala, Acp
Ile	D-Ile, Val, D-Val, AdaA, AdaG, Leu, D-Leu, Met, D-Met
Phe	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
Tyr	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Cys	D-Cys, S--Me--Cys, Met, D-Met, Thr, D-Thr
Gln	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Asn	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Lys	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Asp	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Glu	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Met	D-Met, S--Me--Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val

Examples

Example 1: INSP170 Protein BLAST Results

The INSP170 polypeptide sequence, shown in SEQ ID NO:12, was used as a BLAST query against the NCBI non-redundant sequence database. The top ten hits are shown in Figure 1. The top hit is to a double-headed protease inhibitor found in the submandibular gland of cat. The alignment of INSP170 with this top hit is shown in Figure 2. INSP170 has 59% identity over its 143 residues to the top hit. The top hit has a highly significant expectation value of $1e^{-32}$.

The fact that all the top ten hits are from members of the Kazal family of serine protease inhibitors having two Kazal domains (i.e. double-headed serine protease inhibitors), together with the highly significant expectation values given for these hits (the top ten hits all have expectation values of below $2e^{-26}$), indicates that INSP170 is also a double-headed member of the Kazal family of serine protease inhibitors.

An alignment of INSP170 (SEQ ID NO:12) to secreted double-headed ("bikazal") Kazal domain containing protein family members is shown in Figure 3. INSP170 shows strong homology to known double-headed ("bikazal") Kazal domain-containing secreted proteins. The six structurally significant cysteine residues found in Kazal domains are also present in INSP170 and are highlighted by asterisks in Figure 3. The residues highlighted by '#' represent the P1 residues from the reactive sites of each Kazal domain. As mentioned above, P1 residues are believed to confer proteinase inhibitor specificity. The N-terminal Kazal domain of INSP170 contains an arginine at the P1 position, suggesting that this domain inhibits trypsin and trypsin-like enzymes. The C-terminal Kazal domain of INSP170 contains a methionine at the P1 position, suggesting that this domain inhibits chymotrypsin and chymotrypsin-like enzymes, (Nirmala, *et al* 2001).

The INSP170 polypeptide sequence (SEQ ID NO:12) was analysed against the NCBI Conserved Domain Database http://www.ncbi.nih.gov/Structure/cdd/cdd.shtml. The results are shown in Figure 4. Analysis against the Smart, Pfam, CDD and SPAN collections all returned matches to proteins having Kazal domains with significant E-values. This supports the finding that the INSP170 polypeptide also contains two Kazal domains and is thus a double-headed member of the Kazal family of serine protease inhibitors.

SignalP analysis (Nielsen *et al.* 1997 Protein Eng 1:1-6) suggests that amino acid residues 1-18 of the INSP170 polypeptide (SEQ ID NO:12) form an N-terminal signal peptide (see Figure 5).

In summary, these results indicate that the INSP170 polypeptide is a double-headed member of the Kazal family of serine protease inhibitors and thus acts as a member of the Kazal family of serine protease inhibitors.

Example 2: Autoimmunity Assays Suitable for Exploration of the Biological Relevance of INSP170, INSP222 and INSP223 Function

A number of autoimmunity-related assays have been developed by the Applicant and are of use in the investigation of the biological relevance of INSP170, INSP222 and INSP223 function. 13 cellular assays with 10 different read-outs have been developed. These assays address generic biological responses such as survival and proliferation as well as specific cellular responses such as cytokine secretion or calcium mobilization.

The majority of the assays focus on the major immune system cell types, namely T and B lymphocytes, macrophages and neutrophils. Further assays concentrate on the biology of microglial cells.

Assays targeting endothelial cell responses use cultures of primary endothelial cells such as human embryonic umbilical endothelial cells (HUVEC).

Further assays are envisaged which are more specific to diseases such as rheumatoid arthritis, Crohn's disease and Multiple Sclerosis to investigate the therapeutic relevance of selected proteins.

A. Assays targeting T lymphocyte responses

i) Fas-Ligand-induced T cell death assay:

This assay will reveal new modulators of receptor mediated cell death. In this assay, T cell apoptosis is induced by stimulating Jurkat cells (a human T cell line) with recombinant 6 Histidine-tagged Fas Ligand combined with a monoclonal anti 6-his antibody. Death is quantified by release of LDH, a cytoplasmic enzyme released in the culture medium when cells are dying. The read out is a colorimetric assay read at 490nm. T cells have been shown to be pathogenic in many autoimmune diseases and being able to control antigen-specific T cell death is a therapeutic strategy (e.g. anti-TNFα treatment in a patient with Crohn's disease).

ii) Human-MLR: proliferation and cytokine secretion:

This cell-based assay measures the effects of novel proteins on lymphocyte proliferation and cytokine secretion or inhibition upon stimulation by PBMC from another donor (alloreactivity). Proliferation and cytokine secretion are independent responses. These assays address antigen-specific T cell and antigen presenting cell functions, which are

crucial cellular responses in any autoimmune diseases. Secreted cytokines (IL-2, 4, 5, 10, TNF-α and IFN-γ) are quantified by a cytokine bead array (CBA).

iii) Mouse-MLR: proliferation.

This cell-based assay measures the effects of novel proteins on lymphocyte proliferation or inhibition of mouse spleen cells following stimulation by spleen cells from another donor (mouse strain). This cell-based assay measures the effect of novel proteins on T lymphocyte and antigen presenting cell responses and will be used to confirm activity of positives and hits identified in the human-MLR assays. This assay will be used to select proteins that will be tested in murine models of human diseases.

iv) Human PBMC stimulated with the superantigen TSST:

Superantigens are strong modulators of the immune system affecting T cells. Superantigens influence immunologically mediated disorders such as inflammatory bowel disease (IBD), inflammatory skin diseases like atopic dermatitis and psoriasis. In this cellular assay, we are specifically targeting T lymphocyte activation via the TCR but with different requirements than the T cell response to classical antigens, in particular in respect to co-stimulatory molecules.

v) Human PBMC stimulated with either ConA or PHA:

These cell-based assays measure the effects of novel proteins on cytokine secretion induced by two different stimuli acting on different cells as measured by a cytokine bead array (CBA) assay (IL-2, IFN-γ, TNF-α, IL-5, IL-4 and IL-10). Most cytokines can have dual actions, pro or anti-inflammatory, depending of the injury, milieu and cellular target. Any protein with the capability to modulate cytokine secretion may have a therapeutic potential (e.g. decreasing IFN-γ and TNF-α would be beneficial in Th1-mediated autoimmune disease. In contrast, decreasing IL-4, IL-5 may be beneficial in Th2-mediated-diseases. Inducing IL-10 would influential in MS and systemic lupus erithematosus (SLE)).

B. Assays targeting monocyte/macrophages and granulocyte responses

i) Human PBMC stimulated with LPS.

This cell-based assay measures the effects of novel proteins on cytokine secretion (IFN-γ, TNF-α) induced by LPS acting on monocytes/macrophages and granulocytes.

Any protein with the capability to modulate IFN-γ and TNF-α secretion would be beneficial in Th1-mediated autoimmune diseases.

C. Assays targeting neutrophil responses

Neutrophils are important in inflammation and autoimmune diseases such as Rheumatoid Arthritis. Leukocyte chemo-attractants such as IL-8 initiate a sequence of adhesive interactions between cells and the micro-vascular endothelium, resulting in activation, adhesion and finally migration of neutrophils. The tissue infiltration of neutrophils depends on a reorganisation of cytoskeleton elements associated with specific changes in cell morphology of these cells. This cell-based assay measures the effect of novel proteins on cytoskeleton reorganization of human neutrophils.

D. Assays targeting B lymphocyte responses

Autoantibodies, as well as infiltrating B cells, are thought to be important in the pathogenesis of various autoimmune diseases, such as systemic lupus erithematosus (SLE), rheumatoid arthritis (RA), Sjogren's syndrome and myasthenia gravis. Compelling evidence indicates that a disregulation in B cell homeostasis could affect immune tolerance leading to the inappropriate survival of autoreactive B cells producing pathogenic antibodies and sustained inflammation. The identification of new factors that play critical roles in the regulation of B cell proliferation, survival and differentiation following B cell receptor triggering are of high relevance in the development of novel therapies.

i) B cell proliferation assay:

This cell-based assay measures the effect of novel proteins on B cell survival.

ii) B cell co-stimulation assay:

This cell-based assay measures the effect of novel proteins on B cell co-stimulation.

E. Assays targeting monocytes and microglial responses

i) THP-1 calcium flux assay:

The Ca^{+}-flux in THP1-cell assay measures the effects of novel proteins on their ability to trigger an intracellular calcium release (a generic second messenger event) from the endoplasmic reticulum.

ii) Microglia cell proliferation:

During proliferation of microglial progenitors, a number of colony-stimulating factors, including some cytokines, are known to play key roles. Among them, M-CSF is crucial for the final step of maturation of macrophages/microglia and is not replaceable by any other factor. The evaluation of this biological response may represent a way to influence the microglial activity and therefore an opportunity to identify molecules with therapeutic potential fro MS.

A cell-based assay has been developed to measure the proliferative response of a microglia cell line to M-CSF. The feasibility and the robustness phases showed optimal results. This assay is in 96 well plates and is easily automated. Non-radioactive substrate is required.

F. Further assays

The usefulness of the up- or down-regulation of MHC II in a microglia cell line as a response of IFN-γ, TGFβ and IL-10 can be evaluated for high throughput screening tests. Any protein regulating MHC II expression may have an important role in controlling inflammation associated with CNS (the concept has been validated with statins).

The induction/inhibition of NO in microglia in a high throughput-screening test can be evaluated. The modulation of the production of active molecules by microglia cells may represent an opportunity to diminish the cellular damage in the CNS.

Example 3: Cloning of INSP222

3.1 Preparation of human cDNA templates

First strand cDNA was prepared from human prostate total RNA sample (Clontech) using Thermoscript RT-PCR System (Invitrogen) according to the manufacturer's protocol.

Oligo $(dT)_{20}$ primer (1μl at 500μg/ml) , 2.5μg human total RNA, 1μl 10mM dNTP mix (10mM each of dATP, dGTP, dCTP and dTTP at neutral pH) and sterile DEPC-treated water to a final volume of 12μl were combined in a 0.2ml PCR tube, heated to 65°C for 5min and chilled on ice. The contents were collected by brief centrifugation and 4 μl of 5X cDNA Synthesis Buffer, 2μl 0.1M DTT, 1μl of sterile DEPC-treated water, 1μl RnaseOUT™ Recombinant Ribonuclease Inhibitor (40units/μl, Invitrogen) and 1μl of ThermoScript RT enzyme (15units/μl) were added. The contents of the tube were mixed gently and incubated at 55°C for 1 hour and then inactivated by heating at 85°C for 5min.

To remove RNA complementary to the cDNA, 1µl (2units) of *E. coli* RNase H (Invitrogen) was added and the reaction mixture incubated at 37°C for 30min.

3.2 Gene specific cloning primers for PCR

PCR primer pairs having a length of between 18 and 30 bases were designed to amplify the predicted INSP170 cds. PCR primers were optimized to have a Tm close to 55 ± 5°C and a GC content of 40-60%. Primers were selected which had high selectivity for the target sequence (INSP170/INSP222) with little or no non-specific priming.

3.3 PCR amplification of INSP222 from human cDNA templates

Gene-specific cloning primers INSP170fwd and INSP170rev were designed to amplify cDNA fragments of 429 bp spanning the INSP170 cds and 264 bp spanning the INSP222 cds. (Table 4, Figure 6).

As INSP170 had been found to be over-expressed in prostate tissue, the primer pair was used to amplify products from prostate cDNA. PCR was performed in a final volume of 25µl containing 1X Expand High Fidelity™ buffer, 200µM dNTPs, 500nM of each cloning primer, 1unit of Expand High Fidelity enzyme mix, and 1µl of human prostate cDNA. Cycling was performed using an MJ Research DNA Engine, programmed as follows: 94°C, 2min; 35 cycles of 94°C, 45sec, 56°C, 1min, and 72°C, 1min 30sec; followed by 1 cycle at 72°C for 10min and a holding cycle at 4°C.

PCR amplification reaction was visualised on a 2% agarose gel in 1 X TAE buffer (Promega). Products of approximately the expected molecular mass for INSP222 were identified. PCR reactions were then purified using the QIAquick Gel Extraction Kit (Qiagen), eluted in 30µl of EB buffer (10mM Tris) and exposed to a second round of PCR.

Product from the first round of PCR was then used as the template for the second round of PCR using the same amplification primers INSP170fwd and INSP170rev. The second PCR reaction was performed in a final volume of 25µl containing 1X Expand High Fidelity™ buffer, 200µM dNTPs, 50pmoles of each cloning primer, 2.5unit of Expand High Fidelity™ enzyme mix, (Roche), and 14 µl of gel purified PCR product from the first round of PCR. Cycling was performed using an MJ Research DNA Engine, programmed as follows: 94°C, 2min; 30 cycles of 94°C, 45sec, 56°C, 1min, and 72°C, 1 min 30 sec; followed by 1 cycle at 72°C for 10min and a holding cycle at 4°C.

10µl of the final round of PCR was visualized on a 2% agarose gel in 1 X TAE buffer (Promega). A single PCR product could be seen of approximately the expected molecular weight, 268 bp corresponding to INSP222. The band was purified from the gel using the Qiagen Gel Extraction Kit (Qiagen), eluted in 30µl of of EB buffer (10mM Tris), and subcloned directly.

3.5 Subcloning of PCR Products

The PCR products were subcloned using TA cloning into the pGEM-T Easy cloning vector purchased from the Promega Corporation using a modification of the conditions specified by the manufacturer. Briefly, 10µl of gel purified PCR product was incubated over night at 4°C with 1 µl of pGEM-T Easy vector, 2µl of T4 DNA ligase buffer, 6µl of sterile water and 1µl of T4 DNA ligase (New England Biolabs). The reaction mixture was then transformed into *E. coli* strain DH5α (Invitrogen) as follows: a 50µl aliquot of DH5α cells was thawed on ice and 5µl of pGEM-T Easy reaction was added. The mixture was incubated for 1 hour on ice and then heat shocked by incubation at 42°C for exactly 1min. Samples were returned to ice and 500µl of warm (room temperature) LB media was added. Samples were incubated with shaking (220 rpm) for 1h at 37°C. The transformation mixture was then plated on L-broth (LB) plates containing ampicillin (100µg/ml) and incubated overnight at 37°C.

3.6 Colony PCR

Colonies were inoculated into 22.5µl of PCR mix; 19.5µl of sterile water, 2.5µl of 10x Thermopol buffer (New England Biolabs) and 0.5µl of 10mM dNTPs. PCR mix was subsequently heated to 99°C for 10 min and then allowed to cool to room temperature. This was made up to 25µl volume containing 1unit of T4 DNA Polymerase (New England Biolabs), 500nM INSP170fwd and 500nM INSP170rev, and subjected to PCR using an MJ Research DNA Engine. The cycling conditions were as follows: 30 cycles of 94°C, 45sec, 52°C, 1min and 72°C for 1min. Samples were maintained at 4°C (holding cycle) before further analysis.

PCR reaction products were analyzed on 2% agarose gels in 1 X TAE buffer. Positive colonies, which gave PCR products of approximately the expected molecular weight, 264bp, were grown up overnight at 37°C in 3ml Terrific Broth containing ampicillin (100µg/ml), with shaking at 220 rpm.

3.7 Plasmid DNA preparation and sequencing

Miniprep plasmid DNA was prepared from 3ml cultures using QIAGEN Plasmid Mini Kit according to the manufacturer's instructions. Plasmid DNA was eluted in 50µl of EB buffer (10mM Tris). The DNA concentration was measured using a Beckman Coulter DU800 Spectrophotometer. Plasmid DNA (500ng) was sent to MRC-Geneservices and subjected to DNA sequencing with the sequencing primers M13F and M13R (Table 4).

All colonies sequenced contained the full length predicted INSP222 cds. 170A2 was chosen and subsequently termed pGEM INSP222. A comparison of the cloned INSP222 sequence and the predicted INSP222 sequence is shown in Figure 8.

Table 4

Primer	Sequence (5'-3')
INSP170fwd	ATG AAG GGC TTC ACT GCT TTT GC (SEQ ID NO:48)
INSP170rev	GCA CAA TCC ATA ATT TTT CAA CCA (SEQ ID NO:49)
M13F	GCT GCT GGC CCT TGG CCT GG (SEQ ID NO:50)
M13R	TAT GTT GAA GAC CGG GGC TTT CTG (SEQ ID NO:51)

Example 4 : Exon structure of predicted and cloned sequences.

Sequence comparison between the predicted INSP170 and cloned INSP222 sequence reveal that INSP222 has an alternative splice site in exon one, resulting in a reading frame which provides an alternative, shorter exon two. (Figures 9 & 10) The alternative splice site of INSP222 results in a protein lacking both the signal peptide and the N-terminal portion of the 1st Kazal domain, provided for by exon two of INSP170. Although the Applicant does not wish to be bound by this theory, it is possible that the cloned INSP222 protein is a shorter, non-secreted form of INSP170.

INSP223 is an amalgamation of the INSP170 and INSP222 sequences and, although the Applicant does not wish to be bound by this theory, is predicted to be a full length, secreted variant of INSP170, containing a shorter exon one, an alternative exon two, but maintaining both the Kazal domains and signal peptide of INSP170.

Example 5: Expression and purification

Further experiments may now be performed to determine the tissue distribution and expression levels of the INSP170, INSP222 and INSP223 polypeptides *in vivo*, on the basis of the nucleotide and amino acid sequence disclosed herein.

The presence of the transcripts for INSP170, INSP222 and INSP223 may be investigated by PCR of cDNA from different human tissues. The INSP170, INSP222 and INSP223 transcripts may be present at very low levels in the samples tested. Therefore, extreme care is needed in the design of experiments to establish the presence of a transcript in various human tissues as a small amount of genomic contamination in the RNA preparation will provide a false positive result. Thus, all RNA should be treated with DNAse prior to use for reverse transcription. In addition, for each tissue a control reaction may be set up in which reverse transcription was not undertaken (a -ve RT control).

For example, 1μg of total RNA from each tissue may be used to generate cDNA using Multiscript reverse transcriptase (ABI) and random hexamer primers. For each tissue, a control reaction is set up in which all the constituents are added except the reverse transcriptase (-ve RT control). PCR reactions are set up for each tissue on the reverse transcribed RNA samples and the minus RT controls. INSP167-specific and INSP170-specific primers may readily be designed on the basis of the sequence information provided herein. The presence of a product of the correct molecular weight in the reverse transcribed sample together with the absence of a product in the minus RT control may be taken as evidence for the presence of a transcript in that tissue. Any suitable cDNA libraries may be used to screen for the INSP170, INSP222 and INSP223 transcripts, not only those generated as described above.

The tissue distribution pattern of the INSP170, INSP222 and INSP223 polypeptides will provide further useful information in relation to the function of those polypeptides.

In addition, further experiments may now be performed using the pEAK12d-INSP167-6HIS and pEAK12d-INSP170-6HIS expression vectors. Transfection of mammalian cell lines with these vectors may enable the high level expression of the INSP170, INSP222 and INSP223 proteins and thus enable the continued investigation of the functional characteristics of the INSP170, INSP222 and INSP223 polypeptides. The following material and methods are an example of those suitable in such experiments:

Cell Culture

Human Embryonic Kidney 293 cells expressing the Epstein-Barr virus Nuclear Antigen (HEK293-EBNA, Invitrogen) are maintained in suspension in Ex-cell VPRO serum-free medium (seed stock, maintenance medium, JRH). Sixteen to 20 hours prior to transfection (Day-1), cells are seeded in 2x T225 flasks (50ml per flask in DMEM / F12 (1:1)

containing 2% FBS seeding medium (JRH) at a density of $2x10^5$ cells/ml). The next day (transfection day 0) transfection takes place using the JetPEITM reagent (2µl/µg of plasmid DNA, PolyPlus-transfection). For each flask, plasmid DNA is co-transfected with GFP (fluorescent reporter gene) DNA. The transfection mix is then added to the 2xT225 flasks and incubated at 37°C (5%CO_2) for 6 days. Confirmation of positive transfection may be carried out by qualitative fluorescence examination at day 1 and day 6 (Axiovert 10 Zeiss).

On day 6 (harvest day), supernatants from the two flasks are pooled and centrifuged (*e.g.* 4°C, 400g) and placed into a pot bearing a unique identifier. One aliquot (500µl) is kept for QC of the 6His-tagged protein (internal bioprocessing QC).

Scale-up batches may be produced by following the protocol called "PEI transfection of suspension cells", referenced BP/PEI/HH/02/04, with PolyEthyleneImine from Polysciences as transfection agent.

Purification process

The culture medium sample containing the recombinant protein with a C-terminal 6His tag is diluted with cold buffer A (50mM NaH_2PO_4; 600mM NaCl; 8.7% (w/v) glycerol, pH 7.5). The sample is filtered then through a sterile filter (Millipore) and kept at 4°C in a sterile square media bottle (Nalgene).

The purification is performed at 4°C on the VISION workstation (Applied Biosystems) connected to an automatic sample loader (Labomatic). The purification procedure is composed of two sequential steps, metal affinity chromatography on a Poros 20 MC (Applied Biosystems) column charged with Ni ions (4.6 x 50 mm, 0.83ml), followed by gel filtration on a Sephadex G-25 medium (Amersham Pharmacia) column (1,0 x 10cm).

For the first chromatography step the metal affinity column is regenerated with 30 column volumes of EDTA solution (100mM EDTA; 1M NaCl; pH 8.0), recharged with Ni ions through washing with 15 column volumes of a 100mM $NiSO_4$ solution, washed with 10 column volumes of buffer A, followed by 7 column volumes of buffer B (50mM NaH_2PO_4; 600mM NaCl; 8.7% (w/v) glycerol, 400mM; imidazole, pH 7.5), and finally equilibrated with 15 column volumes of buffer A containing 15mM imidazole. The sample is transferred, by the Labomatic sample loader, into a 200ml sample loop and subsequently charged onto the Ni metal affinity column at a flow rate of 10ml/min. The column is

washed with 12 column volumes of buffer A, followed by 28 column volumes of buffer A containing 20mM imidazole. During the 20mM imidazole wash loosely attached contaminating proteins are eluted from the column. The recombinant His-tagged protein is finally eluted with 10 column volumes of buffer B at a flow rate of 2ml/min, and the eluted protein is collected.

For the second chromatography step, the Sephadex G-25 gel-filtration column is regenerated with 2ml of buffer D (1.137M NaCl; 2.7mM KCl; 1.5mM KH_2PO_4; 8mM Na_2HPO_4; pH 7.2), and subsequently equilibrated with 4 column volumes of buffer C (137mM NaCl; 2.7mM KCl; 1.5mM KH_2PO_4; 8mM Na_2HPO_4; 20% (w/v) glycerol; pH 7.4). The peak fraction eluted from the Ni-column is automatically loaded onto the Sephadex G-25 column through the integrated sample loader on the VISION and the protein is eluted with buffer C at a flow rate of 2 ml/min. The fraction was filtered through a sterile centrifugation filter (Millipore), frozen and stored at –80°C. An aliquot of the sample is analyzed on SDS-PAGE (4-12% NuPAGE gel; Novex) Western blot with anti-His antibodies. The NuPAGE gel may be stained in a 0.1% Coomassie blue R250 staining solution (30% methanol, 10% acetic acid) at room temperature for 1h and subsequently destained in 20% methanol, 7.5% acetic acid until the background is clear and the protein bands clearly visible.

Following the electrophoresis the proteins are electrotransferred from the gel to a nitrocellulose membrane. The membrane is blocked with 5% milk powder in buffer E (137mM NaCl; 2.7mM KCl; 1.5mM KH_2PO_4; 8mM Na_2HPO_4; 0.1 % Tween 20, pH 7.4) for 1h at room temperature, and subsequently incubated with a mixture of 2 rabbit polyclonal anti-His antibodies (G-18 and H-15, 0.2µg/ml each; Santa Cruz) in 2.5% milk powder in buffer E overnight at 4°C. After a further 1 hour incubation at room temperature, the membrane is washed with buffer E (3 x 10min), and then incubated with a secondary HRP-conjugated anti-rabbit antibody (DAKO, HRP 0399) diluted 1/3000 in buffer E containing 2.5% milk powder for 2 hours at room temperature. After washing with buffer E (3 x 10 minutes), the membrane is developed with the ECL kit (Amersham Pharmacia) for 1min. The membrane is subsequently exposed to a Hyperfilm (Amersham Pharmacia), the film developed and the western blot image visually analysed.

For samples that showed detectable protein bands by Coomassie staining, the protein concentration may be determined using the BCA protein assay kit (Pierce) with bovine serum albumin as standard.

Furthermore, overexpression or knock-down of the expression of the polypeptides in cell lines may be used to determine the effect on transcriptional activation of the host cell genome. Dimerisation partners, co-activators and co-repressors of the INSP170, INSP222 and INSP223 polypeptides may be identified by immunoprecipitation combined with Western blotting and immunoprecipitation combined with mass spectroscopy.

Example 6: 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 on the array

- 5557 of the probes designed to specifically detect collaboration sequences
- 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.

INSP170:

INSP170 is a 5 exon prediction. We have profiled the chips using probes synthesized from 10 normal tissues: bone marrow, brain, lung, ovary, PBMCs, placenta, prostate, spleen and testis.

Examination of the data on an exon by exon basis indicates INSP170 is highest expressed in the prostate sample relative to the other tissues examined and is fairly selectively expressed in the prostate (relative to the other tissues examined) (Figures 9-13).

CLAIMS

1. A polypeptide, which polypeptide

   (i) comprises the amino acid sequence as recited in SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26 and/or SEQ ID NO:28;

   (ii) is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or having an antigenic determinant in common with the polypeptide of (i); or

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

2. A polypeptide according to claim 1 which comprises the amino acid sequence as recited in SEQ ID NO:28.

3. A polypeptide according to claim 1 or 2 which consists of the amino acid sequence as recited in SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26 and/or SEQ ID NO:28.

4. A polypeptide, which polypeptide:

   (i) comprises the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and/or SEQ ID NO:16;

   (ii); is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or having an antigenic determinant in common with the polypeptide of (i); or

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

5. A polypeptide, which comprises the amino acid sequence as recited in SEQ ID NO:12 or SEQ ID NO:16.

6. A polypeptide according to claim 4 or claim 5 which consists of the amino acid sequence as recited in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14 and/or SEQ ID NO:16.

7. A polypeptide, which polypeptide:

   (i) comprises the amino acid sequence as recited in 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 and/or SEQ ID NO:46;

(ii) is a fragment thereof which is a member of the Kazal family of serine protease inhibitors, or having an antigenic determinant in common with the polypeptide of (i); or

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

8. A polypeptide, which comprises the amino acid sequence as recited in SEQ ID NO:42 or SEQ ID NO:42.

9. A polypeptide according to claim 7 or claim 8 which consists of the amino acid sequence as recited in 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 and/or SEQ ID NO:46.

10. A polypeptide which is a functional equivalent according to part (iii) of any of the above claims, 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: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 and/or SEQ ID NO:46 and is a member of the Kazal family of serine protease inhibitors.

11. A polypeptide which is a fragment or a functional equivalent as recited in any one of claims 1 to 11, 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: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 and/or SEQ ID NO:46 or with an active fragment thereof, preferably greater than 85%, 90%, 95%, 98% or 99% sequence identity.

12. A polypeptide which is a functional equivalent as recited in any one of claims 1 to 12, 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: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 and/or SEQ ID NO:46.

13. A polypeptide which is a fragment as recited in claims 1-9 or claim 11 having an antigenic determinant in common with the polypeptide of part (i) of any one of claim 1 to claim 9 which consists of 7 or more amino acid residues from the amino acid sequence 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 and/or SEQ ID NO:46.

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

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

16. A purified nucleic acid molecule according to claim 15, which comprises the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43 and/or SEQ ID NO:45 or is a redundant equivalent or fragment thereof.

17. A purified nucleic acid molecule according to claim 15 which consists of the nucleic acid sequence as recited in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43 and/or SEQ ID NO:45 or is a redundant equivalent or fragment thereof.

18. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claims 15 to 17.

19. A vector comprising a nucleic acid molecule as recited in any one of claims 15 to 18.

20. A host cell transformed with a vector according to claim 19.

21. A ligand which binds specifically to the polypeptide according to any one of claims 1 to 14.

22. A ligand according to claim 21, which is an antibody.

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

24. A compound according to claim 23 that binds to a polypeptide according to any one of claims 1 to 14 without inducing any of the biological effects of the polypeptide.

25. A compound according to claim 24, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.

26. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, or a compound according to any one of claims 23 to 25, for use in therapy or diagnosis of disease.

27. 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 14, or assessing the activity of a polypeptide according to any one of claims 1 to 14, 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.

28. A method according to claim 27 that is carried out *in vitro*.

29. A method according to claim 27 or claim 28, which comprises the steps of:

a) contacting a ligand according to claim 21 or claim 22 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and

b) detecting said complex

30. A method according to claim 27 or claim 28, 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 15 to 18 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.

31. A method according to claim 27 or claim 28, 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 15 to 18 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.

32. A method according to claim 27 or claim 28 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 15 to 18 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

33. The method of claim 32, 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.

34. The method of claim 32 or claim 33, 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.

35. A method according to any one of claims 27 to 34, wherein said disease is selected from the group consisting of cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma; autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation and organ transplant rejection; cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, and ischemia; neurological disorders including central nervous system disease, brain injury, amyotrophic lateral sclerosis, and pain; developmental disorders; metabolic disorders including diabetes mellitus, osteoporosis, and obesity, AIDS and renal disease; infections including viral infection, bacterial infection, fungal infection and parasitic infection.

36. A method according to any one of claims 27 to 34, wherein said disease is selected from the group consisting of chronic inflammatory diseases such as chronic pancreatitis, including idiopathic, hereditary and alcoholic chronic pancreatitis, fibrocalculous pancreatic diabetes, asthma, interstitial lung disease, psoriasis, arthritis gingivitis, peridontitis, cancers such as genitourinary tumours including ovarian cancer and also breast cancer and esophageal cancer, Netherton Syndrome, fertility disorders including male (in)fertility and female (in)fertility, arthritis, Alzheimer's disease and transmissible prion encephalopathies, liver disease including hepatitis, juvenile cirrhosis and hepatocellular carcinoma, cardiovascular disease, hereditary thrombophilia, bleeding, coronary atherosclerosis, increased risk or coronary heart disease, myocardial infarction, hereditary angioedema, acquired angioedema, hereditary angioneurotic edema, capillary leak syndrome, essential hypertension, pregnancy-induced hypertension (preeclampsia), allergenic hypersensitivity, cerebral ischemia/reperfusion, hemorrhagic stroke, Parkinson's disease, chronic obstructive pulmonary disease, congenital alpha 1-antitrypsin deficiency, familial emphysema, lung disease including cystic fibrosis, adult respiratory distress syndrome, chronic

obstructive pulmonary disease, blood clotting disorders, various disorders affecting medium sized arteries including intracranial aneurysms and cervicocephalic arterial dissections, fibromuscular dysplasia, Hepatitis C, familial dementia, and other disorders of coagulation and fibrinolysis.

37. A method according to any one of claims 27 to 34, wherein said disease is a disease in which a member of the Kazal family of serine protease inhibitors is implicated.

38. Use of a polypeptide according to any one of claims 1 to 14 as a Kazal domain-containing protein .

39. Use of a polypeptide according to any one of claims 1 to 14 as a member of the Kazal family of serine protease inhibitors.

40. Use of a polypeptide according to any one of claims 1 to 14 as a serine protease inhibitor.

41. A pharmaceutical composition comprising a polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 20 or claim 21, or a compound according to any one of claims 23 to 25.

42. A vaccine composition comprising a polypeptide according to any one of claims 1 to 14 or a nucleic acid molecule according to any one of claims 15 to 18.

43. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41, for use in the manufacture of a medicament for the treatment of a disease selected from the group consisting of cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposis' sarcoma; autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, and organ transplant rejection; cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, and ischemia; neurological disorders

including central nervous system disease, brain injury, amyotrophic lateral sclerosis, and pain; developmental disorders; metabolic disorders including diabetes mellitus, osteoporosis, and obesity, AIDS and renal disease; infections including viral infection, bacterial infection, fungal infection and parasitic infection.

44. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41, for use in the manufacture of a medicament for the treatment of a disease selected from the group consisting of chronic inflammatory diseases such as chronic pancreatitis, including idiopathic, hereditary and alcoholic chronic pancreatitis, fibrocalculous pancreatic diabetes, asthma, interstitial lung disease, psoriasis, arthritis gingivitis, peridontitis, cancers such as genitourinary tumours including ovarian cancer and also breast cancer and esophageal cancer, Netherton Syndrome, fertility disorders including male (in)fertility and female (in)fertility, arthritis, Alzheimer's disease and transmissible prion encephalopathies, liver disease including hepatitis, juvenile cirrhosis and hepatocellular carcinoma, cardiovascular disease, hereditary thrombophilia, bleeding, coronary atherosclerosis, increased risk or coronary heart disease, myocardial infarction, hereditary angioedema, acquired angioedema, hereditary angioneurotic edema, capillary leak syndrome, essential hypertension, pregnancy-induced hypertension (preeclampsia), allergenic hypersensitivity, cerebral ischemia/reperfusion, hemorrhagic stroke, Parkinson's disease, chronic obstructive pulmonary disease, congenital alpha 1-antitrypsin deficiency, familial emphysema, lung disease including cystic fibrosis, adult respiratory distress syndrome, chronic obstructive pulmonary disease, blood clotting disorders, various disorders affecting medium sized arteries including intracranial aneurysms and cervicocephalic arterial dissections, fibromuscular dysplasia, Hepatitis C, familial dementia, and other disorders of coagulation and fibrinolysis.

45. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41, for use in the manufacture of a medicament for the treatment of a disease in

which a Kazal domain-containing protein is implicated.

46. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41, for use in the manufacture of a medicament for the treatment of a disease in which a member of the Kazal family of serine protease inhibitors is implicated.

47. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41.

48. A method according to claim 47, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

49. A method according to claim 47, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.

50. 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 14, or the level of expression of a nucleic acid molecule according to any one of claims 15 to 18 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.

51. 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 14 or a nucleic acid molecule according to any one of claims 15 to 18 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.

52. 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 15 to 18; 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.

53. The kit of claim 52, further comprising a third container holding an agent for digesting unhybridised RNA.

54. 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 15 to 18.

55. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1 to 14; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

56. 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 14.

57. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 56 with a candidate compound and determining the effect of the compound on the disease of the animal.

58. A method according to any one of claims 48 to 51 or claim 57, wherein said disease is one of the diseases set forth in claim 44 or claim 45.

59. A polypeptide according to any one of claims 1 to 14, a nucleic acid molecule according to any one of claims 15 to 18, a vector according to claim 19, a host cell according to claim 20, a ligand according to claim 21 or claim 22, a compound according to any one of claims 23 to 25, or a pharmaceutical composition according to claim 41 for use in contraception.

60. The use of a INSP170, INSP222 or INSP223 polypeptide as a target for screening candidate drugs for treating or preventing a Kazal family of serine protease inhibitor related disorder.

61. Method of selecting biologically active compounds comprising:

(i) contacting a candidate compound with recombinant host cells expressing an INSP170, INSP222 or INSP223 polypeptide;

(ii) selecting compounds that bind said INSP170, INSP222 or INSP223 polypeptide at the surface of said cells and/or that modulate the activity of the INSP170, INSP222 or INSP223 polypeptide.

FIG. 1

Query= INSP170.pp

(143 letters)

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

1,539,396 sequences; 498,883,957 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
sp|P08480|IPSG_FELCA Double-headed protease inhibitor, submandib...	138	1e-32
sp|P08481|IPSG_PANLE Double-headed protease inhibitor, submandib...	135	1e-31
gb|AAB28294.1| bikazin=Kazal-type proteinase inhibitor [Martes m...	133	6e-31
sp|P16226|IPSG_MELME Double-headed protease inhibitor, submandib...	129	6e-30
sp|P08479|IPSG_VULVU Double-headed protease inhibitor, submandib...	128	2e-29
sp|P01002|IPSG_CANFA Double-headed protease inhibitor, submandib...	127	3e-29
sp|P81481|IPSG_MUSLU Double-headed protease inhibitor, submandib...	126	7e-29
prf||754238A inhibitor DSI-1,proteinase	121	2e-27
sp|P81262|IPSG_UNCUN Double-headed protease inhibitor, submandib...	120	4e-27
sp|P81482|IPSG_MARMT Double-headed protease inhibitor, submandib...	118	2e-26

FIG. 2

>sp|P08480|IPSG_FELCA Double-headed protease inhibitor, submandibular gland
pir||A29654 proteinase inhibitor (PSTI type), submandibular - cat

Length = 111

Score = 138 bits (348), Expect = 1e-32

Identities = 64/108 (59%), Positives = 77/108 (71%), Gaps = 4/108 (3%)

```
Query: 38  VNCSRYI-KGLKTACPRDWRPICGTDQKTYSNECMFCMQNQEVDQEFQLRKLHEGKC-IQ 95
           VNCS+Y  KG    C ++W+PICG D KTYSNECMFC  NQ  ++ FQLRKLH+ KC I+
Sbjct: 6   VNCSQYNRKGSGITCSKEWKPICGIDHKTYSNECMFCQLNQ--NKRFQLRKLHDNKCQIE 63

Query: 96  CTRYSEACTMDYTLHCGSDGNVYSNRCTFCNTVVKSQGAIWLKNYGLC 143
           CT YS  CTM+Y  CGSDG  YSN+C FCN VVK +G ++L  YG C
Sbjct: 64  CTNYSAICTMEYFPLCGSDGQEYSNKCLFCNEVVKRRGTLFLAKYGQC 111
```

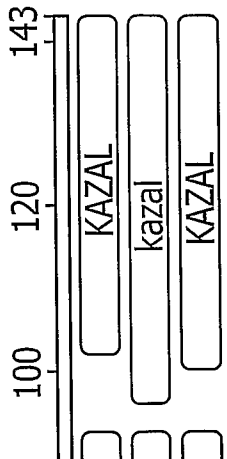
FIG. 3


```
P81482-pine_marten      ----------------APPPV------GDQAGGRKVDCFKYNTTGSAFACTRHERPVC
P81262-leopard          ----------------APPPV------GDQAGGRKVDCFKYNTKGSAFACTRHERPVC
P81481-mink             ----------------APPPV------GDQAGGRKVDCFKYNTTGSEFACSRKWQPVC
P16226-badger           ----------------APPPV------GDQAGGRKVNCSKYNAKGSQFACSRHLDPVC
P08479-fox              ----------------DPPP---------AIGREVDCSSYKGKGSQIACPRHLQPIC
P01002-dog              ----------------GPPP---------AIGREVDCSNYKGKGSQIACPRLHQPIC
P08480-cat              ----------------ASPP-------------EVNCSQYNRKGSGITCSKEWKPIC
P08481-lion             ----------------ASPP-------------EVNCSQYNRKGSGIACSKQLKPIC
prediction-cow          MNSITTFAILALAAMTWAVTPPVSS----GDQEIGIEVNCTKYNIKGIGIACTKIWSPIC
INSP170                 MKGFTAFAILGLAVTAWATSPYGKSGFLSNSHAFMLTVNCSRY-IKGLKTACPRDWRPIC
                                                        *          * #     *

P81482-pine_marten      GTDHRTYSNECMFCMLTQ--NKGFGVRILQDNEC-DIECTQYSDMCTMEYLPLCGSDGKN
P81262-leopard          GTDHRTYSNECMFCMLTQ--NKGFGVRILQDNEC-DIECTQYSDMCTMEYLPLCGSDGKN
P81481-mink             GTDHRTYSNECMFCMLTQ--NKRFPVRILQDNKC-DIECPQYSDMCTMDYLPLCGSDGKN
P16226-badger           GTDHRTYSNECMFCMLTQ--NKRFSVRILQDNNC-DIECTQYSDMCTMDYLPLCGSDGNN
P08479-fox              GTDHNTYSNECMFCALTL--NKEFEVRKLQDTAC-DIECTEYSDMCTMDYRPLCGSDGKN
P01002-dog              GTDHKTYSNECMFCALTL--NKKFEVRKLQDTAC-DIECTEYSDMCTMDYRPLCGSDGKN
P08480-cat              GIDHKTYSNECMFCQLNQ--NKRFQLRKLHDNKC-QIECTNYSAICTMEYFPLCGSDGQE
P08481-lion             GIDHKTYSNECMFCLLNQ--NKQFQIRKLYDDKC-QIECTNYSAICTMEYFPLCGSDGKV
prediction-cow          GIDMKTYSNECMYCFLNR--DKGSQLRKLHNNECREVECTTYSEICTMEYLPHCGSDGIV
INSP170                 GTDQKTYSNECMFCMQNQEVDQEFQLRKLHEGKC--IQCTRYSEACTMDYTLHCGSDGNV
                                  * *                   *    *      * #     *

P81482-pine_marten      YSNKCLFCNAVMGSRGALFLAKHGQCQSP-
P81262-leopard          YSNKCLFCNAVMGSRGALFLAKHGQCQSP-
P81481-mink             YSNKCLFCNAVLRSRGALFLAKHGQCQSP-
P16226-badger           YSNKCLFCNAVLRSRGALFLAKHGQCESP-
P08479-fox              YSNKCIFCNAVVRSRGTIFLAKHGEC----
P01002-dog              YSNKCSFCNAVKKSRGTIFLAKHGEC----
P08480-cat              YSNKCLFCNEVVKRRGTLFLAKYGQC----
P08481-lion             YSNKCSFCNEVVKRRGTLFLAKYGQCK---
prediction-cow          YGNRCVFCNAVVESRGTLYFVKYGPCSESP
INSP170                 YSNRCTFCNTVVKSQGAIWLKNYGLC----
                                          * *                 *
```

FIG. 4



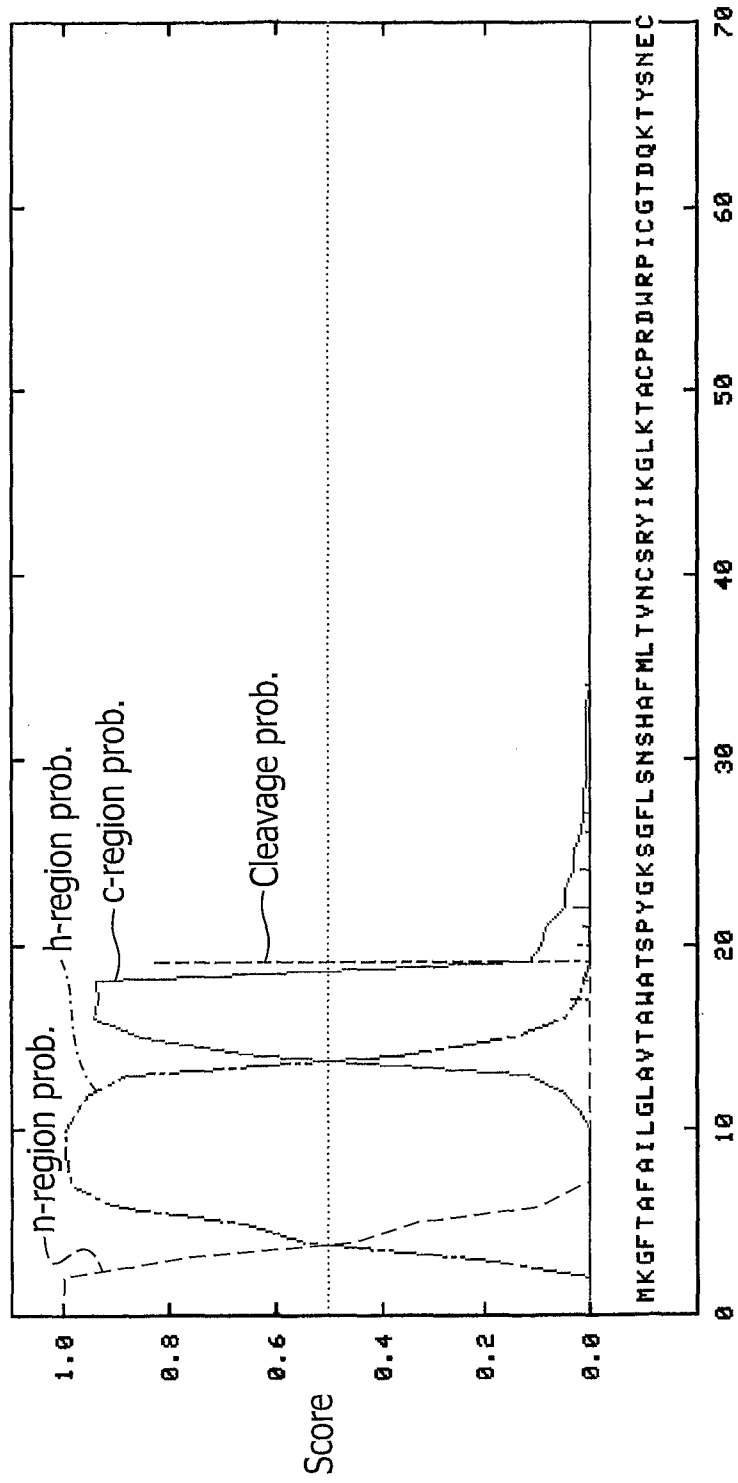
Show Domain Relatives

.. This CD alignment includes 3D structure. To display structure, download Cn3D!

PSSMs producing significant alignments:

	Score (bits)	E value
gnl|CDD|5329 cd00104, KAZAL, Kazal type serine protease inhibitors and foll...	52.1	3e-08
gnl|CDD|5329 cd00104, KAZAL, Kazal type serine protease inhibitors and foll...	40.6	6e-05
gnl|CDD|619 pfam00050, kazal, Kazal-type serine protease inhibitor domain....	49.3	2e-07
gnl|CDD|619 pfam00050, kazal, Kazal-type serine protease inhibitor domain....	45.4	2e-06
gnl|CDD|8961 smart00280, KAZAL, Kazal type serine protease inhibitors; Kaza...	49.1	2e-07
gnl|CDD|8961 smart00280, KAZAL, Kazal type serine protease inhibitors; Kaza...	41.1	5e-05

SignalP-HMM result:



data

```
>INSP170
Prediction: Signal peptide
Signal peptide probability: 0.999
Signal anchor probability: 0.001
Max cleavage site probability: 0.828 between pos. 18 and 19
```

FIG. 5

FIG. 6

INSP170fwd

1 atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg

61 ccat**ATG***gga caaagagatc aagaatttca actcagaaaa cttcatgaag gtaaatgtat*

121 *tcagtgtacc aggtattctg aggcatgtac catggactac acacttcact gtggatctga*

181 *tggaaatgta tattccaaca gatgtacctt ttgcaatacc gttgtgaaga gccaaggtgc*

INSP170rev

241 *aatttggttg aaaaattatg gattgtgc*

INSP222 Start Codon in capitals

INSP222 Coding region in *italics*

Position and sense of PCR primer →

FIG. 7

```
1   atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg

61  ccatatggga caaagagatc aagaatttca actcagaaaa cttcatgaag gtaaatgtat
          M  G  Q  R  D  Q  E  F  Q  L  R  K  L  H  E  G  K  C

121 tcagtgtacc aggtattctg aggcatgtac catggactac acacttcact gtggatctga
    I  Q  C  T  R  Y  S  E  A  C  T  M  D  Y  T  L  H  C  G  S

181 tggaaatgta tattccaaca gatgtacctt ttgcaatacc gttgtgaaga gccaaggtgc
    D  G  N  V  Y  S  N  R  C  T  F  C  N  T  V  V  K  S  Q  G

241 aatttggttg aaaaattatg gattgtgc
                A  I  W  L  K  N  Y  G  L  C
```

FIG. 8

```
Cloned  ATGAAGGGCTTCACTGCTTTTGCCATCCTTGGTCTAGCAGTCACTGCTTGGGCTACCTCG 60
INSP222 ATGAAGGGCTTCACTGCTTTTGCCATCCTTGGTCTAGCAGTCACTGCTTGGGCTACCTCG 60
        ************************************************************

Cloned  CCATATGGGACAAAGAGATCAAGAATTTCAACTCAGAAAACTTCATGAAGGTAAATGTAT 120
INSP222 CCATATGGGACAAAGAGATCAAGAATTTCAACTCAGAAAACTTCATGAAGGTAAATGTAT 120
        ************************************************************

Cloned  TCAGTGTACCAGGTATTCTGAGGCATGTACCATGGACTACACACTTCACTGTGGATCTGA 180
INSP222 TCAGTGTACCAGGTATTCTGAGGCATGTACCATGGACTACACACTTCACTGTGGATCTGA 180
        ************************************************************

Cloned  TGGAAATGTATATTCCAACAGATGTACCTTTTGCAATACCGTTGTGAAGAGCCAAGGTGC 240
INSP222 TGGAAATGTATATTCCAACAGATGTACCTTTTGCAATACCGTTGTGAAGAGCCAAGGTGC 240
        ************************************************************

Cloned  AATTTGGTTGAAAAATTATGGATTGTGC 268
INSP222 AATTTGGTTGAAAAATTATGGATTGTGC 268
        ****************************
```

FIG. 9

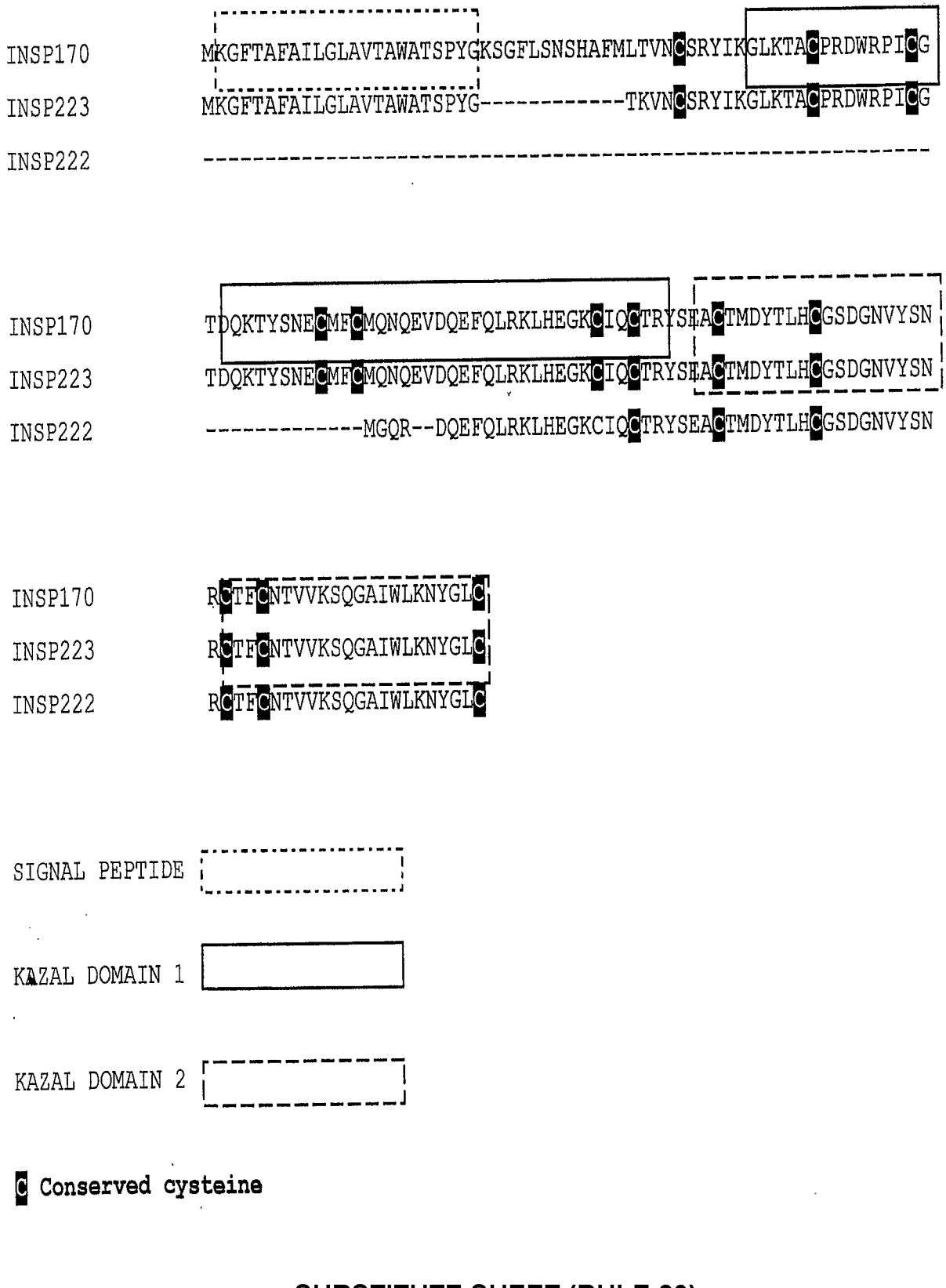
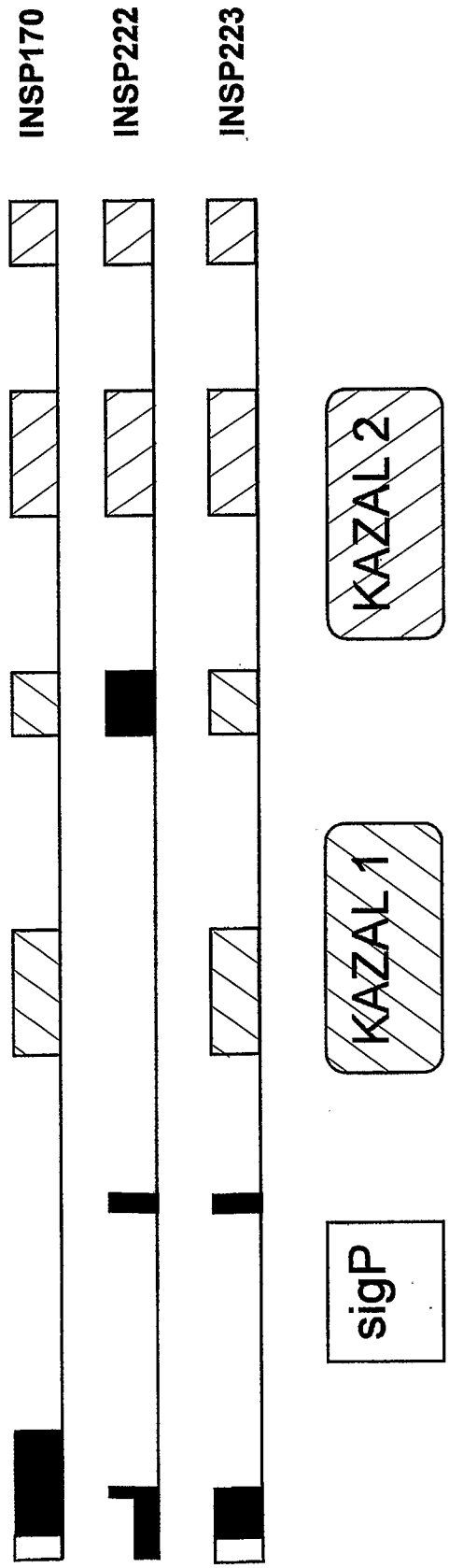
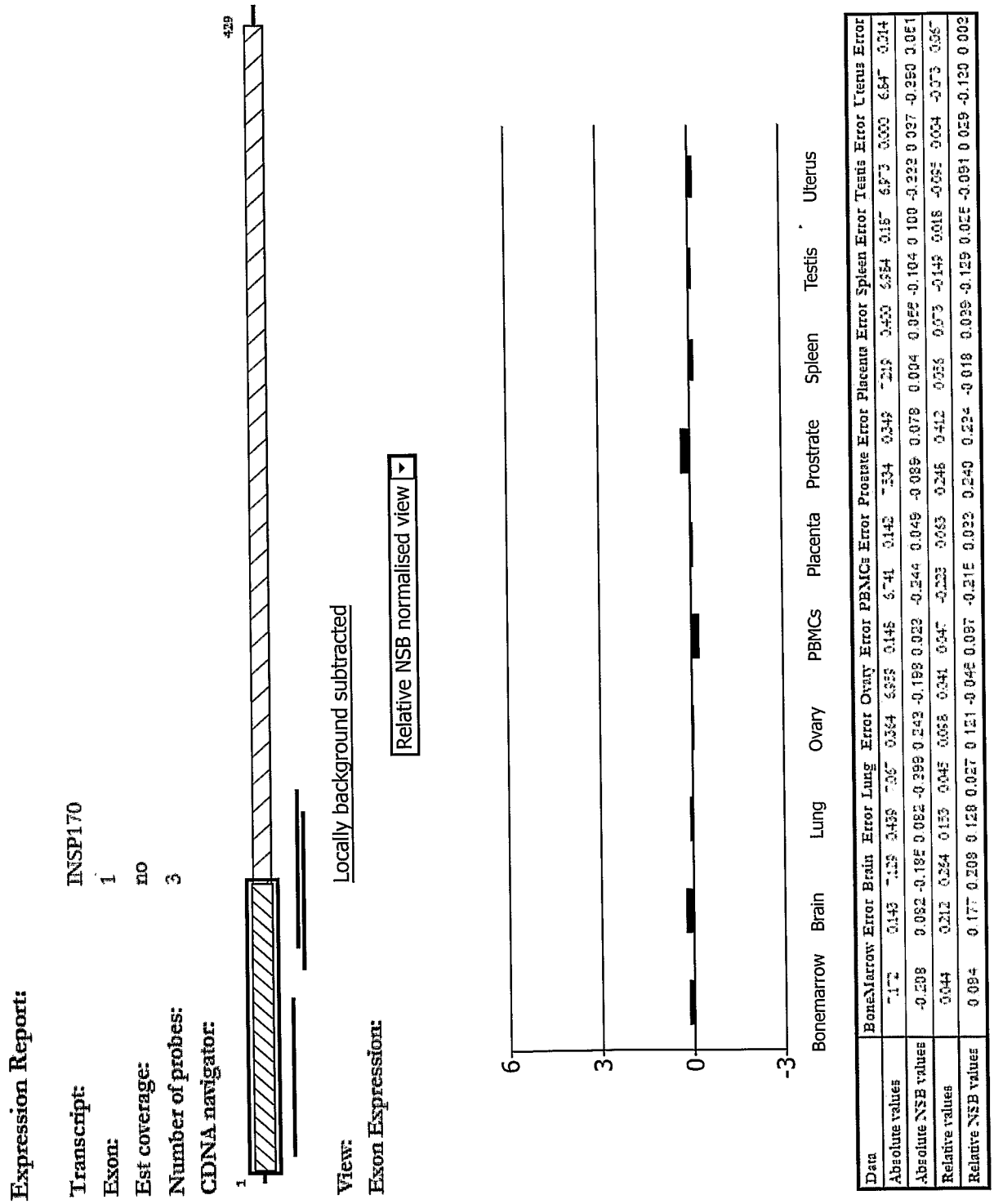

INSP170 MKGFTAFAILGLAVTAWATSPYGKSGFLSNSHAFMLTVNCSRYIKGLKTACPRDWRPICG
INSP223 MKGFTAFAILGLAVTAWATSPYG----------TKVNCSRYIKGLKTACPRDWRPICG
INSP222 ----------
INSP170 TDQKTYSNECMFCMQNQEVDQEFQLRKLHEGKCIQCTRYSEACTMDYTLHCGSDGNVYSN
INSP223 TDQKTYSNECMFCMQNQEVDQEFQLRKLHEGKCIQCTRYSEACTMDYTLHCGSDGNVYSN
INSP222 ----------MGQR--DQEFQLRKLHEGKCIQCTRYSEACTMDYTLHCGSDGNVYSN
INSP170 RCTFCNTVVKSQGAIWLKNYGLC
INSP223 RCTFCNTVVKSQGAIWLKNYGLC
INSP222 RCTFCNTVVKSQGAIWLKNYGLC
SIGNAL PEPTIDE
KAZAL DOMAIN 1
KAZAL DOMAIN 2
C Conserved cysteine

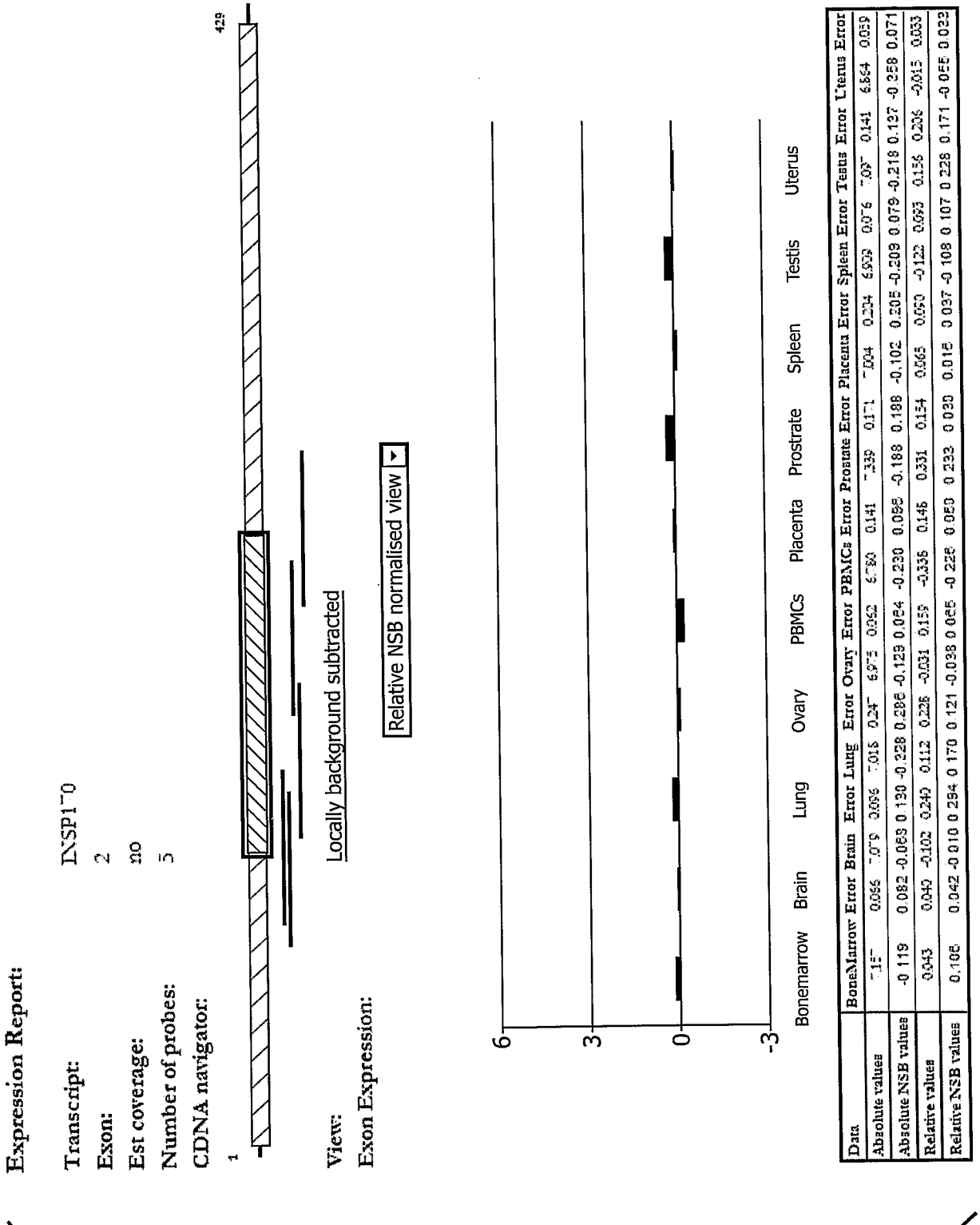
FIG. 10





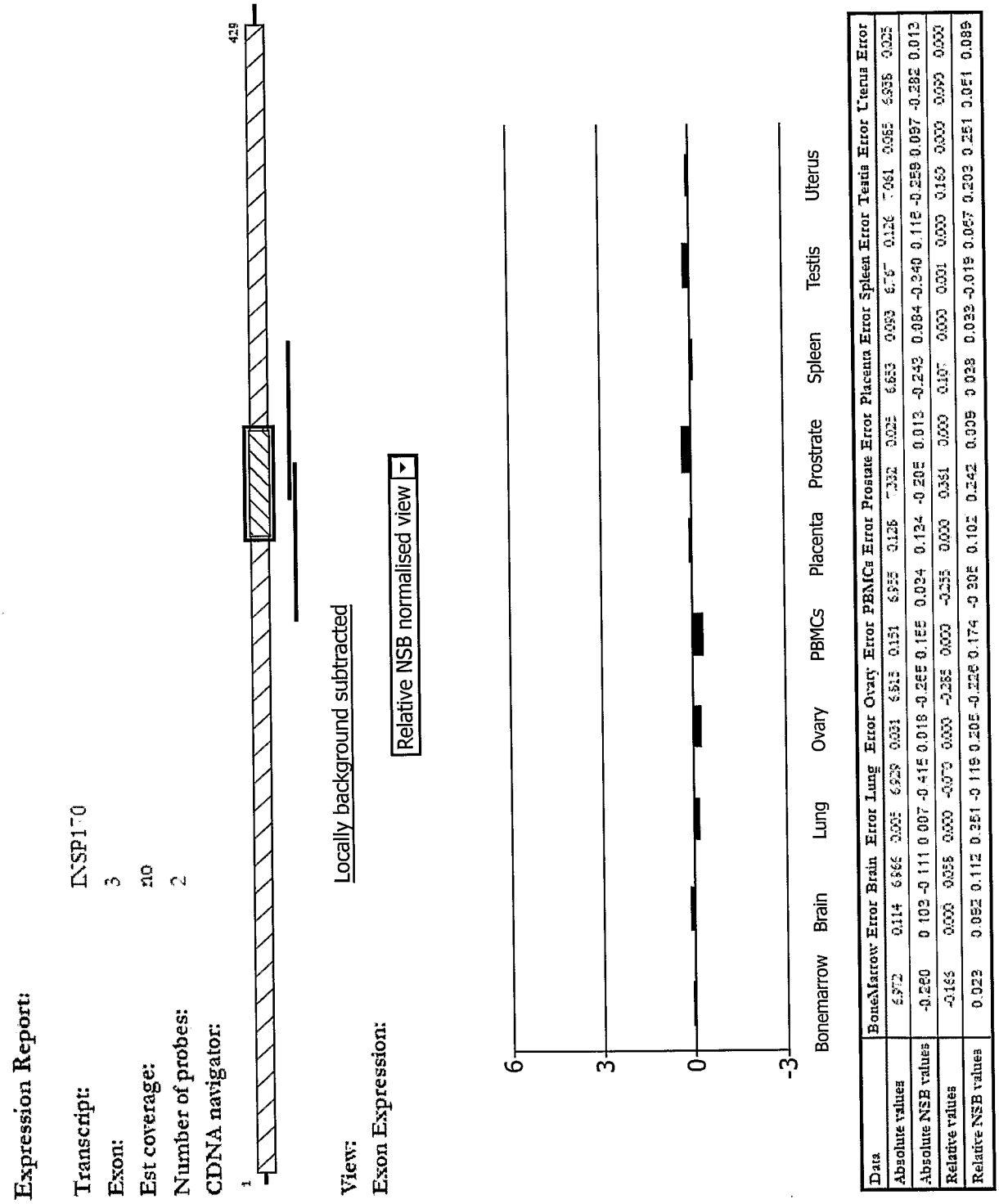
Data	BoneMarrow Error	Brain Error	Lung Error	Ovary Error	PBMCs Error	Prostate Error	Placenta Error	Spleen Error	Testis Error	Uterus Error
Absolute values	[illegible]	0.143	[illegible]	0.384	0.148	[illegible]	0.142	[illegible]	[illegible]	[illegible]
Absolute NSB values	-0.208	0.092	-0.182	0.082	-0.399	0.243	-0.198	0.022	-0.244	0.049	-0.089	0.078	0.004	0.052	-0.104	0.100	-0.222	0.037	-0.250	0.051
Relative values	0.044	0.212	0.264	0.153	0.045	0.068	0.341	0.547	-0.222	0.063	0.245	0.412	0.035	0.073	-0.145	0.018	-0.055	0.024	-0.073	0.067
Relative NSB values	0.094	0.177	0.208	0.128	0.027	0.121	-0.042	0.097	-0.215	0.032	0.240	0.224	-0.018	0.039	-0.129	0.025	-0.091	0.029	-0.120	0.002

FIG. 11



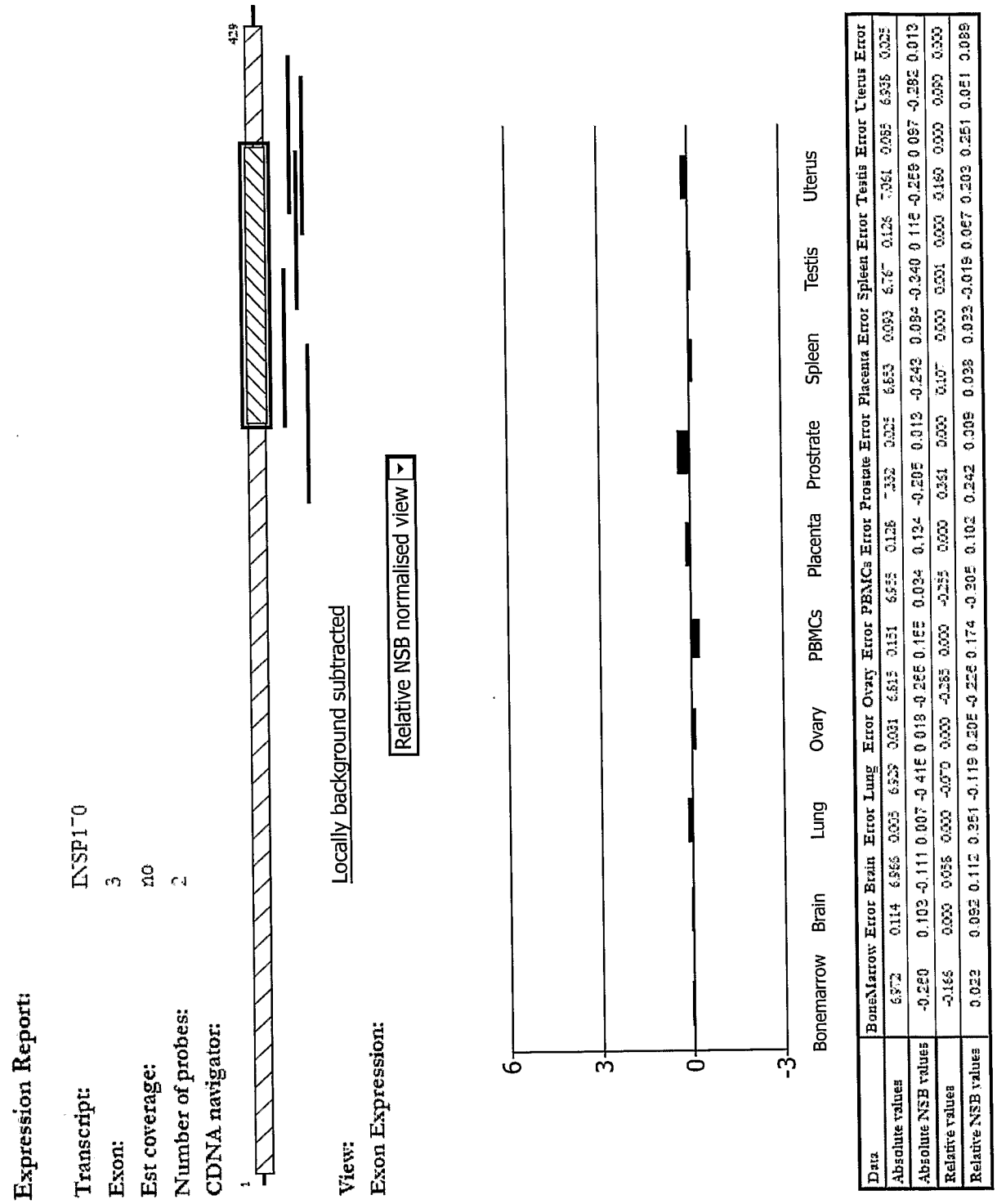
Data	BoneMarrow Error	Brain Error	Lung Error	Ovary Error	PBMCs Error	Placenta Error	Prostate Error	Spleen Error	Testis Error	Uterus Error
Absolute values	[illegible]	0.095	[illegible]	0.247	[illegible]	0.141	[illegible]	0.234	6.939	0.076	[illegible]	0.141	6.864	0.059
Absolute NSB values	-0.119	0.082	-0.063	0.139	-0.228	0.286	-0.129	0.084	-0.230	0.085	-0.188	0.188	-0.102	0.205	-0.209	0.079	-0.218	0.137	-0.358	0.071
Relative values	0.043	0.040	-0.102	0.240	0.112	0.226	-0.031	0.159	-0.335	0.146	0.331	0.154	0.565	0.050	-0.122	0.093	0.136	0.206	-0.015	0.033
Relative NSB values	0.106	0.042	-0.010	0.294	0.170	0.121	-0.038	0.065	-0.228	0.050	0.233	0.030	0.018	0.037	-0.108	0.107	0.228	0.171	-0.055	0.029

FIG. 12



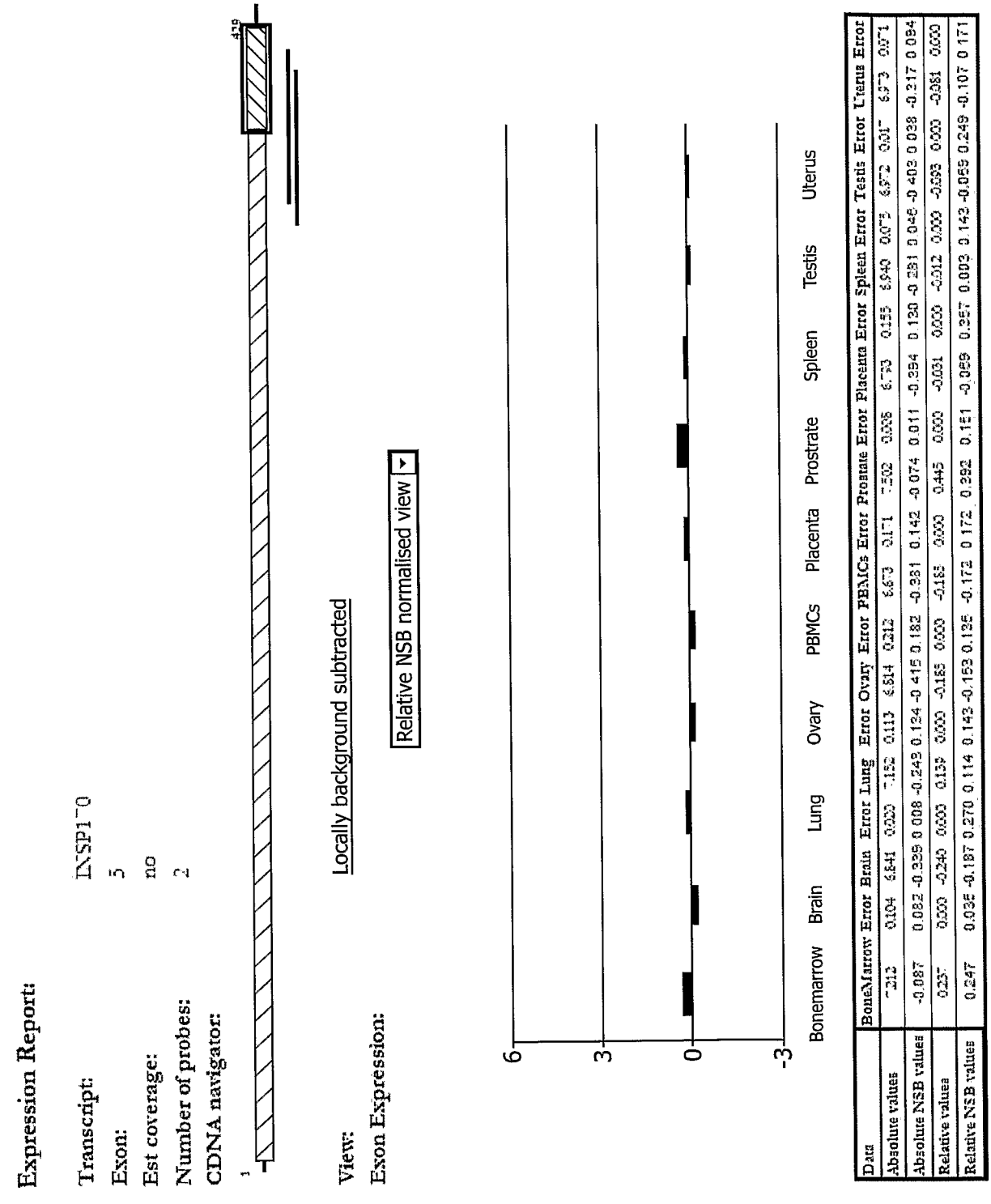
Data	BoneMarrow	Error	Brain	Error	Lung	Error	Ovary	Error	PBMCs	Error	Placenta	Error	Prostate	Error	Spleen	Error	Testis	Error	Uterus	Error
Absolute values	6.972	0.114	6.866	0.005	6.929	0.031	6.813	0.151	6.853	0.126	7.332	0.025	6.853	0.093	6.767	0.126	7.061	0.065	6.935	0.025
Absolute NSB values	-0.260	0.103	-0.111	0.007	-0.415	0.018	-0.265	0.155	0.024	0.124	-0.205	0.013	-0.243	0.084	-0.340	0.112	-0.259	0.097	-0.282	0.013
Relative values	-0.165	0.000	0.056	0.000	-0.070	0.000	-0.285	0.000	-0.255	0.000	0.351	0.000	0.107	0.000	0.001	0.000	0.160	0.000	0.099	0.000
Relative NSB values	0.023	0.052	0.112	0.351	-0.119	0.205	-0.220	0.174	-0.305	0.105	0.242	0.005	0.028	0.039	-0.019	0.067	0.203	0.251	0.051	0.089

FIG. 13



Data	BoneMarrow	Error	Brain	Error	Lung	Error	Ovary	Error	PBMCs	Error	Placenta	Error	Prostate	Error	Spleen	Error	Testis	Error	Uterus	Error
Absolute values	6.972	0.114	6.966	0.005	6.959	0.031	6.815	0.131	6.655	0.128	7.332	0.025	6.653	0.093	6.767	0.126	7.061	0.065	6.936	0.025
Absolute NSB values	-0.280	0.103	-0.111	0.007	-0.416	0.018	-0.285	0.165	0.034	0.124	-0.205	0.018	-0.243	0.084	-0.340	0.112	-0.259	0.097	-0.292	0.012
Relative values	-0.165	0.000	0.056	0.000	-0.070	0.000	-0.285	0.000	-0.255	0.000	0.341	0.000	0.107	0.000	0.001	0.000	0.180	0.000	0.000	0.000
Relative NSB values	0.022	0.092	0.112	0.351	-0.119	0.205	-0.226	0.174	-0.305	0.102	0.242	0.089	0.038	0.033	-0.018	0.067	0.203	0.251	0.051	0.089

FIG. 14



Data	BoneMarrow Error	Brain Error	Lung Error	Ovary Error	PBMCs Error	Placenta Error	Prostate Error	Spleen Error	Testis Error	Uterus Error
Absolute values	7.213	0.104	6.841	0.000	7.150	0.113	6.814	0.212	6.673	0.171	7.502	0.008	6.753	0.155	6.940	0.075	6.972	0.017	6.973	0.071
Absolute NSB values	-0.087	0.082	-0.339	0.008	-0.243	0.124	-0.415	0.182	-0.331	0.142	-0.074	0.011	-0.334	0.130	-0.231	0.046	-0.408	0.038	-0.317	0.084
Relative values	0.257	0.000	-0.240	0.000	0.139	0.000	-0.165	0.000	-0.165	0.000	0.445	0.000	-0.031	0.000	-0.012	0.000	-0.593	0.000	-0.381	0.000
Relative NSB values	0.247	0.035	-0.187	0.270	0.114	0.143	-0.153	0.135	-0.172	0.172	0.292	0.151	-0.069	0.357	0.003	0.143	-0.059	0.249	-0.107	0.171

FIG. 15

SEQUENCE LISTING

<110> ARES TRADING S.A.

<120> NOVEL HUMAN MEMBERS OF THE KAZAL FAMILY OF SERINE PROTEASE INHIBITORS

<130> P042764WO

<150> 0525999.9
<151> 2005-12-21

<160> 51

<170> SeqWin99, version 1.02

<210> 1
<211> 111
<212> DNA
<213> Homo sapiens

<400> 1
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatggca agtcagggtt tctttccaat agtcacgcgt ttatgttaac t 111

<210> 2
<211> 37
<212> PRT
<213> Homo sapiens

<400> 2
Met Lys Gly Phe Thr Ala Phe Ala Ile Leu Gly Leu Ala Val Thr Ala
1 5 10 15
Trp Ala Thr Ser Pro Tyr Gly Lys Ser Gly Phe Leu Ser Asn Ser His
20 25 30
Ala Phe Met Leu Thr
35

<210> 3
<211> 125
<212> DNA
<213> Homo sapiens

<400> 3
gtgaactgtt ccagatacat taaaggctta aagactgcat gcccaagaga ctggagacca 60
atatgtggca cagatcagaa aacttacagt aatgaatgca tgttttgcat gcaaaaccag 120
gaagt 125

<210> 4
<211> 42
<212> PRT
<213> Homo sapiens

<400> 4
Val Asn Cys Ser Arg Tyr Ile Lys Gly Leu Lys Thr Ala Cys Pro Arg
1 5 10 15
Asp Trp Arg Pro Ile Cys Gly Thr Asp Gln Lys Thr Tyr Ser Asn Glu
20 25 30
Cys Met Phe Cys Met Gln Asn Gln Glu Val
35 40

<210> 5
<211> 43
<212> DNA
<213> Homo sapiens

<400> 5
agatcaagaa tttcaactca gaaaacttca tgaaggtaaa tgt 43

<210> 6
<211> 14
<212> PRT
<213> Homo sapiens

<400> 6
Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys
1 5 10

<210> 7
<211> 107
<212> DNA
<213> Homo sapiens

<400> 7
attcagtgta ccaggtattc tgaggcatgt accatggact acacacttca ctgtggatct 60
gatggaaatg tatattccaa cagatgtacc ttttgcaata ccgttgt 107

<210> 8
<211> 36
<212> PRT
<213> Homo sapiens

<400> 8
Ile Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu
1 5 10 15
His Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys
20 25 30
Asn Thr Val Val
35

<210> 9
<211> 43
<212> DNA
<213> Homo sapiens

<400> 9
gaagagccaa ggtgcaattt ggttgaaaaa ttatggattg tgc 43

<210> 10
<211> 14
<212> PRT
<213> Homo sapiens

<400> 10
Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu Cys
1 5 10

<210> 11
<211> 429
<212> DNA
<213> Homo sapiens

<400> 11
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatggca agtcagggtt tctttccaat agtcacgcgt ttatgttaac tgtgaactgt 120
tccagataca ttaaaggctt aaagactgca tgcccaagag actggagacc aatatgtggc 180
acagatcaga aaacttacag taatgaatgc atgttttgca tgcaaaacca ggaagtagat 240
caagaatttc aactcagaaa acttcatgaa ggtaaatgta ttcagtgtac caggtattct 300
gaggcatgta ccatggacta cacacttcac tgtggatctg atggaaatgt atattccaac 360
agatgtacct tttgcaatac cgttgtgaag agccaaggtg caatttggtt gaaaaattat 420
ggattgtgc 429

<210> 12
<211> 143
<212> PRT
<213> Homo sapiens

<400> 12
Met Lys Gly Phe Thr Ala Phe Ala Ile Leu Gly Leu Ala Val Thr Ala
1 5 10 15
Trp Ala Thr Ser Pro Tyr Gly Lys Ser Gly Phe Leu Ser Asn Ser His
20 25 30
Ala Phe Met Leu Thr Val Asn Cys Ser Arg Tyr Ile Lys Gly Leu Lys
35 40 45
Thr Ala Cys Pro Arg Asp Trp Arg Pro Ile Cys Gly Thr Asp Gln Lys
50 55 60
Thr Tyr Ser Asn Glu Cys Met Phe Cys Met Gln Asn Gln Glu Val Asp
65 70 75 80
Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys Ile Gln Cys
85 90 95
Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu His Cys Gly
100 105 110
Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys Asn Thr Val
115 120 125
Val Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu Cys
130 135 140

<210> 13
<211> 57
<212> DNA
<213> Homo sapiens

<400> 13
acctcgccat atggcaagtc agggtttctt tccaatagtc acgcgtttat gttaact 57

<210> 14
<211> 19
<212> PRT
<213> Homo sapiens

<400> 14
Thr Ser Pro Tyr Gly Lys Ser Gly Phe Leu Ser Asn Ser His Ala Phe
1 5 10 15
Met Leu Thr

<210> 15
<211> 375
<212> DNA
<213> Homo sapiens

<400> 15

acctcgccat atggcaagtc agggtttctt tccaatagtc acgcgtttat gttaactgtg 60
aactgttcca gatacattaa aggcttaaag actgcatgcc caagagactg gagaccaata 120
tgtggcacag atcagaaaac ttacagtaat gaatgcatgt tttgcatgca aaaccaggaa 180
gtagatcaag aatttcaact cagaaaactt catgaaggta aatgtattca gtgtaccagg 240
tattctgagg catgtaccat ggactacaca cttcactgtg gatctgatgg aaatgtatat 300
tccaacagat gtacctttg caataccgtt gtgaagagcc aaggtgcaat ttggttgaaa 360
aattatggat tgtgc 375

<210> 16
<211> 125
<212> PRT
<213> Homo sapiens

<400> 16
Thr Ser Pro Tyr Gly Lys Ser Gly Phe Leu Ser Asn Ser His Ala Phe
1 5 10 15
Met Leu Thr Val Asn Cys Ser Arg Tyr Ile Lys Gly Leu Lys Thr Ala
20 25 30
Cys Pro Arg Asp Trp Arg Pro Ile Cys Gly Thr Asp Gln Lys Thr Tyr
35 40 45
Ser Asn Glu Cys Met Phe Cys Met Gln Asn Gln Glu Val Asp Gln Glu
50 55 60
Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys Ile Gln Cys Thr Arg
65 70 75 80
Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu His Cys Gly Ser Asp
85 90 95
Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys Asn Thr Val Val Lys
100 105 110
Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu Cys
115 120 125

<210> 17
<211> 67
<212> DNA
<213> Homo sapiens

<400> 17
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatg 67

<210> 18
<211> 1
<212> PRT
<213> Homo sapiens

<400> 18
Met
1

<210> 19
<211> 8
<212> DNA
<213> Homo sapiens

<400> 19
ggacaaag 8

<210> 20
<211> 3
<212> PRT
<213> Homo sapiens

<400> 20
Gly Gln Arg
1

<210> 21
<211> 43
<212> DNA
<213> Homo sapiens

<400> 21
agatcaagaa tttcaactca gaaaacttca tgaaggtaaa tgt 43

<210> 22
<211> 14
<212> PRT
<213> Homo sapiens

<400> 22
Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys
1 5 10

<210> 23
<211> 107
<212> DNA
<213> Homo sapiens

<400> 23
attcagtgta ccaggtattc tgaggcatgt accatggact acacacttca ctgtggatct 60
gatggaaatg tatattccaa cagatgtacc ttttgcaata ccgttgt 107

<210> 24
<211> 36
<212> PRT
<213> Homo sapiens

<400> 24
Ile Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu
1 5 10 15
His Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys
20 25 30
Asn Thr Val Val
35

<210> 25
<211> 43
<212> DNA
<213> Homo sapiens

<400> 25
gaagagccaa ggtgcaattt ggttgaaaaa ttatggattg tgc 43

<210> 26
<211> 14
<212> PRT
<213> Homo sapiens

<400> 26
Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu Cys
1 5 10

<210> 27
<211> 268
<212> DNA
<213> Homo sapiens

<400> 27
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatggga caaagagatc aagaatttca actcagaaaa cttcatgaag gtaaatgtat 120
tcagtgtacc aggtattctg aggcatgtac catggactac acacttcact gtggatctga 180
tggaaatgta tattccaaca gatgtacctt ttgcaatacc gttgtgaaga gccaaggtgc 240
aatttggttg aaaaattatg gattgtgc 268

<210> 28
<211> 68
<212> PRT
<213> Homo sapiens

<400> 28
Met Gly Gln Arg Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly
1 5 10 15
Lys Cys Ile Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr
20 25 30
Thr Leu His Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr
35 40 45
Phe Cys Asn Thr Val Val Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn
50 55 60
Tyr Gly Leu Cys
65

<210> 29
<211> 67
<212> DNA
<213> Homo sapiens

<400> 29
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatg 67

<210> 30
<211> 23
<212> PRT
<213> Homo sapiens

<400> 30
Met Lys Gly Phe Thr Ala Phe Ala Ile Leu Gly Leu Ala Val Thr Ala
1 5 10 15
Trp Ala Thr Ser Pro Tyr Gly
20

<210> 31
<211> 8
<212> DNA
<213> Homo sapiens

<400> 31
ggacaaag 8

<210> 32
<211> 2
<212> PRT
<213> Homo sapiens

<400> 32
Thr Lys
1

<210> 33
<211> 125
<212> DNA
<213> Homo sapiens

<400> 33
gtgaactgtt ccagatacat taaaggctta aagactgcat gcccaagaga ctggagacca 60
atatgtggca cagatcagaa aacttacagt aatgaatgca tgttttgcat gcaaaaccag 120
gaagt 125

<210> 34
<211> 42
<212> PRT
<213> Homo sapiens

<400> 34
Val Asn Cys Ser Arg Tyr Ile Lys Gly Leu Lys Thr Ala Cys Pro Arg
1 5 10 15
Asp Trp Arg Pro Ile Cys Gly Thr Asp Gln Lys Thr Tyr Ser Asn Glu
20 25 30
Cys Met Phe Cys Met Gln Asn Gln Glu Val
35 40

<210> 35
<211> 43
<212> DNA
<213> Homo sapiens

<400> 35
agatcaagaa tttcaactca gaaaacttca tgaaggtaaa tgt 43

<210> 36
<211> 14
<212> PRT
<213> Homo sapiens

<400> 36
Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys
1 5 10

<210> 37
<211> 107
<212> DNA
<213> Homo sapiens

<400> 37
attcagtgta ccaggtattc tgaggcatgt accatggact acacacttca ctgtggatct 60
gatggaaatg tatattccaa cagatgtacc ttttgcaata ccgttgt 107

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

<400> 38
Ile Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu

```
1               5                   10                  15
His Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys
            20                  25                  30
Asn Thr Val Val
        35
```

<210> 39
<211> 43
<212> DNA
<213> Homo sapiens

<400> 39
gaagagccaa ggtgcaattt ggttgaaaaa ttatggattg tgc 43

<210> 40
<211> 14
<212> PRT
<213> Homo sapiens

<400> 40
Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu Cys
1   5   10

<210> 41
<211> 393
<212> DNA
<213> Homo sapiens

<400> 41
atgaagggct tcactgcttt tgccatcctt ggtctagcag tcactgcttg ggctacctcg 60
ccatatggga caaaggtgaa ctgttccaga tacattaaag gcttaaagac tgcatgccca 120
agagactgga gaccaatatg tggcacagat cagaaaactt acagtaatga atgcatgttt 180
tgcatgcaaa accaggaagt agatcaagaa tttcaactca gaaaacttca tgaaggtaaa 240
tgtattcagt gtaccaggta ttctgaggca tgtaccatgg actacacact tcactgtgga 300
tctgatggaa atgtatattc caacagatgt acctttgca ataccgttgt gaagagccaa 360
ggtgcaattt ggttgaaaaa ttatggattg tgc 393

<210> 42
<211> 131
<212> PRT
<213> Homo sapiens

<400> 42

```
Met Lys Gly Phe Thr Ala Phe Ala Ile Leu Gly Leu Ala Val Thr Ala
1               5                   10                  15
Trp Ala Thr Ser Pro Tyr Gly Thr Lys Val Asn Cys Ser Arg Tyr Ile
            20                  25                  30
Lys Gly Leu Lys Thr Ala Cys Pro Arg Asp Trp Arg Pro Ile Cys Gly
        35                  40                  45
Thr Asp Gln Lys Thr Tyr Ser Asn Glu Cys Met Phe Cys Met Gln Asn
    50                  55                  60
Gln Glu Val Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys
65                  70                  75                  80
Cys Ile Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr
                85                  90                  95
Leu His Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe
            100                 105                 110
Cys Asn Thr Val Val Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr
        115                 120                 125
Gly Leu Cys
    130
```

<210> 43
<211> 13
<212> DNA
<213> Homo sapiens

<400> 43
acctcgccat atg 13

<210> 44
<211> 5
<212> PRT
<213> Homo sapiens

<400> 44
Thr Ser Pro Tyr Gly
1 5

<210> 45
<211> 339
<212> DNA
<213> Homo sapiens

<400> 45
acctcgccat atgggacaaa ggtgaactgt tccagataca ttaaaggctt aaagactgca 60
tgcccaagag actggagacc aatatgtggc acagatcaga aaacttacag taatgaatgc 120
atgttttgca tgcaaaacca ggaagtagat caagaatttc aactcagaaa acttcatgaa 180
ggtaaatgta ttcagtgtac caggtattct gaggcatgta ccatggacta cacacttcac 240
tgtggatctg atggaaatgt atattccaac agatgtacct tttgcaatac cgttgtgaag 300
agccaaggtg caatttggtt gaaaaattat ggattgtgc 339

<210> 46
<211> 113
<212> PRT
<213> Homo sapiens

<400> 46
Thr Ser Pro Tyr Gly Thr Lys Val Asn Cys Ser Arg Tyr Ile Lys Gly
1 5 10 15
Leu Lys Thr Ala Cys Pro Arg Asp Trp Arg Pro Ile Cys Gly Thr Asp
20 25 30
Gln Lys Thr Tyr Ser Asn Glu Cys Met Phe Cys Met Gln Asn Gln Glu
35 40 45
Val Asp Gln Glu Phe Gln Leu Arg Lys Leu His Glu Gly Lys Cys Ile
50 55 60
Gln Cys Thr Arg Tyr Ser Glu Ala Cys Thr Met Asp Tyr Thr Leu His
65 70 75 80
Cys Gly Ser Asp Gly Asn Val Tyr Ser Asn Arg Cys Thr Phe Cys Asn
85 90 95
Thr Val Val Lys Ser Gln Gly Ala Ile Trp Leu Lys Asn Tyr Gly Leu
100 105 110
Cys

<210> 47
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion Protein Linker Sequence

<400> 47
Glu Phe Gly Ala Gly Leu Val Leu Gly Gly Gln Phe Met
1 5 10

<210> 48
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 48
atgaagggct tcactgcttt tgc 23

<210> 49
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 49
gcacaatcca taatttttca acca 24

<210> 50
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 50
gctgctggcc cttggcctgg 20

<210> 51
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 51
tatgttgaag accggggctt tctg 24

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/004846

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/81

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, Sequence Search, WPI Data, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/054152 A (HYSEQ INC [US]; TANG Y TOM [US]; ASUNDI VINOD [US]; GOODRICH RYLE W [U) 3 July 2003 (2003-07-03) abstract page 3, line 32 - page 4, line 26 page 6, line 16 - page 8, line 27 page 42, line 31 - page 44, line 10 page 112, line 12 - page 113, line 4 page 170; table 2A page 260; table 2B page 332; table 3A page 399; table 3B page 458; table 4A page 481; table 4B	1-3, 10-61

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but 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.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 March 2007	11/05/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Mundel, Christophe

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2006/004846

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. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 21, 23-26, 29, 41, 43-47 and 59 (partially)
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. ☐ 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:

see additional sheet

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

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

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ 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.:
1-3 (completely) and 10-61 (partially)

Remark on Protest

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

☐ No protest accompanied the payment of additional search fees.

Continuation of Box II.2

Claims Nos.: 21, 23-26, 29, 41, 43-47 and 59 (partially)

The present claims 21, 23-26, 29, 41, 43-47 and 59 encompass compounds defined only by their desired function, contrary to the requirements of clarity of Article 6 PCT, because the result-to-be-achieved type of definition does not allow the scope of the claim to be ascertained. The fact that any compound could be screened does not overcome this objection, as the skilled person would not have knowledge beforehand as to whether it would fall within the scope claimed, except for compounds like antibodies directed against the polypeptide and inhibiting RNAs. Undue experimentation would be required to screen compounds randomly. This non-compliance with the substantive provisions is to such an extent, that the search was performed taking into consideration the non-compliance in determining the extent of the search for claims 21, 23-26, 29, 41, 43-47 and 59.

The search of claims 21, 23-26, 29, 41, 43-47 and 59 was consequently restricted to antibodies and inhibiting RNAs.

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.

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-3 (completely) and 10-61 (partially)

A polypeptide comprising or consisting of the amino acid sequence as recited in SEQ ID NO: 18, 20, 22, 24, 26 and/or 28 (INSP222), a fusion protein comprising such a polypeptide, a polynucleotide encoding such a polypeptide, a vector comprising such a polynucleotide, a host cell transformed with such a vector, a ligand which specifically binds to the polypeptide, a compound that either increase or decrease the level of expression or activity of said polypeptide, uses of the polypeptide, polynucleotide, vector in diagnostic and/or treatment and a transgenic or knock-out non human animal which expresses higher, lower or absent levels of said polypeptide.

---

2. claims: 4-6 (completely) and 10-61 (partially)

A polypeptide comprising or consisting of the amino acid sequence as recited in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 and/or 16 (INSP170), a fusion protein comprising such a polypeptide, a polynucleotide encoding such a polypeptide, a vector comprising such a polynucleotide, a host cell transformed with such a vector, a ligand which specifically binds to the polypeptide, a compound that either increase or decrease the level of expression or activity of said polypeptide, uses of the polypeptide, polynucleotide, vector in diagnostic and/or treatment and a transgenic or knock-out non human animal which expresses higher, lower or absent levels of said polypeptide.

---

3. claims: 7-9 (completely) and 10-61 (partially)

A polypeptide comprising or consisting of the amino acid sequence as recited in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44 and/or 46 (INSP223), a fusion protein comprising such a polypeptide, a polynucleotide encoding such a polypeptide, a vector comprising such a polynucleotide, a host cell transformed with such a vector, a ligand which specifically binds to the polypeptide, a compound that either increase or decrease the level of expression or activity of said polypeptide, uses of the polypeptide, polynucleotide, vector in diagnostic and/or treatment and a transgenic or knock-out non human animal which expresses higher, lower or absent levels of said polypeptide.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2006/004846

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03054152 A	03-07-2003	AU 2002366951 A1	09-07-2003
		CA 2469941 A1	03-07-2003
		EP 1504099 A2	09-02-2005