



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁷ : C12N 9/50, 1/20, 15/00, C12P 21/06</p>	A1	<p>(11) International Publication Number: WO 00/70033</p> <p>(43) International Publication Date: 23 November 2000 (23.11.00)</p>						
<p>(21) International Application Number: PCT/US00/14037</p> <p>(22) International Filing Date: 19 May 2000 (19.05.00)</p> <p>(30) Priority Data:</p> <table border="0" style="width: 100%;"> <tr> <td style="width: 30%;">60/134,886</td> <td style="width: 40%;">19 May 1999 (19.05.99)</td> <td style="width: 30%;">US</td> </tr> <tr> <td>09/573,274</td> <td>18 May 2000 (18.05.00)</td> <td>US</td> </tr> </table> <p>(71) Applicants: SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHKLINE BEECHAM PLC [GB/GB]; Two New Horizons Court, Great West Road, Brentford, Middlesex TW8 9EP (GB).</p> <p>(72) Inventor: CREASY, Caretha, Lee; 25 E. Bells Mill Road, Erdenheim, PA 19038 (US).</p> <p>(74) Agents: ANDERSEN, Robert, L. et al.; Ratner & Prestia, 301 One Westlakes (Berwyn), P.O. Box 980, Valley Forge, PA 19482-0980 (US).</p>		60/134,886	19 May 1999 (19.05.99)	US	09/573,274	18 May 2000 (18.05.00)	US	<p>(81) Designated States: JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report.</i></p>
60/134,886	19 May 1999 (19.05.99)	US						
09/573,274	18 May 2000 (18.05.00)	US						
<p>(54) Title: MURINE HTRA2</p>								
<p>(57) Abstract</p> <p>Mus musculus htra2 polypeptides and polynucleotides and method for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for screening for compounds which either agonize or antagonize Mus musculus htra2. Such compounds are expected to be useful in treatment of human diseases, including, but not limited to: Alzheimer's Disease; bone loss including osteoporosis; inflammatory diseases such as Adult Respiratory Disease Syndrome (ARDS), Rheumatoid arthritis, Osteoarthritis, Inflammatory Bowel Disease (IBD), psoriasis, dermatitis, asthma, allergies; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; HIV-associated cachexia and other immunodeficiency disorders; septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's disease; cardiovascular disease including restenosis, atherosclerosis, acute heart failure, myocardial infarction; hypotension; hypertension; urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome.</p>								

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Murine HTRA2

This application claims the benefit of U.S. Provisional Application No. 60/134,886, filed May 19, 1999, the entire contents of which are incorporated herein in its entirety.

Field of the Invention

5 This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in identifying compounds that may be agonists and/or antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

10 The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach is rapidly superseding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

15 Functional genomics relies heavily on the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterize further genes and their related polypeptides/proteins, as targets for drug discovery.

20 The *E. coli* htrA gene product (high temperature requirement) encodes a serine protease required for growth at elevated temperatures [Seol et al., Biochem. Biophys. Res. Com. 176, 730-736 (1991)]. Null mutants in *E. coli* htrA show decreased ability to degrade abnormal periplasmic proteins and the protein is required for the degradation of mis-folded proteins formed at high temperatures [Strauch, et al., J. Bacteriol. 171, 2689-2696(1989); Laskoska, et al., Mol. Micro. 22, 555-571 (1996); Kolmar et al., J. Bacteriol. 178, 5925-5929 (1996)]. Recently, the first human homolog to *E. coli* htrA was identified and found to be up-regulated in osteoarthritic tissue; therefore, it has been suggested that mammalian homologs to *E. coli* htrA may function in the stress response which occurs in many disease situations [Hu et al., J. Biol. Chem. 273, 34406-34412 (1998)]. Based on this belief, it can be suggested that other mammalian homologs to *E. coli* htrA will play a similar role in stress response.

Summary of the Invention

30 The present invention relates to *Mus musculus* htra2, in particular *Mus musculus* htra2 polypeptides and *Mus musculus* htra2 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors of the *Mus musculus* htra2 gene. This invention further relates to the

generation of *in vitro* and *in vivo* comparison data relating to the polynucleotides and polypeptides in order to predict oral absorption and pharmacokinetics in man of compounds that either agonize or antagonize the biological activity of such polynucleotides or polypeptides. Such a comparison of data will enable the selection of drugs with optimal pharmacokinetics in man, *i.e.*, good oral
5 bioavailability, blood-brain barrier penetration, plasma half life, and minimum drug interaction.

The present invention further relates to methods for creating transgenic animals, which overexpress or underexpress or have regulatable expression of a *htra2* gene and "knock-out" animals, preferably mice, in which an animal no longer expresses a *htra2* gene. Furthermore, this invention relates to transgenic and knock-out animals obtained by using these methods. Such animal models are
10 expected to provide valuable insight into the potential pharmacological and toxicological effects in humans of compounds that are discovered by the aforementioned screening methods as well as other methods. An understanding of how a *Mus musculus htra2* gene functions in these animal models is expected to provide an insight into treating and preventing human diseases including, but not limited to:
15 Alzheimer's Disease; bone loss including osteoporosis; inflammatory diseases such as Adult Respiratory Disease Syndrome (ARDS), Rheumatoid arthritis, Osteoarthritis, Inflammatory Bowel Disease (IBD), psoriasis, dermatitis, asthma, allergies; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; HIV-associated cachexia and other immunodeficiency disorders; septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's disease; cardiovascular disease including restenosis, atherosclerosis, acute heart failure, myocardial
20 infarction; hypotension; hypertension; urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, hereinafter referred to as "the Diseases", amongst others.

Description of the Invention

25 In a first aspect, the present invention relates to *Mus musculus htra2* polypeptides. Such polypeptides include isolated polypeptides comprising an amino acid sequence having at least a 95% identity, most preferably at least a 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

(a) an isolated polypeptide encoded by a polynucleotide comprising the sequence contained in
30 SEQ ID NO:1;

(b) an isolated polypeptide comprising a polypeptide sequence having at least a 95%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(c) an isolated polypeptide comprising the polypeptide sequence of SEQ ID NO:2;(d) an isolated polypeptide having at least a 95%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(e) the polypeptide sequence of SEQ ID NO:2; and

5 (f) variants and fragments thereof; and portions of such polypeptides in (a) to (e) that generally contain at least 30 amino acids, more preferably at least 50 amino acids, thereof.

Polypeptides of the present invention are believed to be members of the serine protease family of polypeptides. They are, therefore, of interest, because agonists and antagonists of serine proteases have proven to be useful therapeutics. Furthermore, the polypeptides of the present invention can be used to establish assays to predict oral absorption and pharmacokinetics in man and thus enhance compound and formulation design, among others. These properties, either alone or in the aggregate, are hereinafter referred to as "Mus musculus htra2 activity" or "Mus musculus htra2 polypeptide activity" or "biological activity of htra2." Preferably, a polypeptide of the present invention exhibits at least one biological activity of Mus musculus htra2.

15 Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including alleles and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative. Particularly preferred variants are those in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination. Particularly preferred primers will have between 20 and 25 nucleotides.

25 Preferred fragments of polypeptides of the present invention include an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO:2, or an isolated polypeptide comprising an amino acid sequence having at least 15, 20, 30, 40, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO:2.

30 Also preferred are biologically active fragments which are those fragments that mediate activities of htra2, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of Mus musculus or the ability to initiate, or maintain cause the Diseases in an individual, particularly a human.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

5 The polypeptides of the present invention may be in the form of a "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-sequences, sequences that aid in purification, for instance, multiple histidine residues, or an additional sequence for stability during recombinant production.

10 The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

15 Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

20 In a further aspect, the present invention relates to *Mus musculus* htra2 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide having at least a 95% identity, to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least a 97% identity are highly preferred, while those with at least a 98-99% identity are more highly preferred, and those with at least a 99% identity are most highly preferred. Such polynucleotides include a polynucleotide
25 comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

30 Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence having at least a 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least a 97% identity are highly preferred, while those with at least a 98-99% identity are more highly preferred, and those with at least a 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence having at least a 95% identity, to SEQ ID NO:1 over the entire length of SEQ

ID NO:1. In this regard, polynucleotides which have at least a 97% identity are highly preferred, while those with at least a 98-99% identify are more highly preferred, and those with at least a 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1, as well as the polynucleotide of SEQ ID NO:1.

5 The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

 The nucleotide sequence of SEQ ID NO:1 shows homology with PSP-1 (GPS accession number, V2953; patent number EP-828003-A2). The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotides 1 to 1374) encoding a
10 polypeptide of 458 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of
15 SEQ ID NO:2 is structurally related to other proteins of the serine protease family, having homology and/or structural similarity with PSP-1 (GPS accession number, W56769; patent number EP-828003-A3).

 Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one htra2
20 activity.

 Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of *Mus musculus* adult brain, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al. Science* (1991) 252:1651-1656; Adams, M.D. *et al., Nature* (1992) 355:632-634; Adams, M.D., *et al., Nature* (1995) 377 Supp.: 3-174).
25 Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

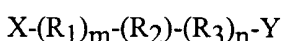
 When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with
30 other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz, *et al., Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The

polynucleotide may also comprise non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance
5 from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 1 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination. Particularly preferred probes will have between 30 and 50 nucleotides, but may have between 100 and 200 contiguous nucleotides of the polynucleotide of SEQ ID NO:1.

10 A preferred embodiment of the invention is a polynucleotide of consisting of or comprising nucleotide 1 to the nucleotide immediately upstream of or including nucleotide 1377 set forth in SEQ ID NO:1, which encodes a htra2 polypeptide.

The invention also includes a polynucleotide consisting of or comprising a polynucleotide of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, a metal or a modified nucleotide residue, or together with Y defines a covalent bond, and at the 3' end of the molecule, Y is hydrogen, a metal, or a modified nucleotide residue, or together with X defines the covalent bond, each occurrence of R₁ and R₃ is independently any nucleic acid residue or modified nucleic acid residue, m is an integer between
20 1 and 3000 or zero, n is an integer between 1 and 3000 or zero, and R₂ is a nucleic acid sequence or modified nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from SEQ ID NO: 1 or a modified nucleic acid sequence thereof. In the polynucleotide formula above, R₂ is oriented so that its 5' end nucleic acid residue is at the left, bound to R₁, and its 3' end nucleic acid residue is at the right, bound to R₃. Any stretch of nucleic acid residues denoted by either R₁ and/or
25 R₂, where m and/or n is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. Where, in a preferred embodiment, X and Y together define a covalent bond, the polynucleotide of the above formula is a closed, circular polynucleotide, which can be a double-stranded polynucleotide wherein the formula shows a first strand to which the second strand is complementary. In another preferred embodiment m and/or n is an integer between 1 and 1000. Other
30 preferred embodiments of the invention are provided where m is an integer between 1 and 50, 100 or 500, and n is an integer between 1 and 50, 100, or 500.

Polynucleotides that are identical, or are substantially identical to a nucleotide sequence of SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including
5 genes encoding homologs and orthologs from species other than *Mus musculus*) that have a high sequence identity to SEQ ID NO:1. Typically these nucleotide sequences are 95% identical to that of the referent. Preferred probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides, and may even have at least 100 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

10 A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from a species other than *Mus musculus*, may be obtained by a process comprising the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides in length; and isolating full-length cDNA and genomic clones comprising said polynucleotide sequence. Such
15 hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus, the present invention also includes isolated
20 polynucleotides, preferably of at least 100 nucleotides in length, obtained by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This
25 is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerization reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example, those based on the method of Rapid
30 Amplification of cDNA ends (RACE) (see, for example, Frohman, *et al.*, *Proc. Natl. Acad. Sci., USA* 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification

(PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that
5 anneals further 5' in the known gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known
10 in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems comprising a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from
15 the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides 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
20 Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

25 Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and
30 virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The

expression systems may comprise control regions that regulate as well as engender expression.

Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a 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,
5 those set forth in Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL (*supra*).

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced
10 intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography.
15 Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and/or purification.

Mus musculus htra2 gene products can be expressed in transgenic animals. Animals of any species, including, but not limited to: mice, rats, rabbits, guinea pigs, dogs, cats, pigs, micro-pigs, goats,
20 and non-human primates, *e.g.*, baboons, monkeys, chimpanzees, may be used to generate htra2 transgenic animals.

This invention further relates to a method of producing transgenic animals, preferably Mus musculus, over-expressing htra2, which method may comprise the introduction of several copies of a segment comprising at least the polynucleotide sequence encoding SEQ ID NO:2 with a suitable
25 promoter into the cells of a Mus musculus embryo, or the cells of another species, at an early stage.

This invention further relates to a method of producing transgenic animals, preferably Mus musculus, under-expressing or regulatably expressing htra2, which method may comprise the introduction of a weak promoter or a regulatable promoter (*e.g.*, an inducible or repressible promoter) respectively, expressibly linked to the polynucleotide sequence of SEQ ID NO:1 into the cells of a Mus
30 musculus embryo at an early stage.

This invention also relates to transgenic animals, characterized in that they are obtained by a method, as defined above.

Any technique known in the art may be used to introduce a *Mus musculus* *htra2* transgene into animals to produce a founder line of animals. Such techniques include, but are not limited to: pronuclear microinjection (U.S. Patent No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten, *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985); gene targeting in embryonic stem cells (Thompson, *et al.*, *Cell* 56: 313-321 (1989); electroporation of embryos (Lo, *Mol. Cell Biol.* 3: 1803-1814 (1983); and sperm-mediated gene transfer (Lavitrano, *et al.*, *Cell* 57: 717-723 (1989); *etc.* For a review of such techniques, see Gordon, *Intl. Rev. Cytol.* 115: 171-229 (1989).

A further aspect of the present invention involves gene targeting by homologous recombination in embryonic stem cells to produce a transgenic animal with a mutation in a *htra2* gene ("knock-out" mutation). In such so-called "knock-out" animals, there is inactivation of the *htra2* gene or altered gene expression, such that the animals are useful to study the function of the *htra2* gene, thus providing animals models of human disease, which are otherwise not readily available through spontaneous, chemical or irradiation mutagenesis. Another aspect of the present invention involves the generation of so-called "knock-in" animals in which a portion of a wild-type gene is fused to the cDNA of a heterologous gene.

This invention further relates to a method of producing "knock-out" animals, preferably mice, no longer expressing *htra2*. By using standard cloning techniques, a *Mus musculus* *htra2* cDNA (SEQ ID NO:1) can be used as a probe to screen suitable libraries to obtain the murine *htra2* genomic DNA clone. Using the murine genomic clone, the method used to create a knockout mouse is characterized in that:

a suitable mutation is produced in the polynucleotide sequence of the murine *htra2* genomic clone, which inhibits the expression of a gene encoding murine *htra2*, or inhibits the activity of the gene product;

said modified murine *htra2* polynucleotide is introduced into a homologous segment of murine genomic DNA, combined with an appropriate marker, so as to obtain a labeled sequence comprising said modified murine genomic DNA;

said modified murine genomic DNA comprising the modified polynucleotide is transfected into embryonic stem cells and correctly targeted events selected *in vitro*; then

said stem cells are re injected into a mouse embryo; then

said embryo is implanted into a female recipient and brought to term as a chimera which transmits said mutation through the germline; and

homozygous recombinant mice are obtained at the F2 generation which are recognizable by the presence of the marker.

Various methods for producing mutations in non-human animals are contemplated and well known in the art. In a preferred method, a mutation is generated in a murine *htra2* allele by the introduction of a DNA construct comprising DNA of a gene encoding murine *htra2*, which murine gene contains the mutation. The mutation is targeted to the allele by way of the DNA construct. The DNA
5 of the gene encoding murine *htra2* comprised in the construct may be foreign to the species of which the recipient is a member, may be native to the species and foreign only to the individual recipient, may be a construct comprised of synthetic or natural genetic components, or a mixture of these. The mutation may constitute an insertion, deletion, substitution, or combination thereof. The DNA construct can be introduced into cells by, for example, calcium-phosphate DNA co-precipitation. It is preferred that a
10 mutation be introduced into cells using electroporation, microinjection, virus infection, ligand-DNA conjugation, virus-ligand-DNA conjugation, or liposomes.

Another embodiment of the instant invention relates to "knock-out" animals, preferably mice, obtained by a method of producing recombinant mice as defined above, among others.

Another aspect of this invention provides for *in vitro* *htra2* "knock-outs", *i.e.*, tissue cultures.
15 Animals of any species, including, but not limited to: mice, rats, rabbits, guinea pigs, dogs, cats, pigs, micro-pigs, goats, and non-human primates, *e.g.*, baboons, monkeys, chimpanzees, may be used to generate *in vitro* *htra2* "knock-outs". Methods for "knocking out" genes *in vitro* are described in Galli-Taliadoros, *et al.*, *Journal of Immunological Methods* 181: 1-15 (1995).

Transgenic, "knock-in", and "knock-out" animals, as defined above, are a particularly
20 advantageous model, from a physiological point of view, for studying serine protease. Such animals will be valuable tools to study the functions of a *htra2* gene. Moreover, such animal models are expected to provide information about potential toxicological effects in humans of any compounds discovered by an aforementioned screening method, among others. An understanding of how a *Mus musculus* *htra2* gene functions in these animal models is expected to provide an insight into treating and
25 preventing human diseases including, but not limited to: Alzheimer's Disease; bone loss including osteoporosis; inflammatory diseases such as Adult Respiratory Disease Syndrome (ARDS), Rheumatoid arthritis, Osteoarthritis, Inflammatory Bowel Disease (IBD), psoriasis, dermatitis, asthma, allergies; infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; HIV-associated cachexia and other immunodeficiency disorders; septic
30 shock; pain; injury; cancers; anorexia; bulimia; Parkinson's disease; cardiovascular disease including restenosis, atherosclerosis, acute heart failure, myocardial infarction; hypotension; hypertension; urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome.

Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases mentioned herein. It is, therefore, an aspect of the invention to devise screening methods to identify compounds that stimulate (agonists) or that inhibit (antagonists) the function of the polypeptide, such as agonists, antagonists and inhibitors. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those that stimulate or inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for the Diseases mentioned herein mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists and antagonists so-identified may be natural or modified substrates, ligands, receptors, enzymes, *etc.*, as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan, *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY 1(2): Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, a screening method may involve measuring or, qualitatively or quantitatively, detecting the competition of binding of a candidate compound to the polypeptide with a labeled competitor (*e.g.*, agonist or antagonist). Further, screening methods may test whether the candidate compound results in a signal generated by an agonist or antagonist of the polypeptide, using detection systems appropriate to cells bearing the polypeptide. Antagonists are generally assayed in the presence of a known agonist and an effect on activation by the agonist by the presence of the candidate compound is observed. Further, screening methods may simply comprise the steps of mixing a candidate compound with a solution comprising a polypeptide of the present invention, to form a mixture, measuring *Mus musculus htra2* activity in the mixture, and comparing a *Mus musculus htra2* activity of the mixture to a control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventional low capacity screening methods and also in high-throughput screening (HTS) formats. Such HTS formats include not only the well-established use of 96- and, more recently, 384-well microtiter plates but also emerging methods such as the nanowell method described by Schullek, *et al.*, *Anal Biochem.*, 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and *Mus musculus htra2* polypeptide, as herein described, can also be used for high-throughput screening assays to identify antagonists of antagonists of the polypeptide of the present invention (see D. Bennett, *et al.*, *J. Mol. Recognition*, 8:52-58 (1995); and K. Johanson, *et al.*, *J. Biol. Chem.*, 270(16):9459-9471 (1995)).

Another aspect of the invention is a method for assaying for the presence of a substance that modulates htra2 activity by direct binding to htra2 protein. Examples of modulators include, but are not limited to, peptides and small organic molecules including peptidomimetics. Modulator candidates are synthesized on a solid support by techniques such as those disclosed in Lam *et al.*,
5 *Nature* 354, 82 (1991) or Burbaum *et al.*, *Proc. Natl. Acad. Sci. USA* 92, 6027 (1995) to provide solid support-associated modulator candidates. A labeled htra2 protein is provided having the amino acid sequence of htra2 (SEQ ID NO: 2) or a functional derivative thereof. Exemplary labels include directly attached fluorescent or colored dyes, biotin, radioisotopes or epitope tags, which are detectable by a suitable antibody. A mixture of solid support-associated modulator candidates and
10 labeled htra2 protein is incubated under conditions which can permit the formation of a htra2 protein/modulator candidate complex. The solid support is separated from free soluble labeled htra2 protein. An assay is performed for the presence of solid support-associated labeled protein. Solid supports complexed with labeled protein are isolated and the identity of the modulator candidate determined by techniques well known to those skilled in the art, such as the TOF-SIMS method in
15 Brummel *et al.*, *Science* 264, 399-402(1994).

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligopeptides or proteins that are closely related to ligands, substrates, receptors, enzymes, *etc.*, as the case may be, of a htra2 polypeptide, *e.g.*, a fragment of a ligand, substrate, receptor, enzyme, *etc.*; or small molecules which bind to a htra2 polypeptide but do not elicit a response, so that an activity of a
20 htra2 polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, inhibitors, ligands, receptors, substrates, enzymes, *etc.* for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which compounds comprise a member selected from the group consisting of:

- 25
- (a) a polypeptide of the present invention;
 - (b) a recombinant cell expressing a polypeptide of the present invention; or
 - (c) a cell membrane expressing a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

30 It will be appreciated that in any such kit, (a), (b) or (c) may comprise a substantial component.

It will also be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

- 14 -

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

5 (c) synthesizing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an iterative process.

10 In an alternative preferred embodiment, the present invention relates to the use of *Mus musculus* htra2 polypeptides, polynucleotides, and recombinant materials thereof in selection screens to identify compounds which are neither agonists nor antagonist/inhibitors of *Mus musculus* htra2. The data from such a selection screen is expected to provide *in vitro* and *in vivo* comparisons and to predict oral absorption, pharmacokinetics in humans. The ability to make such a comparison of data will enhance formulation design through the identification of compounds with optimal development characteristics, *i.e.*, high oral bioavailability, UID (once a day) dosing, reduced drug interactions, reduced variability, and reduced food effects, among others.

15

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Allele" refers to one or more alternative forms of a gene occurring at a given locus in the genome.

20

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1.

25

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, *e.g.*, EP-A 0232 262]. On the other hand, for some uses, it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected, and purified.

30

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such

relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms, "ortholog", and "paralog". "Ortholog" refers to polynucleotides/genes or polypeptide that are homologs via speciation, that is closely related and assumed to have common descent based on structural and functional considerations. "Paralog" refers to polynucleotides/genes or polypeptide that are homologs via gene duplication, for instance, duplicated variants within a genome.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectively, over the length of the sequences being compared. For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an associated 'score' from which the "% similarity" of the two sequences can then be determined.

Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J., *et al*, *Nucleic Acids Res*, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (*J. Mol. Biol.*, 147:195-197, 1981, *Advances in Applied Mathematics*, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison,

GAP aligns two sequences, finding a "maximum similarity", according to the algorithm of Needleman and Wunsch (*J. Mol. Biol.*, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S.F., *et al.*, *J. Mol. Biol.*, 215, 403-410, 1990, Altschul S.F., *et al.*, *Nucleic Acids Res.*, 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov) and FASTA (Pearson W R, *Methods in Enzymology*, 183: 63-99 (1990); Pearson W R and Lipman D.J., *Proc Nat Acad Sci USA*, 85: 2444-2448 (1988) (available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S. and Henikoff J.G., *Proc. Nat. Acad Sci. USA*, 89: 10915-10919 (1992)) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a polynucleotide or a polypeptide sequence of the present invention, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

Alternatively, for instance, for the purposes of interpreting the scope of a claim including mention of a "% identity" to a reference polynucleotide, a polynucleotide sequence having, for example, at least 95% identity to a reference polynucleotide sequence is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference sequence. Such point mutations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These point mutations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polynucleotide sequence having at least 95% identity to a reference polynucleotide sequence, up to 5% of the nucleotides of the in the reference sequence may be

deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other % identities such as 96%, 97%, 98%, 99% and 100%.

For the purposes of interpreting the scope of a claim including mention of a "% identity" to a reference polypeptide, a polypeptide sequence having, for example, at least 95% identity to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include up to five point mutations per each 100 amino acids of the reference sequence. Such point mutations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These point mutations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between these terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a sequence polypeptide sequence having at least 95% identity to a reference polypeptide sequence, up to 5% of the amino acids of the in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other % identities such as 96%, 97%, 98%, 99%, and 100%.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO:1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1, y is 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to the nearest integer prior

to subtracting it from x_n . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO:2 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2, y is 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Knock-in" refers to the fusion of a portion of a wild-type gene to the cDNA of a heterologous gene

"Knock-out" refers to partial or complete suppression of the expression of a protein encoded by an endogenous DNA sequence in a cell. The "knock-out" can be affected by targeted deletion of the whole or part of a gene encoding a protein, in an embryonic stem cell. As a result, the deletion

may prevent or reduce the expression of the protein in any cell in the whole animal in which it is normally expressed.

"Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

"Transgenic animal" refers to an animal to which exogenous DNA has been introduced while the animal is still in its embryonic stage. In most cases, the transgenic approach aims at specific modifications of the genome, *e.g.*, by introducing whole transcriptional units into the genome, or by up- or down-regulating pre-existing cellular genes. The targeted character of certain of these procedures sets transgenic technologies apart from experimental methods in which random mutations are conferred to the germline, such as administration of chemical mutagens or treatment with ionizing solution.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs comprising one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to 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. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may comprise amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural

processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may comprise many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter, *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth. Enzymol.* (1990) 182:626-646 and Rattan, *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one

encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

5 All publications including, but not limited to, patents and patent applications, cited in this specification or to which this patent application claims priority, are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Examples

10 Example 1:

PSP-1 Phenotypic Analysis

The PSP-1 N4F1 heterozygous animals display no overt phenotype. By two weeks of age N4F1 homozygous animals are noticeably smaller than their litter-mates. Further, when moving, the homozygous animals display an abnormal shifting gait and tend to clench their paws and appear
15 unstable on their feet. By 3-4 weeks this instability is more pronounced and have an impaired righting reflex. The impairments are so severe that the animals must be culled.

Example 2:

Gross Pathology

Macroscopic organ lesions were limited to moderate thymic atrophy present in one male homozygous
20 animal.

Example 3:

Histopathology

Significant microscopic changes were not apparent in heterozygous PSP-1 mice when compared to age-matched littermate controls. Lesions were present in one male homozygous animal and included
25 moderate to marked cortical atrophy of the thymus and moderately reduced cellularity of the bone marrow. Both myeloid and erythroid precursors were present in the bone marrow. However, these lesions were not present in the other homozygous animals evaluated. Thymic and myeloid tissues of the other age-matched homozygous male and female animals were normal. Testes of both male
30 homozygous animals were considered immature, when compared to age-matched heterozygous and control animals.

What is claimed is:

1. An isolated polynucleotide selected from the group consisting of:

5 (i) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having at least a 95% identity to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2;

(ii) an isolated polynucleotide comprising a nucleotide sequence having at least a 95% identity over its entire length to a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

10 (iii) an isolated polynucleotide comprising a nucleotide sequence having at least a 95% identity to that of SEQ ID NO:1 over the entire length of SEQ ID NO:1;

(iv) an isolated polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(v) an isolated polynucleotide that is the polynucleotide of SEQ ID NO:1; or

15 (vi) an isolated polynucleotide with a nucleotide sequence of at least 100 nucleotides in length obtained by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof of at least 15 nucleotides; or a nucleotide sequence complementary to said isolated polynucleotide.

2. An isolated polypeptide selected from the group consisting of:

20 (i) an isolated polypeptide having at least a 95% identity to the amino acid sequence of SEQ ID NO: 2 over the entire length of SEQ ID NO:2;

(ii) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2; or

(iii) an isolated polypeptide that is the amino acid sequence of SEQ ID NO:2.

25 3. A method for screening to identify compounds that stimulate or that inhibit a function or level of the polypeptide of Claim 2, comprising a method selected from the group consisting of:

30 (a) measuring or, quantitatively or qualitatively, detecting the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

- (b) measuring the competition of the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presence of a labeled competitor;
- 5 (c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;
- (d) mixing a candidate compound with a solution comprising a polypeptide of Claim 2, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a to a control mixture which contains no candidate compound; or
- 10 (e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells.
4. An agonist or an antagonist of the polypeptide of Claim 2.
- 15 5. An agonist or an antagonist of the *Mus musculus* htra2 identified by the method of Claim 3.
6. An expression system comprising a polynucleotide capable of producing a polypeptide of Claim 2 when said expression system is present in a compatible host cell.
- 20 7. A process for producing a recombinant host cell comprising the step of introducing the expression vector of Claim 6 into a cell, such that the host cell, under appropriate culture conditions, produces said polypeptide.
8. A recombinant host cell produced by the process of Claim 7.
- 25 9. A membrane of a recombinant host cell of Claim 8 expressing said polypeptide.
10. A process for producing a polypeptide comprising culturing a host cell of Claim 9 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the
- 30 culture.

SEQUENCE LISTING

<110> SmithKline Beecham Corporation
SmithKline Beecham plc

<120> MURINE HTRA2

<130> GP-70621WO

<140> TO BE ASSIGNED

<141> 2000-05-19

<150> TO BE ASSIGNED

<151> 2000-05-18

<150> 60/134,886

<151> 1999-05-19

<160> 2

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 1377

<212> DNA

<213> MUS MUSCULUS

<400> 1

atggctgcgc	tgaaagcggg	gcgggggtgca	aactggagcc	ttcgggcatg	gcgggctttg	60
gggggcatct	tctggaggaa	gcctcccctg	ttagctcctg	acctccgggc	tctgctgacg	120
tcaggaactc	ctgactctca	gatctggatg	acttatggga	ccccagtct	cccgccag	180
gtgcctgagg	ggttccttgc	atcacgggca	gatctgacgt	ctaggacccc	ggatctctgg	240
gcacgattga	atgtggggac	ttcaggctcc	agtgaccagg	aggcccgcag	gagccctgga	300
agccgtcggc	gcgaatggct	ggcgggtggcg	gtgggcgcag	gaggtgcagt	ggtgctgctg	360
ttgtgggggt	ggggtcgggg	tctttccacg	gtgctggctg	ctgttcctgc	tccgccaccc	420
acttctcccc	ggagccagta	caatttcac	gcagatgtgg	tggagaagac	agcccctgct	480
gtggtctata	tcgagatcct	agaccggcac	cctttctccg	gccgtgaagt	ccccatctca	540
aacggatcag	gattcgtagt	ggcttcagat	gggctcatcg	ttaccaacgc	ccacgtgggtg	600
gctgatcggc	gccgagtacg	agtgaggctg	cctagcgggg	atacttatga	ggccatggtc	660
acagctgtgg	atcccgtagc	agacattgcc	acactgagga	ttcaaaccac	ggagcctctt	720
cccacactgc	ccctcggccg	ctctgctgat	gtccggcaag	gggagtttgt	tgttgccatg	780
ggaagcccct	ttgcaactgca	gaacacgac	acatctggta	ttgtcagctc	tgctcagcgc	840
ccagccaggg	acctgggact	ccctcaaac	aacgtggaat	acattcagac	cgatgcagct	900
attgattttg	gaaattctgg	tggctcccctg	gttaacctgg	atggggaggt	gattggagtg	960
aacaccatga	aggtgacagc	tggaatctcc	tttgccatcc	cttctgatcg	ccttagggag	1020
tttctgcatc	gcggggaaaa	gaaaaattcc	tggtttggaa	ccagtgggtc	ccagcgcgcg	1080
tacattggag	tgatgatgct	gaccctgact	cccagcatcc	ttattgaact	acagctccgt	1140
gagccaagct	tccctgatgt	tcagcatggt	gtcctcattc	ataaagttat	cctgggctcc	1200
cccgcacaca	gggctggctc	gcggcctggt	gatgtgatct	tggccattgg	ggagaaattg	1260
gcacaaaatg	ctgaagatgt	ttatgaagct	gttcgaaccc	aatcacagct	ggcagtgccg	1320
atccggcgcg	gatcagagac	actgatctta	tatgtgaccc	ccgaggtcac	agaatga	1377

<210> 2

<211> 458

<212> PRT

<213> MUS MUSCULUS

<400> 2

2/2

Met Ala Ala Leu Lys Ala Gly Arg Gly Ala Asn Trp Ser Leu Arg Ala
 1 5 10 15
 Trp Arg Ala Leu Gly Gly Ile Phe Trp Arg Lys Pro Pro Leu Leu Ala
 20 25 30
 Pro Asp Leu Arg Ala Leu Leu Thr Ser Gly Thr Pro Asp Ser Gln Ile
 35 40 45
 Trp Met Thr Tyr Gly Thr Pro Ser Leu Pro Ala Gln Val Pro Glu Gly
 50 55 60
 Phe Leu Ala Ser Arg Ala Asp Leu Thr Ser Arg Thr Pro Asp Leu Trp
 65 70 75 80
 Ala Arg Leu Asn Val Gly Thr Ser Gly Ser Ser Asp Gln Glu Ala Arg
 85 90 95
 Arg Ser Pro Gly Ser Arg Arg Arg Glu Trp Leu Ala Val Ala Val Gly
 100 105 110
 Ala Gly Gly Ala Val Val Leu Leu Leu Trp Gly Trp Gly Arg Gly Leu
 115 120 125
 Ser Thr Val Leu Ala Ala Val Pro Ala Pro Pro Pro Thr Ser Pro Arg
 130 135 140
 Ser Gln Tyr Asn Phe Ile Ala Asp Val Val Glu Lys Thr Ala Pro Ala
 145 150 155 160
 Val Val Tyr Ile Glu Ile Leu Asp Arg His Pro Phe Ser Gly Arg Glu
 165 170 175
 Val Pro Ile Ser Asn Gly Ser Gly Phe Val Val Ala Ser Asp Gly Leu
 180 185 190
 Ile Val Thr Asn Ala His Val Val Ala Asp Arg Arg Arg Val Arg Val
 195 200 205
 Arg Leu Pro Ser Gly Asp Thr Tyr Glu Ala Met Val Thr Ala Val Asp
 210 215 220
 Pro Val Ala Asp Ile Ala Thr Leu Arg Ile Gln Thr Lys Glu Pro Leu
 225 230 235 240
 Pro Thr Leu Pro Leu Gly Arg Ser Ala Asp Val Arg Gln Gly Glu Phe
 245 250 255
 Val Val Ala Met Gly Ser Pro Phe Ala Leu Gln Asn Thr Ile Thr Ser
 260 265 270
 Gly Ile Val Ser Ser Ala Gln Arg Pro Ala Arg Asp Leu Gly Leu Pro
 275 280 285
 Gln Asn Asn Val Glu Tyr Ile Gln Thr Asp Ala Ala Ile Asp Phe Gly
 290 295 300
 Asn Ser Gly Gly Pro Leu Val Asn Leu Asp Gly Glu Val Ile Gly Val
 305 310 315 320
 Asn Thr Met Lys Val Thr Ala Gly Ile Ser Phe Ala Ile Pro Ser Asp
 325 330 335
 Arg Leu Arg Glu Phe Leu His Arg Gly Glu Lys Lys Asn Ser Trp Phe
 340 345 350
 Gly Thr Ser Gly Ser Gln Arg Arg Tyr Ile Gly Val Met Met Leu Thr
 355 360 365
 Leu Thr Pro Ser Ile Leu Ile Glu Leu Gln Leu Arg Glu Pro Ser Phe
 370 375 380
 Pro Asp Val Gln His Gly Val Leu Ile His Lys Val Ile Leu Gly Ser
 385 390 395 400
 Pro Ala His Arg Ala Gly Leu Arg Pro Gly Asp Val Ile Leu Ala Ile
 405 410 415
 Gly Glu Lys Leu Ala Gln Asn Ala Glu Asp Val Tyr Glu Ala Val Arg
 420 425 430
 Thr Gln Ser Gln Leu Ala Val Arg Ile Arg Arg Gly Ser Glu Thr Leu
 435 440 445
 Ile Leu Tyr Val Thr Pro Glu Val Thr Glu
 450 455

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/14037**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : C12N 9/50, 1/20, 15/00, C12P 21/06

US CL : 435/69.1, 183, 219, 252.1, 320.1

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)

U.S. : 435/69.1, 183, 219, 252.1, 320.1

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 practicable, search terms used)

MEDLINE, EMBASE, CAPLUS, BIOSIS, SCISEARCH, CANCERLIT

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PALLEN et al. The HtrA family of serine proteases. Molecular Microbiology. October 1997, Vol. 26, No.2, pages 209-221, see entire article.	1-2, 6-10
Y	HU et al. Human HtrA, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage. Journal of Biological Chemistry. December 1998, Vol. 273, No. 51, pages 34406-34412, see abstract and figure 7.	1-2 and 6-10
X	Database EST in GenBank, AI481710, Marra, M. et al. The Washington University-HHMI Mouse EST project. 08 July 1998.	1 and 6-10

 Further documents are listed in the continuation of Box C.
 See patent family annex.

* Special categories of cited documents:	*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
A document defining the general state of the art which is not considered to be of particular relevance	*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
E earlier document published on or after the international filing date	*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
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)	*Z* document member of the same patent family
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	

Date of the actual completion of the international search

01 AUGUST 2000

Date of mailing of the international search report

30 AUG 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer


MANJUNATH RAO

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/14037

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Database EST in GenBank, AA289213, Marra et al. The Washington U-HHMI Mouse EST project. 14 April 1997.	1 and 6-10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/14037

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international 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.:
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:

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 II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

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

Please See Extra 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, 2 and 6-10

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/14037

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1, 2 and 6-10, drawn to polynucleotides, polypeptides, vectors, host cells and method of making.

Group II, claim 3, drawn to a method of screening to identify compounds.

Group III, claims 4-5, drawn to an agonist or an antagonist.

The inventions listed as Groups I-III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Polynucleotides encoding htrA (high temperature requirement) polypeptide are known in prior art and thus the above invention considered as a whole does not contribute over the prior art.

(See: Pallen et al.(1997) The HtrA family of serine proteases. Molecular Microbiology, Vol. 26(2):209-221).