

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number
WO 02/08253 A2

(51) International Patent Classification⁷: **C07K 14/00**

(21) International Application Number: PCT/EP01/08210

(22) International Filing Date: 17 July 2001 (17.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00116044.9 26 July 2000 (26.07.2000) EP

(71) Applicant (*for all designated States except US*): **MERCK PATENT GMBH** [DE/DE]; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **HOCK, Björn** [DE/DE]; Nordstrasse 2a, 63477 Maintal (DE). **DÜCKER, Klaus** [DE/DE]; Ettesterstrasse 5, D-64291 Darmstadt (DE).

(74) Common Representative: **MERCK PATENT GMBH**; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

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

(54) Title: A NOVEL MEMBER OF THE EphA RECEPTOR FAMILY

(57) Abstract: EphA9 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing EphA9 polypeptides and polynucleotides in diagnostic assays.



WO 02/08253 A2

A novel member of the EphA receptor family

Field of the Invention

5 This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides sometimes hereinafter referred to as „novel member of the EphA receptor family (EphA9)“, to their use in diagnosis and in identifying compounds that may be agonists, 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 as a means to identify genes and gene products as therapeutic targets is rapidly superceding 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.

20 Functional genomics relies heavily on high-throughput DNA sequencing technologies and 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 characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

Summary of the Invention

30 The present invention relates to EphA9, in particular EphA9 polypeptides and EphA9 polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods of treatment of certain diseases, including, but not

limited to, Walker-Warburg syndrome (WWS), muscle-eye-brain disease (Cormand, B. et al., Am. J. Hum. Genet. 64, 126-35, 1999), Schwartz-Jampel syndrome, type I (Nicole, S. et al., Hum. Molec. Genet. 4, 1633-1636, 1995), cancer, neurodegenerative disorders, neovasculogenesis, neoangiogenesis, metastasis, neuronal plasticity disorders, disorders associated with improper or undesirable axon outgrowth or orientation, ischemia, stroke, immunological disorders, abnormal or disturbed wound healing, cardiovascular disorders, arteriosclerosis, vasculotitis, arthritis, ., hereinafter referred to as " diseases of the invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (e.g., inhibitors) using the materials' provided by the invention, and treating conditions associated with EphA9 imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate EphA9 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to EphA9 polypeptides. Such polypeptides include:

- (a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;
- (b) a polypeptide comprising a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (c) a polypeptide comprising the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (d) a polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (e) the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8; and

(f) a polypeptide having or comprising a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

5 (g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to be members of the ephrin type A receptor family family of polypeptides. In vertebrates, the 14 Eph receptors and 8 ephrin ligands comprise two major specificity subclasses: EphA receptors bind to GPI-anchored ephrinA ligands, and
10 EphB receptors (and EphA4) that bind ephrinB ligands, which possess a transmembrane domain and a short cytoplasmic region (Gale et al., Neuron 17, 9-19, 1996, Eph Nomenclature committee, Cell 90, 403-404, 1997).

15 Expression studies have shown, that Eph receptors and ephrins are expressed in many, if not all embryonic and adult tissues (Gale et al., Neuron 17, 9-19, 1996). This data suggests that they have widespread roles. In some cases, interacting Eph receptors and ephrins have complementary domains or gradients of expression, whereas in other
20 cases, their domains of expression overlap. Functional studies have provided important insights into the role of complementary expression in which interactions between Eph receptors and ephrins occur at boundaries or along a gradient. Juxtacrine interactions between Eph receptors and ephrins on opposing cells were initially implicated in
25 patterning of the brain and somites, and in the process of neural cell guidance during embryonic development (for further reading refer to Mellitzer G. et al., Curr. Opin. Neurobiol. 10, 400 -408, 2000, and references therein). Also, particular members of the Eph receptors and their ligands have been shown to function in the development of the
30 vascular system and play a pivotal role in tumor angiogenesis (Illan, N. and Madri, J.A., Curr. Opin. Biotech. 10, 536-540, 1999).

Herewith we disclose the sequence information of a novel human gene encoding a new member of the Ephrin receptor family, EphA9 and the

identification of 4 differential splicing isoforms. Due to the absence of amino acids known to be crucial for kinase activity it is very likely that this gene encodes an ephrin receptor with no intrinsic catalytic kinase activity (Hanks, S.K. et al., *Science*. 241, 42-52, 1988). Additionally, the two conserved autophosphorylation sites in the juxtamembrane region are substituted by non-phosphorylatable amino acids (Holland, S.J. et al., *EMBO J.* 16, 3877-3888, 1997; Hock, B. et al., *Oncogene* 17, 255-260, 1998).

The identified splicing isoforms affect the cytoplasmic region of the receptor. One splicing event leads to the insertion of a new domain in the juxtamembrane region of the receptor, a region known to be crucial for protein-protein interaction and signalling (see above). A further differential splicing event leads to an insertion in the middle of the ATP binding motif of the cytoplasmic kinase domain, which adds a novel stretch of 29 amino acids before it ends in a translational stop signal. Thereby, this insertion removes nearly the entire cytoplasmic kinase domain.

Herewith we disclose the sequence information of 4 differential isoforms of human EphA9.

The biological properties of the EphA9 are hereinafter referred to as "biological activity of EphA9" or "EphA9 activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of EphA9.

Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. 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.

Preferred fragments of polypeptides of the present invention include a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2, or a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO: 2. Preferred fragments are biologically active fragments that mediate the biological activity of EphA9, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in 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. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or 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.

Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation from naturally occurring sources, from genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesizers, or a combination of such methods.. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to EphA9 polynucleotides. Such polynucleotides include:

- (a) a polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

- (b) a polynucleotide comprising the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;
- (c) a polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;
- (d) the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;
- (e) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (f) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (g) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (h) a polynucleotide encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;
- (i) a polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;
- (j) a polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8; and
- polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include a polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1, or a polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1.

Preferred variants of polynucleotides of the present invention include splice variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8 and 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 acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

(a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

(b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

(c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7; or

(d) is the RNA transcript of the DNA sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

and RNA polynucleotides that are complementary thereto.

5 The polynucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or
SEQ ID NO:5 and/or SEQ ID NO:7 shows homology with HUMRPTK (Fox,
G.M. et al., Oncogene 10, 897-905, 1995). The polynucleotide sequence of
SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID
10 NO:7 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2
and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8. The
polynucleotide sequence encoding the polypeptide of SEQ ID NO:2
and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8 may be
identical to the polypeptide encoding sequence of SEQ ID NO:1 and/or
15 SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7 or it may be a
sequence other than SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID
NO:5 and/or SEQ ID NO:7, which, as a result of the redundancy
(degeneracy) of the genetic code, also encodes the polypeptide of SEQ
ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.
20 The polypeptide of the SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID
NO:6 and/or SEQ ID NO:8 is related to other proteins of the ephrin type A
receptor family family, having homology and/or structural similarity with
Q15375 (Fox, G.M. et al., Oncogene 10, 897-905, 1995).

25 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
EphA9 activity.

30 Polynucleotides of the present invention may be obtained using standard
cloning and screening techniques from a cDNA library derived from mRNA
in cells of human B-cells, brain, colon tumor, colon adenocarcinoma, kidney
tumor, carcinoid lung, fetal lung, testis, (see for instance, Sambrook *et al.*,
Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor
Laboratory Press, Cold Spring Harbor, N.Y. (1989)). 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.

35 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 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 contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used 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 genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a 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; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such 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 polynucleotides, preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7 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 does not extend all the way through to the 5' terminus. This 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 polymerisation reaction), failing to complete a DNA copy of the mRNA template during first 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 Amplification of cDNA ends (RACE) (see, for example, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) 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 anneals further 5' in the known gene sequence). The products of this reaction can then be analysed 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 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 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. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al. (*ibid*). 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.

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 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 contain 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 polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, (*ibid*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

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 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. 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 intracellular synthesis, isolation and/or purification.

Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7 in the cDNA or genomic sequence and 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 well known in the art.

Nucleic acids 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 it may be amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled EphA9 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence difference 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 (see, for instance, 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).

An array of oligonucleotide probes comprising EphA9 polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of *e.g.*, genetic mutations. Such arrays are preferably high density arrays or grids. 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, 274, 610-613 (1996) and other references cited therein.

Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. 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 protein, such as a polypeptide of the present invention, in a sample derived

from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

The polynucleotide sequences of the present invention are valuable for chromosome localisation studies. The sequence 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 first step in correlating those sequences with 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 relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise human chromosomal localisations for a genomic sequence (gene

fragment etc.) can be determined using Radiation Hybrid (RH) Mapping (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, *Nature Genetics* 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (*Hum Mol Genet* 1996 Mar;5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These scores are compared with scores created using PCR products from genomic sequences of known location. This comparison is conducted at <http://www.genome.wi.mit.edu/>. The gene of the present invention maps to human chromosome 1p34.1-1p34.4.

The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and include in situ hybridisation techniques to clones arrayed on a grid, such as cDNA microarray hybridisation (Schena *et al*, *Science*, 270, 467-470, 1995 and Shalon *et al*, *Genome Res*, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the normal function of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate

expression thereof in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The polypeptides of the present invention are expressed in B-cells, brain, colon tumor, colon adenocarcinoma, kidney tumor, carcinoid lung, fetal lung, testis.

A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. 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.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a mammal may also be induced by a method comprises delivering a polypeptide of the present invention *v/a* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. The formulations 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 vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the

diseases of the invention hereinbefore mentioned. It is therefore useful to to identify compounds that stimulate or inhibit the function or level of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those that stimulate or inhibit the function or level of the polypeptide. Such methods identify agonists or antagonists that may be employed for therapeutic and prophylactic purposes for such diseases of the invention as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, collections of chemical compounds, and natural product mixtures. Such agonists or antagonists so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; a structural or functional mimetic thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule.

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, the screening method may involve measuring or detecting (qualitatively or quantitatively) the competitive binding of a candidate compound to the polypeptide against a labeled competitor (e.g. agonist or antagonist). Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring a EphA9 activity in the mixture, and comparing the EphA9 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 micotiter 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 EphA9 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for 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)).

10 Screening techniques

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents that may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

A polypeptide of the present invention may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ¹²⁵I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide that compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of antagonists of polypeptides of the present invention include antibodies or, in some cases, oligonucleotides or proteins that are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, *e.g.*, a fragment of the ligands, substrates, receptors, enzymes, etc.; or a small molecule that bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Screening methods may also involve the use of transgenic technology and EphA9 gene. The art of constructing transgenic animals is well established. For example, the EphA9 gene may be introduced through microinjection into the male pronucleus of fertilized oocytes, retroviral transfer into pre- or post-implantation embryos, or injection of genetically modified, such as by electroporation, embryonic stem cells into host blastocysts. Particularly useful transgenic animals are so-called "knock-in" animals in which an animal gene is replaced by the human equivalent within the genome of that animal. Knock-in transgenic animals are useful in the drug discovery process, for target validation, where the compound is specific for the human target. Other useful transgenic animals are so-called "knock-out" animals in which the expression of the animal ortholog of a polypeptide of the present invention and encoded by an endogenous DNA sequence in a cell is partially or completely annulled. The gene knock-out may be targeted to specific cells or tissues, may occur only in certain cells or tissues as a consequence of the limitations of the technology, or may occur in all, or substantially all, cells in the animal. Transgenic animal technology also offers a whole animal expression-cloning system in which introduced genes are expressed to give large amounts of polypeptides of the present invention

Screening kits for use in the above described methods form a further aspect of the present invention. Such screening kits comprise:

- (a) a polypeptide of the present invention;
- (b) a recombinant cell expressing a polypeptide of the present invention;
- (c) a cell membrane expressing a polypeptide of the present invention; or
- (d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.

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

5

Glossary

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

10 "Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an

Fab or other immunoglobulin expression library.

15 "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.

25 "Polynucleotide" generally refers to any polyribonucleotide (RNA) or polydeoxribonucleotide (DNA), which may be unmodified 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 containing one or more modified bases and

30

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 polypeptide 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 contain 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 that 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 contain 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, biotinylation, 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 cystine, 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, 1-12, in Post-translational 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, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci, 663, 48-62, 1992).

"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 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7..

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the 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 the reference polypeptide. Generally, alterations 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, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, 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. Also included as variants are polypeptides having one or more post-translational modifications, for instance glycosylation, phosphorylation, methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

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

"Polymorphism" refers to a variation in nucleotide sequence (and encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

"Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele Specific Primers.

"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.

"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.

5 "% Identity" - 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
10 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
15 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
20 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
25 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
30 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
35 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 Neddleman 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 reference polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence

may include on average up to five differences per each 100 nucleotides of the referen sequence. Such differences are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences 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 an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5 in every 100 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 values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino acids of the reference sequence. Such differences are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences 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 polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following equation:

$$n_a \leq x_a - (x_a \cdot I),$$

in which:

n_a is the number of nucleotide or amino acid differences,

x_a is the total number of nucleotides or amino acids in SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7 or SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8, respectively,

I is the Identity Index ,

$\bullet$ is the symbol for the multiplication operator, and

in which any non-integer product of x_a and I is rounded down to the nearest integer prior to subtracting it from x_a .

"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 a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotide or polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, unrelated, fused genes or fragments thereof. Examples have been disclosed in US 5541087, 5726044. In the case of Fc-EphA9, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for performing the functional expression of Fc-EphA9 or fragments of - EphA9, to improve pharmacokinetic properties of such a fusion protein when used for therapy and to generate a dimeric EphA9. The Fc-EphA9 DNA construct comprises in 5' to 3' direction, a secretion cassette, i.e. a signal sequence that triggers export from a mammalian cell, DNA encoding an immunoglobulin Fc region fragment, as a fusion partner, and a DNA encoding EphA9 or fragments thereof. In some uses it would be desirable to be able to alter the intrinsic functional properties (complement binding, Fc-Receptor binding) by mutating the functional Fc

sides while leaving the rest of the fusion protein untouched or delete the Fc part completely after expression.

5 All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which
10 this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

Claims

1. A polypeptide selected from one of the groups consisting of:

(a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

(b) a polypeptide comprising a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

(c) a polypeptide having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8; and

(d) the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8 and

(e) fragments and variants of such polypeptides in (a) to (d).

2. The polypeptide as claimed in claim 1 comprising the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.

3. The polypeptide as claimed in claim 1 which is the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.

4. A polynucleotide selected from one of the groups consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

(b) a polynucleotide having at least 95% identity to the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

5 (c) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

10 (d) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8;

15 (e) an polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof having at least 15 nucleotides;

(f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);

or a polynucleotide sequence complementary to said polynucleotide

20 and polynucleotides that are variants and fragments of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

5. A polynucleotide as claimed in claim 4 selected from the group consisting of:

25 (a) a polynucleotide comprising the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

(b) the polynucleotide of SEQ ID NO:1 and/or SEQ ID NO:3 and/or SEQ ID NO:5 and/or SEQ ID NO:7;

30 (c) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8; and

(d) a polynucleotide encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 and/or SEQ ID NO:6 and/or SEQ ID NO:8.

5 6. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression vector is present in a compatible host cell.

7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of claim 1.

10

8. A process for producing a polypeptide of claim 1 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

15

9. A fusion protein consisting of the Immunoglobulin Fc-region and any one polypeptide of claim 1.

20 10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.

11. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of claim 1 comprising a method selected from the group consisting of:

25 (a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing 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 binding of a candidate compound to the polypeptide (or to the cells or membranes expressing 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 expressing the polypeptide;
- 10 (d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or
- (e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay, and
- 15 (f) producing said compound according to biotechnological or chemical standard techniques.

1/23

SEQUENCE LISTING

□

5 <110> Merrck Patent GmbH

□

□

10 <120> Novel protein of the EphA receptor family

□

□

15 <130> EphA9kdws

□

□

<140>

□

<141>

20 □

□

<160> 8

□

25 □

<170> PatentIn Ver. 2.1

<210> 1

30 <211> 2955

<212> DNA

<213> Homo sapiens

<220>

35 <221> CDS

<222> (1)..(2925)

<223> Full length kinase

<400> 1

40 atc ctc ctg gat tcc aaa gcc tcc cag gcc gag ctg ggc tgg act gca 48
 Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala
 1 5 10 15

45 ctg cca agt aat ggg tgg gag gag atc agc ggc gtg gat gaa cac gac 96
 Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp
 20 25 30

50 cgt ccc atc cgc acg tac caa gtg tgc aat gtg ctg gag ccc aac cag 144
 Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln
 35 40 45

55 gac aac tgg ctg cag act ggc tgg ata agc cgt ggc cgc ggg cag cgc 192
 Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg
 50 55 60

60 atc ttc gtg gaa ctg cag ttc aca ctc cgt gac tgc agc agc atc cct 240
 Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro
 65 70 75 80

60 ggc gcc gcg ggt acc tgc aag gag acc ttc aac gtc tac tac ctg gaa 288

2/23

	Gly	Ala	Ala	Gly	Thr	Cys	Lys	Glu	Thr	Phe	Asn	Val	Tyr	Tyr	Leu	Glu	
					85					90					95		
5	act	gag	gcc	gac	ctg	ggc	cgt	ggg	cgt	ccc	cgc	cta	ggc	ggc	agc	cgg	336
	Thr	Glu	Ala	Asp	Leu	Gly	Arg	Gly	Arg	Pro	Arg	Leu	Gly	Gly	Ser	Arg	
				100					105					110			
10	ccc	cgc	aaa	atc	gac	acg	atc	gcg	gcg	gac	gag	agc	ttc	acg	cag	ggc	384
	Pro	Arg	Lys	Ile	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser	Phe	Thr	Gln	Gly	
			115					120					125				
15	gac	ctg	ggt	gag	cgc	aag	atg	aag	ctg	aac	aca	gag	gtg	cgc	gag	atc	432
	Asp	Leu	Gly	Glu	Arg	Lys	Met	Lys	Leu	Asn	Thr	Glu	Val	Arg	Glu	Ile	
			130				135					140					
20	gga	ccg	ctc	agc	cgg	cgg	ggt	ttc	cac	ctg	gcc	ttt	cag	gac	gtg	ggc	480
	Gly	Pro	Leu	Ser	Arg	Arg	Gly	Phe	His	Leu	Ala	Phe	Gln	Asp	Val	Gly	
	145					150					155				160		
25	gca	tgc	gtg	gcg	ctt	gtc	tcg	gtg	cgc	gtc	tac	tac	aag	cag	tgc	cgc	528
	Ala	Cys	Val	Ala	Leu	Val	Ser	Val	Arg	Val	Tyr	Tyr	Lys	Gln	Cys	Arg	
					165				170						175		
30	gcc	acc	gtg	cgg	ggc	ctg	gcc	acg	ttc	cca	gcc	acc	gca	gcc	gag	agc	576
	Ala	Thr	Val	Arg	Gly	Leu	Ala	Thr	Phe	Pro	Ala	Thr	Ala	Ala	Glu	Ser	
				180					185				190				
35	gcc	ttc	tcc	aca	ctg	gtg	gaa	gtg	gcc	gga	acg	tgc	gtg	gcg	cac	tcg	624
	Ala	Phe	Ser	Thr	Leu	Val	Glu	Val	Ala	Gly	Thr	Cys	Val	Ala	His	Ser	
			195				200						205				
40	gaa	ggg	gag	cct	ggc	agc	ccc	cca	cgc	atg	cac	tgc	ggc	gcc	gac	ggc	672
	Glu	Gly	Glu	Pro	Gly	Ser	Pro	Pro	Arg	Met	His	Cys	Gly	Ala	Asp	Gly	
		210					215					220					
45	gag	tgg	ctg	gtg	cct	gtg	ggc	cgc	tgc	agc	tgc	agc	gcg	gga	ttc	cag	720
	Glu	Trp	Leu	Val	Pro	Val	Gly	Arg	Cys	Ser	Cys	Ser	Ala	Gly	Phe	Gln	
						230					235					240	
50	gag	cgt	ggt	gac	ttc	tgc	gaa	gcc	tgt	ccc	cca	ggg	ttt	tac	aag	gtg	768
	Glu	Arg	Gly	Asp	Phe	Cys	Glu	Ala	Cys	Pro	Pro	Gly	Phe	Tyr	Lys	Val	
					245					250					255		
55	tcc	ccg	cgg	cgg	ccc	ctc	tgc	tca	ccg	tgc	cca	gag	cac	agc	cgg	gcc	816
	Ser	Pro	Arg	Arg	Pro	Leu	Cys	Ser	Pro	Cys	Pro	Glu	His	Ser	Arg	Ala	
				260					265					270			
60	ctg	gaa	aac	gcc	tcc	acc	ttc	tgc	gtg	tgc	cag	gac	agc	tat	gcg	cgc	864
	Leu	Glu	Asn	Ala	Ser	Thr	Phe	Cys	Val	Cys	Gln	Asp	Ser	Tyr	Ala	Arg	
			275				280						285				
65	tca	ccc	acc	gac	ccg	ccc	tcg	gct	tcc	tgc	acc	cgg	ccg	ccg	tcg	gcg	912
	Ser	Pro	Thr	Asp	Pro	Pro	Ser	Ala	Ser	Cys	Thr	Arg	Pro	Pro	Ser	Ala	
			290				295					300					
70	ccg	cgg	gac	ctg	cag	tac	agc	ctg	agc	cgc	tcg	ccg	ctg	gtg	ctg	cga	960
	Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg	
						310					315					320	
75	ctg	cgc	tgg	ctg	ccg	ccg	gcc	gac	tcg	gga	ggc	cgc	tcg	gac	gtc	acc	1008
	Leu	Arg	Trp	Leu	Pro	Pro	Ala	Asp	Ser	Gly	Gly	Arg	Ser	Asp	Val	Thr	

3/23

	325							330							335							
5	tac	tcg	ctg	ctg	tgc	ctg	cgc	tgc	ggc	cgc	gag	ggc	ccg	gcg	ggc	gcc	1056					
	Tyr	Ser	Leu	Leu	Cys	Leu	Arg	Cys	Gly	Arg	Glu	Gly	Pro	Ala	Gly	Ala						
				340					345					350								
10	tgc	gag	ccg	tgc	ggg	ccg	cgc	gtg	gcc	ttc	cta	ccg	cgc	cag	gca	ggg	1104					
	Cys	Glu	Pro	Cys	Gly	Pro	Arg	Val	Ala	Phe	Leu	Pro	Arg	Gln	Ala	Gly						
			355					360					365									
15	ctg	cgg	gag	cga	gcc	gcc	acg	ctg	ctg	cac	ctg	cgg	ccc	ggc	gcg	cgc	1152					
	Leu	Arg	Glu	Arg	Ala	Ala	Thr	Leu	Leu	His	Leu	Arg	Pro	Gly	Ala	Arg						
			370				375					380										
20	tac	acc	gtg	cgc	gtg	gcc	gcg	ctc	aac	ggc	gtc	tcg	ggc	ccg	gcg	gcc	1200					
	Tyr	Thr	Val	Arg	Val	Ala	Ala	Leu	Asn	Gly	Val	Ser	Gly	Pro	Ala	Ala						
						390					395					400						
25	gcc	gcg	gga	acc	acc	tac	gcg	cag	gtc	acc	gtc	tcc	acc	ggg	ccc	ggg	1248					
	Ala	Ala	Gly	Thr	Thr	Tyr	Ala	Gln	Val	Thr	Val	Ser	Thr	Gly	Pro	Gly						
				405						410					415							
30	gcg	ccc	tgg	gag	gag	gat	gag	atc	cgc	agg	gac	cga	gtg	gaa	ccc	cag	1296					
	Ala	Pro	Trp	Glu	Glu	Asp	Glu	Ile	Arg	Arg	Asp	Arg	Val	Glu	Pro	Gln						
				420					425					430								
35	agc	gtg	tcc	ctg	tcg	tgg	cgg	gag	ccc	atc	cct	gcc	gga	gcc	cct	ggg	1344					
	Ser	Val	Ser	Leu	Ser	Trp	Arg	Glu	Pro	Ile	Pro	Ala	Gly	Ala	Pro	Gly						
			435					440					445									
40	gcc	aat	gac	acg	gag	tac	gag	atc	cga	tac	tac	gag	aag	ggg	cag	agt	1392					
	Ala	Asn	Asp	Thr	Glu	Tyr	Glu	Ile	Arg	Tyr	Tyr	Glu	Lys	Gly	Gln	Ser						
			450				455					460										
45	gag	cag	act	tac	tcc	atg	gtg	aag	aca	ggg	gcg	ccc	aca	gtc	acc	gtc	1440					
	Glu	Gln	Thr	Tyr	Ser	Met	Val	Lys	Thr	Gly	Ala	Pro	Thr	Val	Thr	Val						
						470					475					480						
50	acc	aac	ctg	aag	ccg	gct	acc	cgc	tac	gtc	ttt	cag	atc	cgg	gcc	gct	1488					
	Thr	Asn	Leu	Lys	Pro	Ala	Thr	Arg	Tyr	Val	Phe	Gln	Ile	Arg	Ala	Ala						
					485					490					495							
55	tcc	ccg	ggg	cca	tcc	tgg	gag	gcc	cag	agt	ttt	aac	ccc	agc	att	gaa	1536					
	Ser	Pro	Gly	Pro	Ser	Trp	Glu	Ala	Gln	Ser	Phe	Asn	Pro	Ser	Ile	Glu						
				500					505					510								
60	gta	cag	acc	ctg	ggg	gag	gct	gcc	tca	ggg	tcc	agg	gac	cag	agc	ccc	1584					
	Val	Gln	Thr	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Ser	Arg	Asp	Gln	Ser	Pro						
			515					520					525									
65	gcc	att	gtc	gtc	acc	gta	gtg	acc	atc	tcg	gcc	ctc	ctc	gtc	ctg	ggc	1632					
	Ala	Ile	Val	Val	Thr	Val	Val	Thr	Ile	Ser	Ala	Leu	Leu	Val	Leu	Gly						
			530				535					540										
70	tcc	gtg	atg	agt	gtg	ctg	gcc	att	tgg	agg	agg	ccc	tgc	agc	tat	ggc	1680					
	Ser	Val	Met	Ser	Val	Leu	Ala	Ile	Trp	Arg	Arg	Pro	Cys	Ser	Tyr	Gly						
						550					555					560						
75	aaa	gga	gga	ggg	gat	gcc	cat	gat	gaa	gag	gag	ctg	tat	ttc	cac	ttc	1728					
	Lys	Gly	Gly	Gly	Asp	Ala	His	Asp	Glu	Glu	Glu	Leu	Tyr	Phe	His	Phe						
					565					570					575							

4/23

	aaa gtc cca aca cgt cgc aca ttc ctg gac ccc cag agc tgt ggg gac	1776
	Lys Val Pro Thr Arg Arg Thr Phe Leu Asp Pro Gln Ser Cys Gly Asp	
	580 585 590	
5	ctg ctg cag gct gtg cat ctg ttc gcc aag gaa ctg gat gcg aaa agc	1824
	Leu Leu Gln Ala Val His Leu Phe Ala Lys Glu Leu Asp Ala Lys Ser	
	595 600 605	
10	gtc acg ctg gag agg agc ctt gga gga ggg cgg ttt ggg gag ctg tgc	1872
	Val Thr Leu Glu Arg Ser Leu Gly Gly Gly Arg Phe Gly Glu Leu Cys	
	610 615 620	
15	tgt ggc tgc ttg cag ctc ccc ggt cgc cag gag ctg ctc gta gcc gtg	1920
	Cys Gly Cys Leu Gln Leu Pro Gly Arg Gln Glu Leu Leu Val Ala Val	
	625 630 635 640	
20	cat atg ctg agg gac agc gcc tcc gac tca cag agg ctc ggc ttc ctg	1968
	His Met Leu Arg Asp Ser Ala Ser Asp Ser Gln Arg Leu Gly Phe Leu	
	645 650 655	
25	gcc gag gcc ctc acg ctg ggc cag ttt gac cat agc cac atc gtg cgg	2016
	Ala Glu Ala Leu Thr Leu Gly Gln Phe Asp His Ser His Ile Val Arg	
	660 665 670	
30	ctg gag ggc gtt gtt acc cga ggt agg gcc cgc acc ttg atg att gtc	2064
	Leu Glu Gly Val Val Thr Arg Gly Arg Ala Arg Thr Leu Met Ile Val	
	675 680 685	
35	acc gag tac atg agc cat ggg gcc ctg gac ggc ttc ctc agg cgg cac	2112
	Thr Glu Tyr Met Ser His Gly Ala Leu Asp Gly Phe Leu Arg Arg His	
	690 695 700	
40	gag ggg cag ctg gtg gct ggg caa ctg atg ggg ttg ctg cct ggg ctg	2160
	Glu Gly Gln Leu Val Ala Gly Gln Leu Met Gly Leu Leu Pro Gly Leu	
	705 710 715 720	
45	gca tca gcc atg aag tat ctg tca gag atg ggc tac gtt cac cgg ggc	2208
	Ala Ser Ala Met Lys Tyr Leu Ser Glu Met Gly Tyr Val His Arg Gly	
	725 730 735	
50	ctg gca gct cgc cat gtg ctg gtc agc agc gac ctt gtc tgc aag atc	2256
	Leu Ala Ala Arg His Val Leu Val Ser Ser Asp Leu Val Cys Lys Ile	
	740 745 750	
55	tct ggc ttc ggg cgg ggc ccc cgg gac cga tca gag gct gtc tac acc	2304
	Ser Gly Phe Gly Arg Gly Pro Arg Asp Arg Ser Glu Ala Val Tyr Thr	
	755 760 765	
60	act atg agt ggc cgg agc cca gcg cta tgg gcc gct ccc gag aca ctt	2352
	Thr Met Ser Gly Arg Ser Pro Ala Leu Trp Ala Ala Pro Glu Thr Leu	
	770 775 780	
65	cag ttt ggc cac ttc agc tct gcc agt gac gtg tgg agc ttc ggc atc	2400
	Gln Phe Gly His Phe Ser Ser Ala Ser Asp Val Trp Ser Phe Gly Ile	
	785 790 795 800	
70	atc atg tgg gag gtg atg gcc ttt ggg gag cgg cct tac tgg gac atg	2448
	Ile Met Trp Glu Val Met Ala Phe Gly Glu Arg Pro Tyr Trp Asp Met	
	805 810 815	

5/23

	tct ggc caa gac gtg atc aag gct gtg gag gat ggc ttc cgg ctg cca	2496
	Ser Gly Gln Asp Val Ile Lys Ala Val Glu Asp Gly Phe Arg Leu Pro	
	820 825 830	
5	ccc ccc agg aac tgt cct aac ctt ctg cac cga cta atg ctc gac tgc	2544
	Pro Pro Arg Asn Cys Pro Asn Leu Leu His Arg Leu Met Leu Asp Cys	
	835 840 845	
10	tgg cag aag gac cca ggt gag cgg ccc agg ttc tcc cag atc cac agc	2592
	Trp Gln Lys Asp Pro Gly Glu Arg Pro Arg Phe Ser Gln Ile His Ser	
	850 855 860	
15	atc ctg agc aag atg gtg cag gac cca gag ccc ccc aag tgt gcc ctg	2640
	Ile Leu Ser Lys Met Val Gln Asp Pro Glu Pro Pro Lys Cys Ala Leu	
	865 870 875 880	
20	act acc tgt ccc agg cct ccc acc cca cta gcg gac cgt gcc ttc tcc	2688
	Thr Thr Cys Pro Arg Pro Pro Thr Pro Leu Ala Asp Arg Ala Phe Ser	
	885 890 895	
	acc ttc ccc tcc ttt ggc tct gtg ggc gcg tgg ctg gag gcc ctg gac	2736
	Thr Phe Pro Ser Phe Gly Ser Val Gly Ala Trp Leu Glu Ala Leu Asp	
	900 905 910	
25	ctg tgc cgc tac aag gac agc ttc gcg gct gct ggc tat ggg agc ctg	2784
	Leu Cys Arg Tyr Lys Asp Ser Phe Ala Ala Ala Gly Tyr Gly Ser Leu	
	915 920 925	
30	gag gcc gtg gcc gag atg act gcc cag gac ctg gtg agc cta gcc atc	2832
	Glu Ala Val Ala Glu Met Thr Ala Gln Asp Leu Val Ser Leu Gly Ile	
	930 935 940	
35	tct ttg gct gaa cat cga gag gcc ctc ctc agc ggg atc agc gcc ctg	2880
	Ser Leu Ala Glu His Arg Glu Ala Leu Leu Ser Gly Ile Ser Ala Leu	
	945 950 955 960	
40	cag gca cga gtg ctc cag ctg cag ggc cag ggg gtg cag gtg tga	2925
	Gln Ala Arg Val Leu Gln Leu Gln Gly Gln Gly Val Gln Val	
	965 970 975	
	gtggacccca ttcttccaag gcaggactcc	2955
45	<210> 2	
	<211> 974	
	<212> PRT	
	<213> Homo sapiens	
	<400> 2	
50	Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala	
	1 5 10 15	
	Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp	
	20 25 30	
55	Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln	
	35 40 45	
	Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg	
	50 55 60	
	Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro	
	65 70 75 80	
60	Gly Ala Ala Gly Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu	
	85 90 95	

6/23

	Thr	Glu	Ala	Asp	Leu	Gly	Arg	Gly	Arg	Pro	Arg	Leu	Gly	Gly	Ser	Arg
				100					105					110		
	Pro	Arg	Lys	Ile	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser	Phe	Thr	Gln	Gly
			115					120					125			
5	Asp	Leu	Gly	Glu	Arg	Lys	Met	Lys	Leu	Asn	Thr	Glu	Val	Arg	Glu	Ile
		130					135					140				
	Gly	Pro	Leu	Ser	Arg	Arg	Gly	Phe	His	Leu	Ala	Phe	Gln	Asp	Val	Gly
		145				150					155					160
	Ala	Cys	Val	Ala	Leu	Val	Ser	Val	Arg	Val	Tyr	Tyr	Lys	Gln	Cys	Arg
10					165					170						175
	Ala	Thr	Val	Arg	Gly	Leu	Ala	Thr	Phe	Pro	Ala	Thr	Ala	Ala	Glu	Ser
				180					185						190	
	Ala	Phe	Ser	Thr	Leu	Val	Glu	Val	Ala	Gly	Thr	Cys	Val	Ala	His	Ser
			195					200					205			
15	Glu	Gly	Glu	Pro	Gly	Ser	Pro	Pro	Arg	Met	His	Cys	Gly	Ala	Asp	Gly
		210					215					220				
	Glu	Trp	Leu	Val	Pro	Val	Gly	Arg	Cys	Ser	Cys	Ser	Ala	Gly	Phe	Gln
		225				230					235					240
	Glu	Arg	Gly	Asp	Phe	Cys	Glu	Ala	Cys	Pro	Pro	Gly	Phe	Tyr	Lys	Val
20					245					250						255
	Ser	Pro	Arg	Arg	Pro	Leu	Cys	Ser	Pro	Cys	Pro	Glu	His	Ser	Arg	Ala
				260					265						270	
	Leu	Glu	Asn	Ala	Ser	Thr	Phe	Cys	Val	Cys	Gln	Asp	Ser	Tyr	Ala	Arg
			275					280					285			
25	Ser	Pro	Thr	Asp	Pro	Pro	Ser	Ala	Ser	Cys	Thr	Arg	Pro	Pro	Ser	Ala
		290					295					300				
	Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg
		305				310					315					320
	Leu	Arg	Trp	Leu	Pro	Pro	Ala	Asp	Ser	Gly	Gly	Arg	Ser	Asp	Val	Thr
30					325					330						335
	Tyr	Ser	Leu	Leu	Cys	Leu	Arg	Cys	Gly	Arg	Glu	Gly	Pro	Ala	Gly	Ala
				340					345					350		
	Cys	Glu	Pro	Cys	Gly	Pro	Arg	Val	Ala	Phe	Leu	Pro	Arg	Gln	Ala	Gly
			355					360					365			
35	Leu	Arg	Glu	Arg	Ala	Ala	Thr	Leu	Leu	His	Leu	Arg	Pro	Gly	Ala	Arg
		370					375					380				
	Tyr	Thr	Val	Arg	Val	Ala	Ala	Leu	Asn	Gly	Val	Ser	Gly	Pro	Ala	Ala
		385				390					395					400
	Ala	Ala	Gly	Thr	Thr	Tyr	Ala	Gln	Val	Thr	Val	Ser	Thr	Gly	Pro	Gly
40					405					410					415	
	Ala	Pro	Trp	Glu	Asp	Glu	Ile	Arg	Arg	Asp	Arg	Val	Glu	Pro	Gln	
				420				425					430			
	Ser	Val	Ser	Leu	Ser	Trp	Arg	Glu	Pro	Ile	Pro	Ala	Gly	Ala	Pro	Gly
				435				440					445			
45	Ala	Asn	Asp	Thr	Glu	Tyr	Glu	Ile	Arg	Tyr	Tyr	Glu	Lys	Gly	Gln	Ser
		450					455					460				
	Glu	Gln	Thr	Tyr	Ser	Met	Val	Lys	Thr	Gly	Ala	Pro	Thr	Val	Thr	Val
		465				470					475					480
	Thr	Asn	Leu	Lys	Pro	Ala	Thr	Arg	Tyr	Val	Phe	Gln	Ile	Arg	Ala	Ala
50					485					490					495	
	Ser	Pro	Gly	Pro	Ser	Trp	Glu	Ala	Gln	Ser	Phe	Asn	Pro	Ser	Ile	Glu
				500					505					510		
	Val	Gln	Thr	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Ser	Arg	Asp	Gln	Ser	Pro
			515				520						525			
55	Ala	Ile	Val	Val	Thr	Val	Val	Thr	Ile	Ser	Ala	Leu	Leu	Val	Leu	Gly
		530					535					540				
	Ser	Val	Met	Ser	Val	Leu	Ala	Ile	Trp	Arg	Arg	Pro	Cys	Ser	Tyr	Gly
		545				550					555					560
	Lys	Gly	Gly	Gly	Asp	Ala	His	Asp	Glu	Glu	Glu	Leu	Tyr	Phe	His	Phe
60					565					570					575	
	Lys	Val	Pro	Thr	Arg	Arg	Thr	Phe	Leu	Asp	Pro	Gln	Ser	Cys	Gly	Asp

7/23

			580				585				590					
	Leu	Leu	Gln	Ala	Val	His	Leu	Phe	Ala	Lys	Glu	Leu	Asp	Ala	Lys	Ser
			595					600					605			
5	Val	Thr	Leu	Glu	Arg	Ser	Leu	Gly	Gly	Gly	Arg	Phe	Gly	Glu	Leu	Cys
	610						615					620				
	Cys	Gly	Cys	Leu	Gln	Leu	Pro	Gly	Arg	Gln	Glu	Leu	Leu	Val	Ala	Val
	625					630					635				640	
	His	Met	Leu	Arg	Asp	Ser	Ala	Ser	Asp	Ser	Gln	Arg	Leu	Gly	Phe	Leu
					645				650						655	
10	Ala	Glu	Ala	Leu	Thr	Leu	Gly	Gln	Phe	Asp	His	Ser	His	Ile	Val	Arg
				660					665					670		
	Leu	Glu	Gly	Val	Val	Thr	Arg	Gly	Arg	Ala	Arg	Thr	Leu	Met	Ile	Val
	675						680						685			
	Thr	Glu	Tyr	Met	Ser	His	Gly	Ala	Leu	Asp	Gly	Phe	Leu	Arg	Arg	His
15	690					695					700					
	Glu	Gly	Gln	Leu	Val	Ala	Gly	Gln	Leu	Met	Gly	Leu	Leu	Pro	Gly	Leu
	705				710						715				720	
	Ala	Ser	Ala	Met	Lys	Tyr	Leu	Ser	Glu	Met	Gly	Tyr	Val	His	Arg	Gly
				725					730						735	
20	Leu	Ala	Ala	Arg	His	Val	Leu	Val	Ser	Ser	Asp	Leu	Val	Cys	Lys	Ile
				740					745					750		
	Ser	Gly	Phe	Gly	Arg	Gly	Pro	Arg	Asp	Arg	Ser	Glu	Ala	Val	Tyr	Thr
	755					760						765				
	Thr	Met	Ser	Gly	Arg	Ser	Pro	Ala	Leu	Trp	Ala	Ala	Pro	Glu	Thr	Leu
25	770					775						780				
	Gln	Phe	Gly	His	Phe	Ser	Ser	Ala	Ser	Asp	Val	Trp	Ser	Phe	Gly	Ile
	785				790					795					800	
	Ile	Met	Trp	Glu	Val	Met	Ala	Phe	Gly	Glu	Arg	Pro	Tyr	Trp	Asp	Met
				805					810						815	
30	Ser	Gly	Gln	Asp	Val	Ile	Lys	Ala	Val	Glu	Asp	Gly	Phe	Arg	Leu	Pro
				820					825					830		
	Pro	Pro	Arg	Asn	Cys	Pro	Asn	Leu	His	Arg	Leu	Met	Leu	Asp	Cys	
	835					840						845				
	Trp	Gln	Lys	Asp	Pro	Gly	Glu	Arg	Pro	Arg	Phe	Ser	Gln	Ile	His	Ser
35	850					855					860					
	Ile	Leu	Ser	Lys	Met	Val	Gln	Asp	Pro	Glu	Pro	Pro	Lys	Cys	Ala	Leu
	865				870					875					880	
	Thr	Thr	Cys	Pro	Arg	Pro	Pro	Thr	Pro	Leu	Ala	Asp	Arg	Ala	Phe	Ser
				885					890						895	
40	Thr	Phe	Pro	Ser	Phe	Gly	Ser	Val	Gly	Ala	Trp	Leu	Glu	Ala	Leu	Asp
				900					905					910		
	Leu	Cys	Arg	Tyr	Lys	Asp	Ser	Phe	Ala	Ala	Ala	Gly	Tyr	Gly	Ser	Leu
				915				920					925			
	Glu	Ala	Val	Ala	Glu	Met	Thr	Ala	Gln	Asp	Leu	Val	Ser	Leu	Gly	Ile
45	930					935					940					
	Ser	Leu	Ala	Glu	His	Arg	Glu	Ala	Leu	Leu	Ser	Gly	Ile	Ser	Ala	Leu
	945				950						955				960	
	Gln	Ala	Arg	Val	Leu	Gln	Leu	Gln	Gly	Gln	Gly	Val	Gln	Val		
				965					970							

50

<210> 3

<211> 3099

55

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

60

<222> (1)..(3069)

<223> Extended full length kinase

8/23

<400> 3																	
	atc	ctc	ctg	gat	tcc	aaa	gcc	tcc	cag	gcc	gag	ctg	ggc	tgg	act	gca	48
	Ile	Leu	Leu	Asp	Ser	Lys	Ala	Ser	Gln	Ala	Glu	Leu	Gly	Trp	Thr	Ala	
5	1				5					10					15		
	ctg	cca	agt	aat	ggg	tgg	gag	gag	atc	agc	ggc	gtg	gat	gaa	cac	gac	96
	Leu	Pro	Ser	Asn	Gly	Trp	Glu	Glu	Ile	Ser	Gly	Val	Asp	Glu	His	Asp	
				20					25					30			
10																	
	cgt	ccc	atc	cgc	acg	tac	caa	gtg	tgc	aat	gtg	ctg	gag	ccc	aac	cag	144
	Arg	Pro	Ile	Arg	Thr	Tyr	Gln	Val	Cys	Asn	Val	Leu	Glu	Pro	Asn	Gln	
				35				40					45				
15																	
	gac	aac	tgg	ctg	cag	act	ggc	tgg	ata	agc	cgt	ggc	cgc	ggg	cag	cgc	192
	Asp	Asn	Trp	Leu	Gln	Thr	Gly	Trp	Ile	Ser	Arg	Gly	Arg	Gly	Gln	Arg	
		50					55					60					
20																	
	atc	ttc	gtg	gaa	ctg	cag	ttc	aca	ctc	cgt	gac	tgc	agc	agc	atc	cct	240
	Ile	Phe	Val	Glu	Leu	Gln	Phe	Thr	Leu	Arg	Asp	Cys	Ser	Ser	Ile	Pro	
	65					70					75					80	
25																	
	ggc	gcc	gcg	ggg	acc	tgc	aag	gag	acc	ttc	aac	gtc	tac	tac	ctg	gaa	288
	Gly	Ala	Ala	Gly	Thr	Cys	Lys	Glu	Thr	Phe	Asn	Val	Tyr	Tyr	Leu	Glu	
					85					90					95		
30																	
	act	gag	gcc	gac	ctg	ggc	cgt	ggg	cgt	ccc	cgc	cta	ggc	ggc	agc	cgg	336
	Thr	Glu	Ala	Asp	Leu	Gly	Arg	Gly	Arg	Pro	Arg	Leu	Gly	Gly	Ser	Arg	
				100					105					110			
35																	
	ccc	cgc	aaa	atc	gac	acg	atc	gcg	gcg	gac	gag	agc	ttc	acg	cag	ggc	384
	Pro	Arg	Lys	Ile	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser	Phe	Thr	Gln	Gly	
			115					120					125				
40																	
	gac	ctg	ggg	gag	cgc	aag	atg	aag	ctg	aac	aca	gag	gtg	cgc	gag	atc	432
	Asp	Leu	Gly	Glu	Arg	Lys	Met	Lys	Leu	Asn	Thr	Glu	Val	Arg	Glu	Ile	
		130					135					140					
45																	
	gga	ccg	ctc	agc	cgg	cgg	ggg	ttc	cac	ctg	gcc	ttt	cag	gac	gtg	ggc	480
	Gly	Pro	Leu	Ser	Arg	Arg	Gly	Phe	His	Leu	Ala	Phe	Gln	Asp	Val	Gly	
	145					150					155					160	
50																	
	gca	tgc	gtg	gcg	ctt	gtc	tgc	gtg	cgc	gtc	tac	tac	aag	cag	tgc	cgc	528
	Ala	Cys	Val	Ala	Leu	Val	Ser	Val	Arg	Val	Tyr	Tyr	Lys	Gln	Cys	Arg	
					165					170					175		
55																	
	gcc	acc	gtg	cgg	ggc	ctg	gcc	acg	ttc	cca	gcc	acc	gca	gcc	gag	agc	576
	Ala	Thr	Val	Arg	Gly	Leu	Ala	Thr	Phe	Pro	Ala	Thr	Ala	Ala	Glu	Ser	
				180					185					190			
60																	
	gcc	ttc	tcc	aca	ctg	gtg	gaa	gtg	gcc	gga	acg	tgc	gtg	gcg	cac	tcg	624
	Ala	Phe	Ser	Thr	Leu	Val	Glu	Val	Ala	Gly	Thr	Cys	Val	Ala	His	Ser	
				195				200					205				
65																	
	gaa	ggg	gag	cct	ggc	agc	ccc	cca	cgc	atg	cac	tgc	ggc	gcc	gac	ggc	672
	Glu	Gly	Glu	Pro	Gly	Ser	Pro	Pro	Arg	Met	His	Cys	Gly	Ala	Asp	Gly	
		210					215					220					
70																	
	gag	tgg	ctg	gtg	cct	gtg	ggc	cgc	tgc	agc	tgc	agc	gcg	gga	ttc	cag	720
	Glu	Trp	Leu	Val	Pro	Val	Gly	Arg	Cys	Ser	Cys	Ser	Ala	Gly	Phe	Gln	
	225					230					235					240	

9/23

		gag	cgt	ggt	gac	ttc	tgc	gaa	gcc	tgt	ccc	cca	ggg	ttt	tac	aag	gtg	768
		Glu	Arg	Gly	Asp	Phe	Cys	Glu	Ala	Cys	Pro	Pro	Gly	Phe	Tyr	Lys	Val	
					245						250					255		
5		tcc	ccg	cgg	cgg	ccc	ctc	tgc	tca	ccg	tgc	cca	gag	cac	agc	cgg	gcc	816
		Ser	Pro	Arg	Arg	Pro	Leu	Cys	Ser	Pro	Cys	Pro	Glu	His	Ser	Arg	Ala	
					260					265					270			
10		ctg	gaa	aac	gcc	tcc	acc	ttc	tgc	gtg	tgc	cag	gac	agc	tat	gcg	cgc	864
		Leu	Glu	Asn	Ala	Ser	Thr	Phe	Cys	Val	Cys	Gln	Asp	Ser	Tyr	Ala	Arg	
				275					280					285				
15		tca	ccc	acc	gac	ccg	ccc	tcg	gct	tcc	tgc	acc	cgg	ccg	ccg	tcg	gcg	912
		Ser	Pro	Thr	Asp	Pro	Pro	Ser	Ala	Ser	Cys	Thr	Arg	Pro	Pro	Ser	Ala	
			290					295					300					
20		ccg	cgg	gac	ctg	cag	tac	agc	ctg	agc	cgc	tcg	ccg	ctg	gtg	ctg	cga	960
		Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg	
		305					310					315					320	
25		ctg	cgc	tgg	ctg	ccg	ccg	gcc	gac	tcg	gga	ggc	cgc	tcg	gac	gtc	acc	1008
		Leu	Arg	Trp	Leu	Pro	Pro	Ala	Asp	Ser	Gly	Gly	Arg	Ser	Asp	Val	Thr	
						325					330					335		
30		tac	tcg	ctg	ctg	tgc	ctg	cgc	tgc	ggc	cgc	gag	ggc	ccg	gcg	ggc	gcc	1056
		Tyr	Ser	Leu	Leu	Cys	Leu	Arg	Cys	Gly	Arg	Glu	Gly	Pro	Ala	Gly	Ala	
						340					345				350			
35		tgc	gag	ccg	tgc	ggg	ccg	cgc	gtg	gcc	ttc	cta	ccg	cgc	cag	gca	ggg	1104
		Cys	Glu	Pro	Cys	Gly	Pro	Arg	Val	Ala	Phe	Leu	Pro	Arg	Gln	Ala	Gly	
				355					360					365				
40		ctg	cgg	gag	cga	gcc	gcc	acg	ctg	ctg	cac	ctg	cgg	ccc	ggc	gcg	cgc	1152
		Leu	Arg	Glu	Arg	Ala	Ala	Thr	Leu	Leu	His	Leu	Arg	Pro	Gly	Ala	Arg	
			370					375					380					
45		tac	acc	gtg	cgc	gtg	gcc	gcg	ctc	aac	ggc	gtc	tcg	ggc	ccg	gcg	gcc	1200
		Tyr	Thr	Val	Arg	Val	Ala	Ala	Leu	Asn	Gly	Val	Ser	Gly	Pro	Ala	Ala	
		385					390					395					400	
50		gcc	gcg	gga	acc	acc	tac	gcg	cag	gtc	acc	gtc	tcc	acc	ggg	ccc	ggg	1248
		Ala	Ala	Gly	Thr	Thr	Tyr	Ala	Gln	Val	Thr	Val	Ser	Thr	Gly	Pro	Gly	
						405					410					415		
55		gcg	ccc	tgg	gag	gag	gat	gag	atc	cgc	agg	gac	cga	gtg	gaa	ccc	cag	1296
		Ala	Pro	Trp	Glu	Glu	Asp	Glu	Ile	Arg	Arg	Asp	Arg	Val	Glu	Pro	Gln	
					420					425					430			
60		agc	gtg	tcc	ctg	tcg	tgg	cgg	gag	ccc	atc	cct	gcc	gga	gcc	cct	ggg	1344
		Ser	Val	Ser	Leu	Ser	Trp	Arg	Glu	Pro	Ile	Pro	Ala	Gly	Ala	Pro	Gly	
				435				440						445				
65		gcc	aat	gac	acg	gag	tac	gag	atc	cga	tac	tac	gag	aag	ggg	cag	agt	1392
		Ala	Asn	Asp	Thr	Glu	Tyr	Glu	Ile	Arg	Tyr	Tyr	Glu	Lys	Gly	Gln	Ser	
				450				455						460				
70		gag	cag	act	tac	tcc	atg	gtg	aag	aca	ggg	gcg	ccc	aca	gtc	acc	gtc	1440
		Glu	Gln	Thr	Tyr	Ser	Met	Val	Lys	Thr	Gly	Ala	Pro	Thr	Val	Thr	Val	
		465					470					475					480	

10/23

	acc aac ctg aag ccg gct acc cgc tac gtc ttt cag atc cgg gcc gct	1488
	Thr Asn Leu Lys Pro Ala Thr Arg Tyr Val Phe Gln Ile Arg Ala Ala	
	485 490 495	
5	tcc ccg ggg cca tcc tgg gag gcc cag agt ttt aac ccc agc att gaa	1536
	Ser Pro Gly Pro Ser Trp Glu Ala Gln Ser Phe Asn Pro Ser Ile Glu	
	500 505 510	
10	gta cag acc ctg ggg gag gct gcc tca ggg tcc agg gac cag agc ccc	1584
	Val Gln Thr Leu Gly Glu Ala Ala Ser Gly Ser Arg Asp Gln Ser Pro	
	515 520 525	
15	gcc att gtc gtc acc gta gtg acc atc tcg gcc ctc ctc gtc ctg ggc	1632
	Ala Ile Val Val Thr Val Thr Ile Ser Ala Leu Leu Val Leu Gly	
	530 535 540	
20	tcc gtg atg agt gtg ctg gcc att tgg agg agg ccc tgc agc tat ggc	1680
	Ser Val Met Ser Val Leu Ala Ile Trp Arg Arg Pro Cys Ser Tyr Gly	
	545 550 555 560	
	aaa gga gga ggg gat gcc cat gat gaa gag gag ctg tat ttc cac tgt	1728
	Lys Gly Gly Gly Asp Ala His Asp Glu Glu Glu Leu Tyr Phe His Cys	
	565 570 575	
25	gag ttg gct ggg cct ggc tgg gga ggg gcg ggc tcc acc ctg cct cac	1776
	Glu Leu Ala Gly Pro Gly Trp Gly Gly Ala Gly Ser Thr Leu Pro His	
	580 585 590	
30	tcc tca gcc ctc tgc ctt gct aca ctg cat gag ccc gct tgt cag ccc	1824
	Ser Ser Ala Leu Cys Leu Ala Thr Leu His Glu Pro Ala Cys Gln Pro	
	595 600 605	
35	cag tgg tgg agg ccc cat tcc tca ggc tca tct ctt tct ctc cca gtc	1872
	Gln Trp Trp Arg Pro His Ser Ser Gly Ser Ser Leu Ser Leu Pro Val	
	610 615 620	
40	aaa gtc cca aca cgt cgc aca ttc ctg gac ccc cag agc tgt ggg gac	1920
	Lys Val Pro Thr Arg Arg Thr Phe Leu Asp Pro Gln Ser Cys Gly Asp	
	625 630 635 640	
	ctg ctg cag gct gtg cat ctg ttc gcc aag gaa ctg gat gcg aaa agc	1968
	Leu Leu Gln Ala Val His Leu Phe Ala Lys Glu Leu Asp Ala Lys Ser	
	645 650 655	
45	gtc acg ctg gag agg agc ctt gga gga ggg cgg ttt ggg gag ctg tgc	2016
	Val Thr Leu Glu Arg Ser Leu Gly Gly Gly Arg Phe Gly Glu Leu Cys	
	660 665 670	
50	tgt ggc tgc ttg cag ctc ccc ggt cgc cag gag ctg ctc gta gcc gtg	2064
	Cys Gly Cys Leu Gln Leu Pro Gly Arg Gln Glu Leu Leu Val Ala Val	
	675 680 685	
55	cat atg ctg agg gac agc gcc tcc gac tca cag agg ctc ggc ttc ctg	2112
	His Met Leu Arg Asp Ser Ala Ser Asp Ser Gln Arg Leu Gly Phe Leu	
	690 695 700	
60	gcc gag gcc ctc acg ctg ggc cag ttt gac cat agc cac atc gtg cgg	2160
	Ala Glu Ala Leu Thr Leu Gly Gln Phe Asp His Ser His Ile Val Arg	
	705 710 715 720	
	ctg gag ggc gtt gtt acc cga ggt agg gcc cgc acc ttg atg att gtc	2208

11/23

	Leu	Glu	Gly	Val	Val	Thr	Arg	Gly	Arg	Ala	Arg	Thr	Leu	Met	Ile	Val	
				725						730					735		
5	acc	gag	tac	atg	agc	cat	ggg	gcc	ctg	gac	ggc	ttc	ctc	agg	cgg	cac	2256
	Thr	Glu	Tyr	Met	Ser	His	Gly	Ala	Leu	Asp	Gly	Phe	Leu	Arg	Arg	His	
				740					745					750			
10	gag	ggg	cag	ctg	gtg	gct	ggg	caa	ctg	atg	ggg	ttg	ctg	cct	ggg	ctg	2304
	Glu	Gly	Gln	Leu	Val	Ala	Gly	Gln	Leu	Met	Gly	Leu	Leu	Pro	Gly	Leu	
			755					760					765				
15	gca	tca	gcc	atg	aag	tat	ctg	tca	gag	atg	ggc	tac	gtt	cac	cgg	ggc	2352
	Ala	Ser	Ala	Met	Lys	Tyr	Leu	Ser	Glu	Met	Gly	Tyr	Val	His	Arg	Gly	
			770				775					780					
	ctg	gca	gct	cgc	cat	gtg	ctg	gtc	agc	agc	gac	ctt	gtc	tgc	aag	atc	2400
	Leu	Ala	Ala	Arg	His	Val	Leu	Val	Ser	Ser	Asp	Leu	Val	Cys	Lys	Ile	
						790					795					800	
20	tct	ggc	ttc	ggg	cgg	ggc	ccc	cgg	gac	cga	tca	gag	gct	gtc	tac	acc	2448
	Ser	Gly	Phe	Gly	Arg	Gly	Pro	Arg	Asp	Arg	Ser	Glu	Ala	Val	Tyr	Thr	
					805					810					815		
25	act	atg	agt	ggc	cgg	agc	cca	gcg	cta	tgg	gcc	gct	ccc	gag	aca	ctt	2496
	Thr	Met	Ser	Gly	Arg	Ser	Pro	Ala	Leu	Trp	Ala	Ala	Pro	Glu	Thr	Leu	
				820						825				830			
30	cag	ttt	ggc	cac	ttc	agc	tct	gcc	agt	gac	gtg	tgg	agc	ttc	ggc	atc	2544
	Gln	Phe	Gly	His	Phe	Ser	Ser	Ala	Ser	Asp	Val	Trp	Ser	Phe	Gly	Ile	
			835					840					845				
35	atc	atg	tgg	gag	gtg	atg	gcc	ttt	ggg	gag	cgg	cct	tac	tgg	gac	atg	2592
	Ile	Met	Trp	Glu	Val	Met	Ala	Phe	Gly	Glu	Arg	Pro	Tyr	Trp	Asp	Met	
			850				855					860					
	tct	ggc	caa	gac	gtg	atc	aag	gct	gtg	gag	gat	ggc	ttc	cgg	ctg	cca	2640
	Ser	Gly	Gln	Asp	Val	Ile	Lys	Ala	Val	Glu	Asp	Gly	Phe	Arg	Leu	Pro	
						870					875				880		
40	ccc	ccc	agg	aac	tgt	cct	aac	ctt	ctg	cac	cga	cta	atg	ctc	gac	tgc	2688
	Pro	Pro	Arg	Asn	Cys	Pro	Asn	Leu	Leu	His	Arg	Leu	Met	Leu	Asp	Cys	
					885					890					895		
45	tgg	cag	aag	gac	cca	ggt	gag	cgg	ccc	agg	ttc	tcc	cag	atc	cac	agc	2736
	Trp	Gln	Lys	Asp	Pro	Gly	Glu	Arg	Pro	Arg	Phe	Ser	Gln	Ile	His	Ser	
				900					905					910			
50	atc	ctg	agc	aag	atg	gtg	cag	gac	cca	gag	ccc	ccc	aag	tgt	gcc	ctg	2784
	Ile	Leu	Ser	Lys	Met	Val	Gln	Asp	Pro	Glu	Pro	Pro	Lys	Cys	Ala	Leu	
				915				920					925				
55	act	acc	tgt	ccc	agg	cct	ccc	acc	cca	cta	gcg	gac	cgt	gcc	ttc	tcc	2832
	Thr	Thr	Cys	Pro	Arg	Pro	Pro	Thr	Pro	Leu	Ala	Asp	Arg	Ala	Phe	Ser	
				930			935					940					
	acc	ttc	ccc	tcc	ttt	ggc	tct	gtg	ggc	gcg	tgg	ctg	gag	gcc	ctg	gac	2880
	Thr	Phe	Pro	Ser	Phe	Gly	Ser	Val	Gly	Ala	Trp	Leu	Glu	Ala	Leu	Asp	
						950					955				960		
60	ctg	tgc	cgc	tac	aag	gac	agc	ttc	gcg	gct	gct	ggc	tat	ggg	agc	ctg	2928
	Leu	Cys	Arg	Tyr	Lys	Asp	Ser	Phe	Ala	Ala	Ala	Gly	Tyr	Gly	Ser	Leu	

12/23

	965	970	975	
5	gag gcc gtg gcc gag atg act gcc Glu Ala Val Ala Glu Met Thr Ala	cag gac ctg gtg agc cta ggc atc Gln Asp Leu Val Ser Leu Gly Ile	2976	
	980	985	990	
10	tct ttg gct gaa cat cga gag gcc Ser Leu Ala Glu His Arg Glu Ala	ctc ctc agc ggg atc agc gcc ctg Leu Leu Ser Gly Ile Ser Ala Leu	3024	
	995	1000	1005	
	cag gca cga gtg ctc cag ctg cag ggc Gln Ala Arg Val Leu Gln Leu Gln Gly	cag ggg gtg cag gtg tga Gln Gly Val Gln Val	3069	
	1010	1015	1020	
15	gtggacccca ttcttccaag gcaggactcc		3099	
	<210> 4			
	<211> 1022			
20	<212> PRT			
	<213> Homo sapiens			
	<400> 4			
25	Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala 1 5 10 15			
	Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp 20 25 30			
	Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln 35 40 45			
30	Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg 50 55 60			
	Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro 65 70 75 80			
35	Gly Ala Ala Gly Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu 85 90 95			
	Thr Glu Ala Asp Leu Gly Arg Gly Arg Pro Arg Leu Gly Gly Ser Arg 100 105 110			
	Pro Arg Lys Ile Asp Thr Ile Ala Ala Asp Glu Ser Phe Thr Gln Gly 115 120 125			
40	Asp Leu Gly Glu Arg Lys Met Lys Leu Asn Thr Glu Val Arg Glu Ile 130 135 140			
	Gly Pro Leu Ser Arg Arg Gly Phe His Leu Ala Phe Gln Asp Val Gly 145 150 155 160			
	Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Tyr Lys Gln Cys Arg 165 170 175			
45	Ala Thr Val Arg Gly Leu Ala Thr Phe Pro Ala Thr Ala Ala Glu Ser 180 185 190			
	Ala Phe Ser Thr Leu Val Glu Val Ala Gly Thr Cys Val Ala His Ser 195 200 205			
50	Glu Gly Glu Pro Gly Ser Pro Pro Arg Met His Cys Gly Ala Asp Gly 210 215 220			
	Glu Trp Leu Val Pro Val Gly Arg Cys Ser Cys Ser Ala Gly Phe Gln 225 230 235 240			
55	Glu Arg Gly Asp Phe Cys Glu Ala Cys Pro Pro Gly Phe Tyr Lys Val 245 250 255			
	Ser Pro Arg Arg Pro Leu Cys Ser Pro Cys Pro Glu His Ser Arg Ala 260 265 270			
	Leu Glu Asn Ala Ser Thr Phe Cys Val Cys Gln Asp Ser Tyr Ala Arg 275 280 285			
60	Ser Pro Thr Asp Pro Pro Ser Ala Ser Cys Thr Arg Pro Pro Ser Ala 290 295 300			

13/23

	Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg
	305					310					315					320
	Leu	Arg	Trp	Leu	Pro	Pro	Ala	Asp	Ser	Gly	Gly	Arg	Ser	Asp	Val	Thr
					325					330					335	
5	Tyr	Ser	Leu	Leu	Cys	Leu	Arg	Cys	Gly	Arg	Glu	Gly	Pro	Ala	Gly	Ala
				340					345					350		
	Cys	Glu	Pro	Cys	Gly	Pro	Arg	Val	Ala	Phe	Leu	Pro	Arg	Gln	Ala	Gly
			355					360					365			
	Leu	Arg	Glu	Arg	Ala	Ala	Thr	Leu	Leu	His	Leu	Arg	Pro	Gly	Ala	Arg
10			370				375					380				
	Tyr	Thr	Val	Arg	Val	Ala	Ala	Leu	Asn	Gly	Val	Ser	Gly	Pro	Ala	Ala
	385					390					395					400
	Ala	Ala	Gly	Thr	Thr	Tyr	Ala	Gln	Val	Thr	Val	Ser	Thr	Gly	Pro	Gly
					405					410					415	
15	Ala	Pro	Trp	Glu	Glu	Asp	Glu	Ile	Arg	Arg	Asp	Arg	Val	Glu	Pro	Gln
				420					425					430		
	Ser	Val	Ser	Leu	Ser	Trp	Arg	Glu	Pro	Ile	Pro	Ala	Gly	Ala	Pro	Gly
			435					440					445			
	Ala	Asn	Asp	Thr	Glu	Tyr	Glu	Ile	Arg	Tyr	Tyr	Glu	Lys	Gly	Gln	Ser
20		450					455					460				
	Glu	Gln	Thr	Tyr	Ser	Met	Val	Lys	Thr	Gly	Ala	Pro	Thr	Val	Thr	Val
	465					470					475					480
	Thr	Asn	Leu	Lys	Pro	Ala	Thr	Arg	Tyr	Val	Phe	Gln	Ile	Arg	Ala	Ala
					485					490					495	
25	Ser	Pro	Gly	Pro	Ser	Trp	Glu	Ala	Gln	Ser	Phe	Asn	Pro	Ser	Ile	Glu
				500					505					510		
	Val	Gln	Thr	Leu	Gly	Glu	Ala	Ala	Ser	Gly	Ser	Arg	Asp	Gln	Ser	Pro
			515					520					525			
	Ala	Ile	Val	Val	Thr	Val	Val	Thr	Ile	Ser	Ala	Leu	Leu	Val	Leu	Gly
30		530					535					540				
	Ser	Val	Met	Ser	Val	Leu	Ala	Ile	Trp	Arg	Arg	Pro	Cys	Ser	Tyr	Gly
	545					550					555					560
	Lys	Gly	Gly	Gly	Asp	Ala	His	Asp	Glu	Glu	Glu	Leu	Tyr	Phe	His	Cys
					565					570					575	
35	Glu	Leu	Ala	Gly	Pro	Gly	Trp	Gly	Gly	Ala	Gly	Ser	Thr	Leu	Pro	His
				580					585					590		
	Ser	Ser	Ala	Leu	Cys	Leu	Ala	Thr	Leu	His	Glu	Pro	Ala	Cys	Gln	Pro
			595					600					605			
	Gln	Trp	Trp	Arg	Pro	His	Ser	Ser	Gly	Ser	Ser	Leu	Ser	Leu	Pro	Val
40		610					615						620			
	Lys	Val	Pro	Thr	Arg	Arg	Thr	Phe	Leu	Asp	Pro	Gln	Ser	Cys	Gly	Asp
	625					630					635					640
	Leu	Leu	Gln	Ala	Val	His	Leu	Phe	Ala	Lys	Glu	Leu	Asp	Ala	Lys	Ser
					645					650					655	
45	Val	Thr	Leu	Glu	Arg	Ser	Leu	Gly	Gly	Arg	Phe	Gly	Glu	Leu	Cys	
				660					665				670			
	Cys	Gly	Cys	Leu	Gln	Leu	Pro	Gly	Arg	Gln	Glu	Leu	Leu	Val	Ala	Val
				675				680					685			
	His	Met	Leu	Arg	Asp	Ser	Ala	Ser	Asp	Ser	Gln	Arg	Leu	Gly	Phe	Leu
50		690					695					700				
	Ala	Glu	Ala	Leu	Thr	Leu	Gly	Gln	Phe	Asp	His	Ser	His	Ile	Val	Arg
	705					710					715					720
	Leu	Glu	Gly	Val	Val	Thr	Arg	Gly	Arg	Ala	Arg	Thr	Leu	Met	Ile	Val
					725					730					735	
55	Thr	Glu	Tyr	Met	Ser	His	Gly	Ala	Leu	Asp	Gly	Phe	Leu	Arg	Arg	His
				740					745					750		
	Glu	Gly	Gln	Leu	Val	Ala	Gly	Gln	Leu	Met	Gly	Leu	Leu	Pro	Gly	Leu
				755				760					765			
	Ala	Ser	Ala	Met	Lys	Tyr	Leu	Ser	Glu	Met	Gly	Tyr	Val	His	Arg	Gly
60		770					775					780				
	Leu	Ala	Ala	Arg	His	Val	Leu	Val	Ser	Ser	Asp	Leu	Val	Cys	Lys	Ile

14/23

	785				790				795				800				
	Ser	Gly	Phe	Gly	Arg	Gly	Pro	Arg	Asp	Arg	Ser	Glu	Ala	Val	Tyr	Thr	
					805					810					815		
5	Thr	Met	Ser	Gly	Arg	Ser	Pro	Ala	Leu	Trp	Ala	Ala	Pro	Glu	Thr	Leu	
				820					825					830			
	Gln	Phe	Gly	His	Phe	Ser	Ser	Ala	Ser	Asp	Val	Trp	Ser	Phe	Gly	Ile	
			835					840					845				
	Ile	Met	Trp	Glu	Val	Met	Ala	Phe	Gly	Glu	Arg	Pro	Tyr	Trp	Asp	Met	
		850					855					860					
10	Ser	Gly	Gln	Asp	Val	Ile	Lys	Ala	Val	Glu	Asp	Gly	Phe	Arg	Leu	Pro	
	865					870					875					880	
	Pro	Pro	Arg	Asn	Cys	Pro	Asn	Leu	Leu	His	Arg	Leu	Met	Leu	Asp	Cys	
				885						890					895		
15	Trp	Gln	Lys	Asp	Pro	Gly	Glu	Arg	Pro	Arg	Phe	Ser	Gln	Ile	His	Ser	
				900					905					910			
	Ile	Leu	Ser	Lys	Met	Val	Gln	Asp	Pro	Glu	Pro	Pro	Lys	Cys	Ala	Leu	
		915						920					925				
	Thr	Thr	Cys	Pro	Arg	Pro	Pro	Thr	Pro	Leu	Ala	Asp	Arg	Ala	Phe	Ser	
		930					935						940				
20	Thr	Phe	Pro	Ser	Phe	Gly	Ser	Val	Gly	Ala	Trp	Leu	Glu	Ala	Leu	Asp	
	945					950					955					960	
	Leu	Cys	Arg	Tyr	Lys	Asp	Ser	Phe	Ala	Ala	Ala	Gly	Tyr	Gly	Ser	Leu	
					965					970					975		
25	Glu	Ala	Val	Ala	Glu	Met	Thr	Ala	Gln	Asp	Leu	Val	Ser	Leu	Gly	Ile	
				980					985					990			
	Ser	Leu	Ala	Glu	His	Arg	Glu	Ala	Leu	Leu	Ser	Gly	Ile	Ser	Ala	Leu	
		995						1000					1005				
	Gln	Ala	Arg	Val	Leu	Gln	Leu	Gln	Gly	Gln	Gly	Val	Gln	Val			
	1010						1015					1020					
30																	
	<210> 5																
	<211> 1968																
35	<212> DNA																
	<213> Homo sapiens																
	<220>																
	<221> CDS																
40	<222> (1)..(1944)																
	<223> Truncated kinase																
	<400> 5																
45	atc	ctc	ctg	gat	tcc	aaa	gcc	tcc	cag	gcc	gag	ctg	ggc	tgg	act	gca	48
	Ile	Leu	Leu	Asp	Ser	Lys	Ala	Ser	Gln	Ala	Glu	Leu	Gly	Trp	Thr	Ala	
	1				5					10					15		
	ctg	cca	agt	aat	ggg	tgg	gag	gag	atc	agc	ggc	gtg	gat	gaa	cac	gac	96
50	Leu	Pro	Ser	Asn	Gly	Trp	Glu	Glu	Ile	Ser	Gly	Val	Asp	Glu	His	Asp	
				20					25					30			
	cgt	ccc	atc	cgc	acg	tac	caa	gtg	tgc	aat	gtg	ctg	gag	ccc	aac	cag	144
	Arg	Pro	Ile	Arg	Thr	Tyr	Gln	Val	Cys	Asn	Val	Leu	Glu	Pro	Asn	Gln	
			35					40					45				
55	gac	aac	tgg	ctg	cag	act	ggc	tgg	ata	agc	cgt	ggc	cgc	ggg	cag	cgc	192
	Asp	Asn	Trp	Leu	Gln	Thr	Gly	Trp	Ile	Ser	Arg	Gly	Arg	Gly	Gln	Arg	
		50					55					60					
60	atc	ttc	gtg	gaa	ctg	cag	ttc	aca	ctc	cgt	gac	tgc	agc	agc	atc	cct	240
	Ile	Phe	Val	Glu	Leu	Gln	Phe	Thr	Leu	Arg	Asp	Cys	Ser	Ser	Ile	Pro	

15/23

	65				70				75				80				
5	ggc	gcc	gcg	ggt	acc	tgc	aag	gag	acc	ttc	aac	gtc	tac	tac	ctg	gaa	288
	Gly	Ala	Ala	Gly	Thr	Cys	Lys	Glu	Thr	Phe	Asn	Val	Tyr	Tyr	Leu	Glu	
					85					90					95		
10	act	gag	gcc	gac	ctg	ggc	cgt	ggg	cgt	ccc	cgc	cta	ggc	ggc	agc	cgg	336
	Thr	Glu	Ala	Asp	Leu	Gly	Arg	Gly	Arg	Pro	Arg	Leu	Gly	Gly	Ser	Arg	
				100					105					110			
15	ccc	cgc	aaa	atc	gac	acg	atc	gcg	gcg	gac	gag	agc	ttc	acg	cag	ggc	384
	Pro	Arg	Lys	Ile	Asp	Thr	Ile	Ala	Ala	Asp	Glu	Ser	Phe	Thr	Gln	Gly	
			115					120					125				
20	gac	ctg	ggt	gag	cgc	aag	atg	aag	ctg	aac	aca	gag	gtg	cgc	gag	atc	432
	Asp	Leu	Gly	Glu	Arg	Lys	Met	Lys	Leu	Asn	Thr	Glu	Val	Arg	Glu	Ile	
		130					135					140					
25	gga	ccg	ctc	agc	cgg	cgg	ggt	ttc	cac	ctg	gcc	ttt	cag	gac	gtg	ggc	480
	Gly	Pro	Leu	Ser	Arg	Arg	Gly	Phe	His	Leu	Ala	Phe	Gln	Asp	Val	Gly	
	145					150					155					160	
30	gca	tgc	gtg	gcg	ctt	gtc	tcg	gtg	cgc	gtc	tac	tac	aag	cag	tgc	cgc	528
	Ala	Cys	Val	Ala	Leu	Val	Ser	Val	Arg	Val	Tyr	Tyr	Lys	Gln	Cys	Arg	
					165					170					175		
35	gcc	acc	gtg	cgg	ggc	ctg	gcc	acg	ttc	cca	gcc	acc	gca	gcc	gag	agc	576
	Ala	Thr	Val	Arg	Gly	Leu	Ala	Thr	Phe	Pro	Ala	Thr	Ala	Ala	Glu	Ser	
				180					185					190			
40	gcc	ttc	tcc	aca	ctg	gtg	gaa	gtg	gcc	gga	acg	tgc	gtg	gcg	cac	tcg	624
	Ala	Phe	Ser	Thr	Leu	Val	Glu	Val	Ala	Gly	Thr	Cys	Val	Ala	His	Ser	
			195				200					205					
45	gaa	ggg	gag	cct	ggc	agc	ccc	cca	cgc	atg	cac	tgc	ggc	gcc	gac	ggc	672
	Glu	Gly	Glu	Pro	Gly	Ser	Pro	Pro	Arg	Met	His	Cys	Gly	Ala	Asp	Gly	
		210					215					220					
50	gag	tgg	ctg	gtg	cct	gtg	ggc	cgc	tgc	agc	tgc	agc	gcg	gga	ttc	cag	720
	Glu	Trp	Leu	Val	Pro	Val	Gly	Arg	Cys	Ser	Cys	Ser	Ala	Gly	Phe	Gln	
	225					230					235					240	
55	gag	cgt	ggt	gac	ttc	tgc	gaa	gcc	tgt	ccc	cca	ggg	ttt	tac	aag	gtg	768
	Glu	Arg	Gly	Asp	Phe	Cys	Glu	Ala	Cys	Pro	Pro	Gly	Phe	Tyr	Lys	Val	
				245						250					255		
60	tcc	ccg	cgg	cgg	ccc	ctc	tgc	tca	ccg	tgc	cca	gag	cac	agc	cgg	gcc	816
	Ser	Pro	Arg	Arg	Pro	Leu	Cys	Ser	Pro	Cys	Pro	Glu	His	Ser	Arg	Ala	
				260					265					270			
65	ctg	gaa	aac	gcc	tcc	acc	ttc	tgc	gtg	tgc	cag	gac	agc	tat	gcg	cgc	864
	Leu	Glu	Asn	Ala	Ser	Thr	Phe	Cys	Val	Cys	Gln	Asp	Ser	Tyr	Ala	Arg	
			275					280					285				
70	tca	ccc	acc	gac	ccg	ccc	tcg	gct	tcc	tgc	acc	cgg	ccg	ccg	tcg	gcg	912
	Ser	Pro	Thr	Asp	Pro	Pro	Ser	Ala	Ser	Cys	Thr	Arg	Pro	Pro	Ser	Ala	
		290					295					300					
75	ccg	cgg	gac	ctg	cag	tac	agc	ctg	agc	cgc	tcg	ccg	ctg	gtg	ctg	cga	960
	Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg	
	305					310					315					320	

16/23

	ctg cgc tgg ctg ccg ccg gcc gac tcg gga ggc cgc tcg gac gtc acc	1008
	Leu Arg Trp Leu Pro Pro Ala Asp Ser Gly Gly Arg Ser Asp Val Thr	
	325 330 335	
5	tac tcg ctg ctg tgc ctg cgc tgc ggc cgc gag ggc ccg gcg ggc gcc	1056
	Tyr Ser Leu Leu Cys Leu Arg Cys Gly Arg Glu Gly Pro Ala Gly Ala	
	340 345 350	
10	tgc gag ccg tgc ggg ccg cgc gtg gcc ttc cta ccg cgc cag gca ggg	1104
	Cys Glu Pro Cys Gly Pro Arg Val Ala Phe Leu Pro Arg Gln Ala Gly	
	355 360 365	
15	ctg cgg gag cga gcc gcc acg ctg ctg cac ctg cgg ccc ggc gcg cgc	1152
	Leu Arg Glu Arg Ala Ala Thr Leu Leu His Leu Arg Pro Gly Ala Arg	
	370 375 380	
20	tac acc gtg cgc gtg gcc gcg ctc aac ggc gtc tcg ggc ccg gcg gcc	1200
	Tyr Thr Val Arg Val Ala Ala Leu Asn Gly Val Ser Gly Pro Ala Ala	
	385 390 395 400	
25	gcc gcg gga acc acc tac gcg cag gtc acc gtc tcc acc ggg ccc ggg	1248
	Ala Ala Gly Thr Thr Tyr Ala Gln Val Thr Val Ser Thr Gly Pro Gly	
	405 410 415	
	gcg ccc tgg gag gag gat gag atc cgc agg gac cga gtg gaa ccc cag	1296
	Ala Pro Trp Glu Glu Asp Glu Ile Arg Arg Asp Arg Val Glu Pro Gln	
	420 425 430	
30	agc gtg tcc ctg tcg tgg ccg gag ccc atc cct gcc gga gcc cct ggg	1344
	Ser Val Ser Leu Ser Trp Arg Glu Pro Ile Pro Ala Gly Ala Pro Gly	
	435 440 445	
35	gcc aat gac acg gag tac gag atc cga tac tac gag aag ggt cag agt	1392
	Ala Asn Asp Thr Glu Tyr Glu Ile Arg Tyr Tyr Glu Lys Gly Gln Ser	
	450 455 460	
40	gag cag act tac tcc atg gtg aag aca ggg gcg ccc aca gtc acc gtc	1440
	Glu Gln Thr Tyr Ser Met Val Lys Thr Gly Ala Pro Thr Val Thr Val	
	465 470 475 480	
45	acc aac ctg aag ccg gct acc cgc tac gtc ttt cag atc ccg gcc gct	1488
	Thr Asn Leu Lys Pro Ala Thr Arg Tyr Val Phe Gln Ile Arg Ala Ala	
	485 490 495	
	tcc ccg ggg cca tcc tgg gag gcc cag agt ttt aac ccc agc att gaa	1536
	Ser Pro Gly Pro Ser Trp Glu Ala Gln Ser Phe Asn Pro Ser Ile Glu	
	500 505 510	
50	gta cag acc ctg ggg gag gct gcc tca ggg tcc agg gac cag agc ccc	1584
	Val Gln Thr Leu Gly Glu Ala Ala Ser Gly Ser Arg Asp Gln Ser Pro	
	515 520 525	
55	gcc att gtc gtc acc gta gtg acc atc tcg gcc ctc ctc gtc ctg ggc	1632
	Ala Ile Val Val Thr Val Val Thr Ile Ser Ala Leu Leu Val Leu Gly	
	530 535 540	
60	toc gtg atg agt gtg ctg gcc att tgg agg agg ccc tgc agc tat ggc	1680
	Ser Val Met Ser Val Leu Ala Ile Trp Arg Arg Pro Cys Ser Tyr Gly	
	545 550 555 560	

17/23

	aaa gga gga ggg gat gcc cat gat gaa gag gag ctg tat ttc cac ttc	1728								
	Lys Gly Gly Gly Asp Ala His Asp Glu Glu Glu Leu Tyr Phe His Phe									
	565 570 575									
5	aaa gtc cca aca cgt cgc aca ttc ctg gac ccc cag agc tgt ggg gac	1776								
	Lys Val Pro Thr Arg Arg Thr Phe Leu Asp Pro Gln Ser Cys Gly Asp									
	580 585 590									
10	ctg ctg cag gct gtg cat ctg ttc gcc aag gaa ctg gat gcg aaa agc	1824								
	Leu Leu Gln Ala Val His Leu Phe Ala Lys Glu Leu Asp Ala Lys Ser									
	595 600 605									
	gtc acg ctg gag agg agc ctt gga gga ggc aag ctg ggc gcc cag gaa	1872								
15	Val Thr Leu Glu Arg Ser Leu Gly Gly Gly Lys Leu Gly Ala Gln Glu									
	610 615 620									
	gcc ttg tcc cca tct gga agc ctc acc cac tct ata ggc ccc gcc cct	1920								
20	Ala Leu Ser Pro Ser Gly Ser Leu Thr His Ser Ile Gly Pro Ala Pro									
	625 630 635 640									
	act ctg tcc aca cct cta tcc tga agccacacctc caccaggtc tctg	1968								
	Thr Leu Ser Thr Pro Leu Ser									
	645									
25	<210> 6									
	<211> 647									
	<212> PRT									
	<213> Homo sapiens									
30	<400> 6									
	Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala									
	1 5 10 15									
35	Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp									
	20 25 30									
	Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln									
	35 40 45									
	Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg									
	50 55 60									
40	Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro									
	65 70 75 80									
	Gly Ala Ala Gly Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu									
	85 90 95									
45	Thr Glu Ala Asp Leu Gly Arg Gly Arg Pro Arg Leu Gly Gly Ser Arg									
	100 105 110									
	Pro Arg Lys Ile Asp Thr Ile Ala Ala Asp Glu Ser Phe Thr Gln Gly									
	115 120 125									
	Asp Leu Gly Glu Arg Lys Met Lys Leu Asn Thr Glu Val Arg Glu Ile									
	130 135 140									
50	Gly Pro Leu Ser Arg Arg Gly Phe His Leu Ala Phe Gln Asp Val Gly									
	145 150 155 160									
	Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Tyr Lys Gln Cys Arg									
	165 170 175									
55	Ala Thr Val Arg Gly Leu Ala Thr Phe Pro Ala Thr Ala Ala Glu Ser									
	180 185 190									
	Ala Phe Ser Thr Leu Val Glu Val Ala Gly Thr Cys Val Ala His Ser									
	195 200 205									
	Glu Gly Glu Pro Gly Ser Pro Pro Arg Met His Cys Gly Ala Asp Gly									
	210 215 220									
60	Glu Trp Leu Val Pro Val Gly Arg Cys Ser Cys Ser Ala Gly Phe Gln									
	225 230 235 240									

18/23

Glu Arg Gly Asp Phe Cys Glu Ala Cys Pro Pro Gly Phe Tyr Lys Val
 245 250 255
 Ser Pro Arg Arg Pro Leu Cys Ser Pro Cys Pro Glu His Ser Arg Ala
 260 265 270
 5 Leu Glu Asn Ala Ser Thr Phe Cys Val Cys Gln Asp Ser Tyr Ala Arg
 275 280 285
 Ser Pro Thr Asp Pro Pro Ser Ala Ser Cys Thr Arg Pro Pro Ser Ala
 290 295 300
 10 Pro Arg Asp Leu Gln Tyr Ser Leu Ser Arg Ser Pro Leu Val Leu Arg
 305 310 315 320
 Leu Arg Trp Leu Pro Pro Ala Asp Ser Gly Gly Arg Ser Asp Val Thr
 325 330 335
 Tyr Ser Leu Leu Cys Leu Arg Cys Gly Arg Glu Gly Pro Ala Gly Ala
 340 345 350
 15 Cys Glu Pro Cys Gly Pro Arg Val Ala Phe Leu Pro Arg Gln Ala Gly
 355 360 365
 Leu Arg Glu Arg Ala Ala Thr Leu Leu His Leu Arg Pro Gly Ala Arg
 370 375 380
 20 Tyr Thr Val Arg Val Ala Ala Leu Asn Gly Val Ser Gly Pro Ala Ala
 385 390 395 400
 Ala Ala Gly Thr Thr Tyr Ala Gln Val Thr Val Ser Thr Gly Pro Gly
 405 410 415
 Ala Pro Trp Glu Glu Asp Glu Ile Arg Arg Asp Arg Val Glu Pro Gln
 420 425 430
 25 Ser Val Ser Leu Ser Trp Arg Glu Pro Ile Pro Ala Gly Ala Pro Gly
 435 440 445
 Ala Asn Asp Thr Glu Tyr Glu Ile Arg Tyr Tyr Glu Lys Gly Gln Ser
 450 455 460
 30 Glu Gln Thr Tyr Ser Met Val Lys Thr Gly Ala Pro Thr Val Thr Val
 465 470 475 480
 Thr Asn Leu Lys Pro Ala Thr Arg Tyr Val Phe Gln Ile Arg Ala Ala
 485 490 495
 Ser Pro Gly Pro Ser Trp Glu Ala Gln Ser Phe Asn Pro Ser Ile Glu
 500 505 510
 35 Val Gln Thr Leu Gly Glu Ala Ala Ser Gly Ser Arg Asp Gln Ser Pro
 515 520 525
 Ala Ile Val Val Thr Val Val Thr Ile Ser Ala Leu Leu Val Leu Gly
 530 535 540
 40 Ser Val Met Ser Val Leu Ala Ile Trp Arg Arg Pro Cys Ser Tyr Gly
 545 550 555 560
 Lys Gly Gly Gly Asp Ala His Asp Glu Glu Glu Leu Tyr Phe His Phe
 565 570 575
 Lys Val Pro Thr Arg Arg Thr Phe Leu Asp Pro Gln Ser Cys Gly Asp
 580 585 590
 45 Leu Leu Gln Ala Val His Leu Phe Ala Lys Glu Leu Asp Ala Lys Ser
 595 600 605
 Val Thr Leu Glu Arg Ser Leu Gly Gly Gly Lys Leu Gly Ala Gln Glu
 610 615 620
 50 Ala Leu Ser Pro Ser Gly Ser Leu Thr His Ser Ile Gly Pro Ala Pro
 625 630 635 640
 Thr Leu Ser Thr Pro Leu Ser
 645

55

<210> 7

<211> 2112

<212> DNA

<213> Homo sapiens

60

<220>

19/23

<221> CDS

<222> (1)..(2088)

<223> Extended truncated kinase

```

5  <400> 7
   atc ctc ctg gat tcc aaa gcc tcc cag gcc gag ctg ggc tgg act gca 48
   Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala
     1             5             10             15

10  ctg cca agt aat ggg tgg gag gag atc agc ggc gtg gat gaa cac gac 96
   Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp
           20             25             30

15  cgt ccc atc cgc acg tac caa gtg tgc aat gtg ctg gag ccc aac cag 144
   Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln
           35             40             45

20  gac aac tgg ctg cag act ggc tgg ata agc cgt ggc cgc ggg cag cgc 192
   Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg
           50             55             60

25  atc ttc gtg gaa ctg cag ttc aca ctc cgt gac tgc agc agc atc cct 240
   Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro
           65             70             75             80

30  ggc gcc gcg ggt acc tgc aag gag acc ttc aac gtc tac tac ctg gaa 288
   Gly Ala Ala Gly Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu
           85             90             95

35  act gag gcc gac ctg ggc cgt ggg cgt ccc cgc cta ggc ggc agc cgg 336
   Thr Glu Ala Asp Leu Gly Arg Gly Arg Pro Arg Leu Gly Gly Ser Arg
           100            105            110

40  ccc cgc aaa atc gac acg atc gcg gcg gac gag agc ttc acg cag ggc 384
   Pro Arg Lys Ile Asp Thr Ile Ala Ala Asp Glu Ser Phe Thr Gln Gly
           115            120            125

45  gac ctg ggt gag cgc aag atg aag ctg aac aca gag gtg cgc gag atc 432
   Asp Leu Gly Glu Arg Lys Met Lys Leu Asn Thr Glu Val Arg Glu Ile
           130            135            140

50  gga ccg ctc agc cgg cgg ggt ttc cac ctg gcc ttt cag gac gtg ggc 480
   Gly Pro Leu Ser Arg Arg Gly Phe His Leu Ala Phe Gln Asp Val Gly
           145            150            155            160

55  gca tgc gtg gcg ctt gtc tcg gtg cgc gtc tac tac aag cag tgc cgc 528
   Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Tyr Lys Gln Cys Arg
           165            170            175

60  gcc acc gtg cgg ggc ctg gcc acg ttc cca gcc acc gca gcc gag agc 576
   Ala Thr Val Arg Gly Leu Ala Thr Phe Pro Ala Thr Ala Ala Glu Ser
           180            185            190

65  gcc ttc tcc aca ctg gtg gaa gtg gcc gga acg tgc gtg gcg cac tcg 624
   Ala Phe Ser Thr Leu Val Glu Val Ala Gly Thr Cys Val Ala His Ser
           195            200            205

70  gaa ggg gag cct ggc agc ccc cca cgc atg cac tgc ggc gcc gac ggc 672
   Glu Gly Glu Pro Gly Ser Pro Pro Arg Met His Cys Gly Ala Asp Gly
           210            215            220

```

20/23

		gag	tgg	ctg	gtg	cct	gtg	ggc	cgc	tgc	agc	tgc	agc	gcg	gga	ttc	cag	720
		Glu	Trp	Leu	Val	Pro	Val	Gly	Arg	Cys	Ser	Cys	Ser	Ala	Gly	Phe	Gln	
		225					230					235					240	
5		gag	cgt	ggt	gac	ttc	tgc	gaa	gcc	tgt	ccc	cca	ggg	ttt	tac	aag	gtg	768
		Glu	Arg	Gly	Asp	Phe	Cys	Glu	Ala	Cys	Pro	Pro	Gly	Phe	Tyr	Lys	Val	
					245					250						255		
		tcc	ccg	cgg	cgg	ccc	ctc	tgc	tca	ccg	tgc	cca	gag	cac	agc	cgg	gcc	816
10		Ser	Pro	Arg	Arg	Pro	Leu	Cys	Ser	Pro	Cys	Pro	Glu	His	Ser	Arg	Ala	
					260					265					270			
		ctg	gaa	aac	gcc	tcc	acc	ttc	tgc	gtg	tgc	cag	gac	agc	tat	gcg	cgc	864
15		Leu	Glu	Asn	Ala	Ser	Thr	Phe	Cys	Val	Cys	Gln	Asp	Ser	Tyr	Ala	Arg	
				275					280					285				
		tca	ccc	acc	gac	ccg	ccc	tcg	gct	tcc	tgc	acc	cgg	ccg	ccg	tcg	gcg	912
		Ser	Pro	Thr	Asp	Pro	Pro	Ser	Ala	Ser	Cys	Thr	Arg	Pro	Pro	Ser	Ala	
		290						295					300					
20		ccg	cgg	gac	ctg	cag	tac	agc	ctg	agc	cgc	tcg	ccg	ctg	gtg	ctg	cga	960
		Pro	Arg	Asp	Leu	Gln	Tyr	Ser	Leu	Ser	Arg	Ser	Pro	Leu	Val	Leu	Arg	
		305				310						315					320	
		ctg	cgc	tgg	ctg	ccg	ccg	gcc	gac	tcg	gga	ggc	cgc	tcg	gac	gtc	acc	1008
25		Leu	Arg	Trp	Leu	Pro	Pro	Ala	Asp	Ser	Gly	Gly	Arg	Ser	Asp	Val	Thr	
					325					330					335			
		tac	tcg	ctg	ctg	tgc	ctg	cgc	tgc	ggc	cgc	gag	ggc	ccg	gcg	ggc	gcc	1056
30		Tyr	Ser	Leu	Leu	Cys	Leu	Arg	Cys	Gly	Arg	Glu	Gly	Pro	Ala	Gly	Ala	
				340					345					350				
		tgc	gag	ccg	tgc	ggg	ccg	cgc	gtg	gcc	ttc	cta	ccg	cgc	cag	gca	ggg	1104
35		Cys	Glu	Pro	Cys	Gly	Pro	Arg	Val	Ala	Phe	Leu	Pro	Arg	Gln	Ala	Gly	
		355						360					365					
		ctg	cgg	gag	cga	gcc	gcc	acg	ctg	ctg	cac	ctg	cgg	ccc	ggc	gcg	cgc	1152
		Leu	Arg	Glu	Arg	Ala	Ala	Thr	Leu	Leu	His	Leu	Arg	Pro	Gly	Ala	Arg	
		370						375					380					
40		tac	acc	gtg	cgc	gtg	gcc	gcg	ctc	aac	ggc	gtc	tcg	ggc	ccg	gcg	gcc	1200
		Tyr	Thr	Val	Arg	Val	Ala	Ala	Leu	Asn	Gly	Val	Ser	Gly	Pro	Ala	Ala	
		385				390						395					400	
		gcc	gcg	gga	acc	acc	tac	gcg	cag	gtc	acc	gtc	tcc	acc	ggg	ccc	ggg	1248
45		Ala	Ala	Gly	Thr	Thr	Tyr	Ala	Gln	Val	Thr	Val	Ser	Thr	Gly	Pro	Gly	
					405					410					415			
		gcg	ccc	tgg	gag	gag	gat	gag	atc	cgc	agg	gac	cga	gtg	gaa	ccc	cag	1296
50		Ala	Pro	Trp	Glu	Glu	Asp	Glu	Ile	Arg	Arg	Asp	Arg	Val	Glu	Pro	Gln	
				420					425					430				
		agc	gtg	tcc	ctg	tcg	tgg	cgg	gag	ccc	atc	cct	gcc	gga	gcc	cct	ggg	1344
55		Ser	Val	Ser	Leu	Ser	Trp	Arg	Glu	Pro	Ile	Pro	Ala	Gly	Ala	Pro	Gly	
			435					440					445					
		gcc	aat	gac	acg	gag	tac	gag	atc	cga	tac	tac	gag	aag	ggg	cag	agt	1392
		Ala	Asn	Asp	Thr	Glu	Tyr	Glu	Ile	Arg	Tyr	Tyr	Glu	Lys	Gly	Gln	Ser	
		450						455					460					
60		gag	cag	act	tac	tcc	atg	gtg	aag	aca	ggg	gcg	ccc	aca	gtc	acc	gtc	1440

[illegible]

22/23

<211> 695

<212> PRT

<213> Homo sapiens

5 <400> 8
 Ile Leu Leu Asp Ser Lys Ala Ser Gln Ala Glu Leu Gly Trp Thr Ala
 1 5 10 15
 Leu Pro Ser Asn Gly Trp Glu Glu Ile Ser Gly Val Asp Glu His Asp
 20 25 30
 10 Arg Pro Ile Arg Thr Tyr Gln Val Cys Asn Val Leu Glu Pro Asn Gln
 35 40 45
 Asp Asn Trp Leu Gln Thr Gly Trp Ile Ser Arg Gly Arg Gly Gln Arg
 50 55 60
 Ile Phe Val Glu Leu Gln Phe Thr Leu Arg Asp Cys Ser Ser Ile Pro
 15 65 70 75 80
 Gly Ala Ala Gly Thr Cys Lys Glu Thr Phe Asn Val Tyr Tyr Leu Glu
 85 90 95
 Thr Glu Ala Asp Leu Gly Arg Gly Arg Pro Arg Leu Gly Gly Ser Arg
 100 105 110
 20 Pro Arg Lys Ile Asp Thr Ile Ala Ala Asp Glu Ser Phe Thr Gln Gly
 115 120 125
 Asp Leu Gly Glu Arg Lys Met Lys Leu Asn Thr Glu Val Arg Glu Ile
 130 135 140
 Gly Pro Leu Ser Arg Arg Gly Phe His Leu Ala Phe Gln Asp Val Gly
 25 145 150 155 160
 Ala Cys Val Ala Leu Val Ser Val Arg Val Tyr Tyr Lys Gln Cys Arg
 165 170 175
 Ala Thr Val Arg Gly Leu Ala Thr Phe Pro Ala Thr Ala Ala Glu Ser
 180 185 190
 30 Ala Phe Ser Thr Leu Val Glu Val Ala Gly Thr Cys Val Ala His Ser
 195 200 205
 Glu Gly Glu Pro Gly Ser Pro Pro Arg Met His Cys Gly Ala Asp Gly
 210 215 220
 Glu Trp Leu Val Pro Val Gly Arg Cys Ser Cys Ser Ala Gly Phe Gln
 35 225 230 235 240
 Glu Arg Gly Asp Phe Cys Glu Ala Cys Pro Pro Gly Phe Tyr Lys Val
 245 250 255
 Ser Pro Arg Arg Pro Leu Cys Ser Pro Cys Pro Glu His Ser Arg Ala
 260 265 270
 40 Leu Glu Asn Ala Ser Thr Phe Cys Val Cys Gln Asp Ser Tyr Ala Arg
 275 280 285
 Ser Pro Thr Asp Pro Pro Ser Ala Ser Cys Thr Arg Pro Pro Ser Ala
 290 295 300
 Pro Arg Asp Leu Gln Tyr Ser Leu Ser Arg Ser Pro Leu Val Leu Arg
 45 305 310 315 320
 Leu Arg Trp Leu Pro Pro Ala Asp Ser Gly Arg Ser Asp Val Thr
 325 330 335
 Tyr Ser Leu Leu Cys Leu Arg Cys Gly Arg Glu Gly Pro Ala Gly Ala
 340 345 350
 50 Cys Glu Pro Cys Gly Pro Arg Val Ala Phe Leu Pro Arg Gln Ala Gly
 355 360 365
 Leu Arg Glu Arg Ala Ala Thr Leu Leu His Leu Arg Pro Gly Ala Arg
 370 375 380
 Tyr Thr Val Arg Val Ala Ala Leu Asn Gly Val Ser Gly Pro Ala Ala
 55 385 390 395 400
 Ala Ala Gly Thr Thr Tyr Ala Gln Val Thr Val Ser Thr Gly Pro Gly
 405 410 415
 Ala Pro Trp Glu Glu Asp Glu Ile Arg Arg Asp Arg Val Glu Pro Gln
 420 425 430
 60 Ser Val Ser Leu Ser Trp Arg Glu Pro Ile Pro Ala Gly Ala Pro Gly
 435 440 445

