

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 October 2002 (03.10.2002)

PCT

(10) International Publication Number
WO 02/076510 A1

(51) International Patent Classification⁷: **A61K 51/00**,
A61M 36/14

(74) Agent: **SHERWOOD, Pamela, J.**; BOZICEVIC, FIELD
& FRANCIS, L.L.P, 200 Middlefield Road, Suite 200,
Menlo Park, CA 94025 (US).

(21) International Application Number: PCT/US02/08992

(22) International Filing Date: 22 March 2002 (22.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/816,703 23 March 2001 (23.03.2001) US
09/983,000 17 October 2001 (17.10.2001) US

(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, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, 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).

(71) Applicant (*for all designated States except US*): **AGY
THERAPEUTICS** [US/US]; 290 Utah Avenue, South
San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **MUELLER, Sabine**
[DE/US]; 343 Clipper Street, San Francisco, CA 94114
(US). **MELCHER, Thorsten** [DE/US]; 1976 10th Av-
enue, San Francisco, CA 94116 (US). **CHIN, Daniel, J.**
[US/US]; 653 Celestial Lane, Foster City, CA 94404 (US).

Published:

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: USE OF BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALIZATION OF BRAIN TUMORS

(57) Abstract: The present invention relates to the use of proteins that are differentially expressed in primary brain tumor tissues, as compared to normal brain tissues, as biomolecular targets for brain tumor treatment therapies. Specifically, the present invention relates to the use of immunotherapeutic and immunoimaging agents which specifically bind to one or more of the identified brain tumor protein targets. The present invention also provides compounds and pharmaceutically acceptable compositions for administration in the methods of the invention. The present invention also provides novel splice variants of protein PTP, PTP SM1 and PTP SM2. Nucleic acid probes specific for the spliced mRNA encoding these variants and affinity reagents specific for the novel proteins are also provided.



WO 02/076510 A1

USE OF BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALIZATION OF BRAIN TUMORS

BACKGROUND OF THE INVENTION

5 Among tumors, those of the brain are considered to have one of the least favorable prognoses for long term survival: the average life expectancy of an individual diagnosed with a central nervous system (CNS) tumor is just eight to twelve months. Several unique characteristics of both the brain and its particular types of neoplastic cells create daunting challenges for the complete treatment and management of brain tumors. Among these are the physical characteristics of the intracranial space;
10 the relative biological isolation of the brain from the rest of the body; the relatively essential and irreplaceable nature of the organ mass; and the unique nature of brain tumor cells.

 The intracranial space and physical layout of the brain create significant obstacles to treatment and recovery. The brain is primarily comprised of astrocytes, which make up the majority of the brain mass, and serve as a scaffold and support for the neurons, neurons, which carry the actual
15 electrical impulses of the nervous system, and a minor contingent of other cells, such as insulating oligodendrocytes that produce myelin. These cell types give rise to primary brain tumors, including astrocytomas, neuroblastomas, glioblastomas, oligodendrogliomas, and the like.

 The brain is encased in the rigid shell of the skull, and is cushioned by the cerebrospinal fluid. Because of the relatively small volume of the skull cavity, minor changes in the volume of
20 tissue in the brain can dramatically increase intracranial pressure, causing damage to the entire organ. Thus, even small tumors can have a profound and adverse affect on the brain's function. The cramped physical location of the cranium also makes surgery and treatment of the brain a difficult and delicate procedure. However, because of the dangers of increased intracranial pressure from the tumor, surgery is often the first strategy of attack in treating brain tumors.

25 In addition to its physical isolation, the brain is chemically and biologically isolated from the rest of the body by the "Blood-Brain-Barrier" (or BBB). This physiological phenomenon is due to the "tightness" of the epithelial cell junctions in the lining of the blood vessels in the brain. Nutrients, which are actively transported across the cell lining, can reach the brain, but other molecules from the bloodstream are excluded. This prevents toxins, viruses, and other potentially
30 dangerous molecules from entering the brain cavity. However, it also prevents therapeutic molecules, including many chemotherapeutic agents that are useful in other types of tumors, from crossing into the brain. Thus, many therapies directed at the brain must be delivered directly into the brain cavity, *e.g.* by an Ommaya reservoir, or administered in elevated dosages to ensure the diffusion of an effective amount across the BBB.

35 With the difficulties of administering chemotherapies to the brain, radiotherapy approaches have also been attempted. However, the amount of radiation necessary to completely destroy potential tumor-producing cells also produce unacceptable losses of healthy brain tissue. The

retention of patient cognitive function while eliminating the tumor mass is another challenge to brain tumor treatment. Neoplastic brain cells are often pervasive, and travel throughout the entire brain mass. Thus, it is impossible to define a true "tumor margin," unlike, for example, in lung or bladder cancers. Unlike reproductive (ovarian, uterine, testicular, prostate, etc.), breast, kidney, or lung
5 cancers, the entire organ, or even significant portions, cannot be removed to prevent the growth of new tumors. In addition, brain tumors are very heterogeneous, with different cell doubling times, treatment resistances, and other biochemical idiosyncrasies between the various cell populations that make up the tumor. This pervasive and variable nature greatly adds to the difficulty of treating brain tumors while preserving the health and function of normal brain tissue.

10 Although current surgical methods offer considerably better post-operative life for patients, current combination therapy methods (surgery, low-dosage radiation, and chemotherapy) have only improved the life expectancy of patients by one month, as compared to the methods of 30 years ago. Without effective agents to prevent the growth of brain tumor cells that are present outside the main tumor mass, the prognosis for these patients cannot be significantly improved. Although some
15 immuno-affinity agents have been proposed and tested for the treatment of brain tumors, see, for example, the tenascin-targeting agents described in U.S. Patent No. 5,624,659, these agents have not proven sufficient for the treatment of brain tumors. Thus, therapeutic agents which are directed towards new molecular targets, and are capable of specifically targeting and killing brain tumor cells, are urgently needed for the treatment of brain tumors.

20 *Relevant literature*

Analysis of differential gene expression in glioblastoma may be found in, for example, Mariani *et al.* (2001) *J Neurooncol* **53**(2):161-76; Markert *et al.* (2001) *Physiol Genomics* **5**(1):21-33; Yano *et al.* (2000) *Neurol Res* **22**(7):650-6; Kroes *et al.* (2000) *Cancer Lett* **156**(2):191-8; and
25 Reis *et al.* (2000) *Am J Pathol* **156**(2):425-32, among others.

SUMMARY OF THE INVENTION

The present invention provides methods and reagents for specifically targeting brain tumor neoplastic cells for both therapeutic and imaging purposes, by targeting brain tumor protein targets
30 (T_{BT}). These targets have been identified as being overexpressed in brain tumors, and thus allow for the selective inhibition of cell function or selective marking for visualization with therapeutic or visualizing compositions which have a specific affinity for these protein targets. The invention also provides methods for the identification of compounds that modulate the expression of genes or the activity of gene products involved in such tumors, as well as methods for the treatment of disease by
35 administering such compounds to individuals suffering from such tumors. Also included in the invention are two novel isoforms of PTP ζ , SM1 and SM2.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1: A diagram of the two newly discovered splicing variant isoforms of PTP ζ . The approximate position of the domains of the isoforms is indicated underneath the isoforms, as well as the approximate exon size (for size reference, exon 12 is 3.6 kilobases.) SM 1 fails to splice correctly after the 9th exon, yielding a mRNA with two extra codons followed by a stop codon after the normal terminus of exon 9. SM 2 contains a 116 nucleotide insertion from between exons 23 & 24.

DETAILED DESCRIPTION OF THE EMBODIMENTS

Brain tumor protein targets and genes that are differentially expressed between brain tumor tissue and normal brain tissue are provided herein. Differential cloning between cancerous and normal brains has identified brain tumor protein target genes by DNA sequence analysis. Genes and their protein products that are up-regulated in glioblastoma are important because they provide a specific marker for neoplastic cells, and are expected to mediate the initiation and progression of brain tumors. Inhibition of the gene and/or protein activity can be advantageous in treating brain tumors, *e.g.* glioblastoma multiforme; ependymoma; glioma; astrocytoma; medulloblastoma; neuroglioma; oligodendroglioma; meningioma, *etc.* The overexpressed brain tumor protein targets provide excellent targets for immunotherapeutic agents that either deliver cytotoxic agents to directly promote tumor cell death, or that alter the function of the brain tumor protein targets to inhibit the normal physiology of the tumor cell. In addition, immunoimaging agents targeted to the brain tumor protein targets can be utilized to visualize the tumor mass in diagnostic methods, *e.g.* magnetic resonance imaging (MRI), radiography, *etc.* and/or in surgery, *e.g.* by the use of optically visible dye moieties in an immunoimaging agent, *etc.*

Therapeutic and prophylactic treatment methods for individuals suffering from, or at risk of a brain tumor, involve administering either a therapeutic or prophylactic amount of an agent that modulates the activity of T_{BT} protein or gene or specifically binds to a T_{BT} protein. For example, a chemotherapeutic agent can be coupled to a T_{BT} specific binding moiety.

Screening methods may involve conducting various types of assays to identify agents that modulate the expression or activity of a T_{BT} gene or protein, or may involve screening for specific binding activity to a T_{BT} gene or protein. Lead compounds and/or binding moieties identified during these screens can serve as the basis for the synthesis of more active analogs. Lead compounds and/or active analogs generated therefrom can be formulated into pharmaceutical compositions effective in treating brain tumors.

DISEASE CONDITIONS

The present methods are applicable to brain tumors, particularly glioblastoma. In general, the goals of brain tumor treatments are to remove as many tumor cells as possible, *e.g.* with surgery,

kill as many of the cells left behind after surgery as possible with radiation and/or chemotherapy, and put remaining tumor cells into a nondividing, quiescent state for as long as possible with radiation and chemotherapy. Careful imaging surveillance is a crucial part of medical care, because tumor regrowth requires alteration of current treatment, or, for patients in the observation phase, restarting
5 treatment.

Brain tumors are classified according to the kind of cell from which the tumor seems to originate. Diffuse, fibrillary astrocytomas are the most common type of primary brain tumor in adults. These tumors are divided histopathologically into three grades of malignancy: World Health Organization (WHO) grade II astrocytoma, WHO grade III anaplastic astrocytoma and WHO grade
10 IV glioblastoma multiforme (GBM). WHO grade II astrocytomas are the most indolent of the diffuse astrocytoma spectrum. Astrocytomas display a remarkable tendency to infiltrate the surrounding brain, confounding therapeutic attempts at local control. These invasive abilities are often apparent in low-grade as well as high-grade tumors.

Glioblastoma multiforme is the most malignant stage of astrocytoma, with survival times of
15 less than 2 years for most patients. Histologically, these tumors are characterized by dense cellularity, high proliferation indices, endothelial proliferation and focal necrosis. The highly proliferative nature of these lesions likely results from multiple mitogenic effects. One of the hallmarks of GBM is endothelial proliferation. A host of angiogenic growth factors and their receptors are found in GBMs.

20 There are biologic subsets of astrocytomas, which may reflect the clinical heterogeneity observed in these tumors. These subsets include brain stem gliomas, which are a form of pediatric diffuse, fibrillary astrocytoma that often follow a malignant course. Brain stem GBMs share genetic features with those adult GBMs that affect younger patients. Pleomorphic xanthoastrocytoma (PXA) is a superficial, low-grade astrocytic tumor that predominantly affects young adults. While these
25 tumors have a bizarre histological appearance, they are typically slow-growing tumors that may be amenable to surgical cure. Some PXAs, however, may recur as GBM. Pilocytic astrocytoma is the most common astrocytic tumor of childhood and differs clinically and histopathologically from the diffuse, fibrillary astrocytoma that affects adults. Pilocytic astrocytomas do not have the same genomic alterations as diffuse, fibrillary astrocytomas. Subependymal giant cell astrocytomas
30 (SEGA) are periventricular, low-grade astrocytic tumors that are usually associated with tuberous sclerosis (TS), and are histologically identical to the so-called "candle-gutterings" that line the ventricles of TS patients. Similar to the other tumorous lesions in TS, these are slowly-growing and may be more akin to hamartomas than true neoplasms. Desmoplastic cerebral astrocytoma of infancy (DCAI) and desmoplastic infantile ganglioglioma (DIGG) are large, superficial, usually
35 cystic, benign astrocytomas that affect children in the first year or two of life.

Oligodendrogliomas and oligoastrocytomas (mixed gliomas) are diffuse, usually cerebral tumors that are clinically and biologically most closely related to the diffuse, fibrillary astrocytomas.

The tumors, however, are far less common than astrocytomas and have generally better prognoses than the diffuse astrocytomas. Oligodendrogliomas and oligoastrocytomas may progress, either to WHO grade III anaplastic oligodendroglioma or anaplastic oligoastrocytoma, or to WHO grade IV GBM. Thus, the genetic changes that lead to oligodendroglial tumors constitute yet another pathway to GBM.

Ependymomas are a clinically diverse group of gliomas that vary from aggressive intraventricular tumors of children to benign spinal cord tumors in adults. Transitions of ependymoma to GBM are rare. Choroid plexus tumors are also a varied group of tumors that preferentially occur in the ventricular system, ranging from aggressive supratentorial intraventricular tumors of children to benign cerebellopontine angle tumors of adults. Choroid plexus tumors have been reported occasionally in patients with Li-Fraumeni syndrome and von Hippel-Lindau (VHL) disease.

Medulloblastomas are highly malignant, primitive tumors that arise in the posterior fossa, primarily in children. Meningiomas are common intracranial tumors that arise in the meninges and compress the underlying brain. Meningiomas are usually benign, but some "atypical" meningiomas may recur locally, and some meningiomas are frankly malignant and may invade the brain or metastasize. Atypical and malignant meningiomas are not as common as benign meningiomas. Schwannomas are benign tumors that arise on peripheral nerves. Schwannomas may arise on cranial nerves, particularly the vestibular portion of the eighth cranial nerve (vestibular schwannomas, acoustic neuromas) where they present as cerebellopontine angle masses. Hemangioblastomas are tumors of uncertain origin that are composed of endothelial cells, pericytes and so-called stromal cells. These benign tumors most frequently occur in the cerebellum and spinal cord of young adults. Multiple hemangioblastomas are characteristic of von Hippel-Lindau disease (VHL). Hemangiopericytomas (HPCs) are dural tumors which may display locally aggressive behavior and may metastasize. The histogenesis of dural-based hemangiopericytoma (HPC) has long been debated, with some authors classifying it as a distinct entity and others classifying it as a subtype of meningioma.

The symptoms of both primary and metastatic brain tumors depend mainly on the location in the brain and the size of the tumor. Since each area of the brain is responsible for specific functions, the symptoms will vary a great deal. Tumors in the frontal lobe of the brain may cause weakness and paralysis, mood disturbances, difficulty in thinking, confusion and disorientation, and wide emotional mood swings. Parietal lobe tumors may cause seizures, numbness or paralysis, difficulty with handwriting, inability to perform simple mathematical problems, difficulty with certain movements, and loss of the sense of touch. Tumors in the occipital lobe can cause loss of vision in half of each visual field, visual hallucinations, and seizures. Temporal lobe tumors can cause seizures, perceptual and spatial disturbances, and receptive aphasia. If a tumor occurs in the cerebellum, the person may have ataxia, loss of coordination, headaches, and vomiting. Tumors in the hypothalamus may cause

emotional changes, and changes in the perception of hot and cold. In addition, hypothalamic tumors may affect growth and nutrition in children. With the exception of the cerebellum, a tumor on one side of the brain causes symptoms and impairment on the opposite side of the body.

Other disorders of the nervous system that may be treated or imaged with the compositions of the present invention include, but are not limited to ischemic stroke, brain cancer, epilepsy, schizophrenia, depression, Alzheimer's Disease, Parkinson's Disease, Huntington's Chorea, traumatic head injury, dementia, coma, stupor, headache (and other neurological pain), vertigo, weakness, myasthenia gravis (and other disorders of the neuromuscular junction), ataxia and cerebellar disorders, cranial nerve disorders (such as Bell's Palsy), cerebrovascular disorders, infectious disorders including bacterial, fungal, viral and parasitic infections, multiple sclerosis, and other complications associated with pregnancy, medical illness, alcohol and substance abuse, toxins and metabolic deficiencies.

IDENTIFICATION OF T_{BT} GENES

A genetic sequence that comprises all or a part of a cDNA sequence that is differentially expressed in brain tumor cells, particularly glioblastoma cells, relative to expression in normal, or non-disease conditions, is herein termed a "T_{BT} gene", which encode "T_{BT} proteins". T_{BT} genes were identified by creating subtracted and normalized cDNA libraries from glioblastoma tissues. The cDNA's from control and disease states were subjected to kinetic re-annealing hybridization during which normalization of transcript abundances and enrichment for low-abundance transcripts-occurs. Differential up- or down-regulated transcripts in tumors can be enriched by a subsequent "forward" or "reverse" subtraction step by using a second driver cDNA. Only clones displaying a significant transcriptional induction and/or repression were sequenced and carried forward for expression profiling, using a variety of temporal, spatial and disease-related probe sets. Selected clones showing a significant transcriptional induction and/or repression were sequenced and functionally annotated in a proprietary database structure (See WO01/13105). Because large sequence fragments were utilized in the sequencing step, the data generated has a much higher fidelity and specificity than other approaches, such as SAGE. The resulting sequence information was compared to public databases using the BLAST (blastn) algorithm for DNA sequence comparisons and iterative-Smith Waterman analysis for protein sequence comparisons. The results are listed in Table 1. Table 1 includes, in some instances, the human and animal counterparts of a sequence, as indicated by a shared internal reference designation.

Table 1

SEQ NO.	ID	INTERNAL REFERENCE	GENBANK ACCESSION	GENE
1, 2		SL00045_CP1_E01	NM_002851	Homo sapiens protein tyrosine phosphatase, receptor-type, Z polypeptide 1 (PTPRZ1), mRNA
3, 4				Homo sapiens protein tyrosine phosphatase zeta SM1 variant
5, 6				Homo sapiens protein tyrosine phosphatase zeta SM2 variant
7				ARP2; (Kim et al. (1999) JBC 274:26523-26528)
8			XM_032759	SPARC (osteonectin).
9				CMET; (Park et al. (1987) PNAS 84:6379-6383)
10		SL00045_CP1_J23	XM_006083.2; U40373.1	Homo sapiens CD44 antigen
11		SL00045_CP2_J23	U46958.1	Rattus norvegicus CD44i mRNA
12		SL00045_CP1_B11	AF228710.1	Homo sapiens chondroitin sulfate proteoglycan BEHAB/brevican
13		SL00045_CP2_D06	BC004280	Homo sapiens, tetraspan 3
14		SL00045_CP3_I19	Y18430.1	Homo sapiens VIP2R gene, exons 9-12
15		SL00043_CP1_D18	U20758.1	Human osteopontin
16		SL00043_CP1_C19	D90226.1 AB004306.1	Human mRNA for pleiotrophin/OSF-1

Transcripts that represent differentially expressed genes may be identified by utilizing a variety of methods known to those of skill in the art, including differential screening, subtractive hybridization, differential display, or hybridization to an array comprising a plurality of gene sequences.

"Differential expression" as used herein refers to both quantitative as well as qualitative differences in the genes' temporal and/or tissue expression patterns. Thus, a differentially expressed gene may have its expression activated or inactivated in normal versus neuronal disease conditions, or in control versus experimental conditions. Such a qualitatively regulated gene will exhibit an expression pattern within a given tissue or cell type that is detectable in either control or tumor samples, but is not detectable in both. Detectable, as used herein, refers to an RNA expression pattern that is detectable via the standard techniques of differential display, reverse transcription- (RT-) PCR and/or Northern analyses, which are well known to those of skill in the art. Generally, differential expression means that there is at least a 20% change, and in other instances at least a 2-, 3-, 5- or 10-fold difference between disease and control tissue expression. The difference usually is one that is statistically significant, meaning that the probability of the difference occurring by chance (the P-value) is less than some predetermined level (e.g., 5%). Usually the confidence level (P value) is <0.05, more typically <0.01, and in other instances, <0.001.

Alternatively, a differentially expressed gene may have its expression modulated, i.e., quantitatively increased or decreased, in normal versus neuronal disease states, or under control versus experimental conditions. The difference in expression need only be large enough to be

visualized via standard detection techniques as described above. Generally the difference in expression levels, measured by either the presence of mRNA or the protein product, will differ from basal levels (i.e. normal tissue) by at least about 2 fold, usually at least about 5 fold, and may be 10 fold, 100 fold, or more.

5 Identification of T_{BT} pathway genes may be performed through physical association of gene products, or through database identification of known physiological pathways. Among the methods for detecting protein-protein association are co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. The two-hybrid system detects the association of proteins *in vivo*, as described by Chien *et al.* (1991) Proc. Natl. Acad. Sci. USA **88**:9578-9582. The
10 two-hybrid system or related methodology may be used to screen activation domain libraries for proteins that interact with a known "bait" gene protein.

Once a sequence has been identified as differentially expressed, the sequence can be subjected to a functional validation process to determine whether the gene plays a role in tumor initiation, progression or maintenance. Such candidate genes can potentially be correlated with a
15 wide variety of cellular states or activities. The term "functional validation" as used herein refers to a process whereby one determines whether modulation of expression or function of a candidate gene or set of such genes causes a detectable change in a cellular activity or cellular state for a reference cell, which can be a population of cells such as a tissue or an entire organism. The detectable change or alteration that is detected can be any activity carried out by the reference cell. Specific examples of
20 activities or states in which alterations can be detected include, but are not limited to, phenotypic changes (e.g., cell morphology, cell proliferation, cell viability and cell death); cells acquiring resistance to a prior sensitivity or acquiring a sensitivity which previously did not exist; protein/protein interactions; cell movement; intracellular or intercellular signaling; cell/cell interactions; cell activation (e.g., T cell activation, B cell activation, mast cell degranulation); release
25 of cellular components (e.g., hormones, chemokines and the like); and metabolic or catabolic reactions.

A variety of options are available for functionally validating candidate genes. Such methods as RNAi technology can be used. Antisense technology can also be utilized to functionally validate a candidate gene. In this approach, an antisense polynucleotide that specifically hybridizes to a
30 segment of the coding sequence for the candidate gene is administered to inhibit expression of the candidate gene in those cells into which it is introduced. The functional role that a candidate gene plays in a cell can also be assessed using gene "knockout" approaches in which the candidate gene is deleted, modified, or inhibited on either a single or both alleles. The cells or animals can be optionally be reconstituted with a wild-type candidate gene as part of a further analysis.

35 In one embodiment of the invention, RNAi technology is used in functional validation. As used herein, RNAi technology refers to a process in which double-stranded RNA is introduced into cells expressing a candidate gene to inhibit expression of the candidate gene, *i.e.*, to "silence" its

expression. The dsRNA is selected to have substantial identity with the candidate gene. In general such methods initially involve transcribing a nucleic acids containing all or part of a candidate gene into single- or double-stranded RNA. Sense and anti-sense RNA strands are allowed to anneal under appropriate conditions to form dsRNA. The resulting dsRNA is introduced into reference cells via various methods and the degree of attenuation in expression of the candidate gene is measured using various techniques. Usually one detects whether inhibition alters a cellular state or cellular activity. The dsRNA is prepared to be substantially identical to at least a segment of a candidate gene. Because only substantial sequence similarity between the candidate gene and the dsRNA is necessary, sequence variations between these two species arising from genetic mutations, evolutionary divergence and polymorphisms can be tolerated. Moreover, the dsRNA can include various modified or nucleotide analogs. Usually the dsRNA consists of two separate complementary RNA strands. However, in some instances, the dsRNA may be formed by a single strand of RNA that is self-complementary, such that the strand loops back upon itself to form a hairpin loop. Regardless of form, RNA duplex formation can occur inside or outside of a cell.

dsRNA can be prepared according to any of a number of methods that are known in the art, including *in vitro* and *in vivo* methods, as well as by synthetic chemistry approaches. Examples of such methods include, but are not limited to, the methods described by Sadher *et al.* (Biochem. Int. 14:1015, 1987); by Bhattacharyya (Nature 343:484, 1990); and by Livache, *et al.* (U.S. Patent No. 5,795,715), each of which is incorporated herein by reference in its entirety. Single-stranded RNA can also be produced using a combination of enzymatic and organic synthesis or by total organic synthesis. The use of synthetic chemical methods enable one to introduce desired modified nucleotides or nucleotide analogs into the dsRNA. dsRNA can also be prepared *in vivo* according to a number of established methods (see, e.g., Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed.; Transcription and Translation (B.D. Hames, and S.J. Higgins, Eds., 1984); DNA Cloning, volumes I and II (D.N. Glover, Ed., 1985); and Oligonucleotide Synthesis (M.J. Gait, Ed., 1984, each of which is incorporated herein by reference in its entirety). Once the dsRNA has been formed, it is introduced into a cell. For example, a neuroblastoma cell line can serve as a model system for investigating the functional relevance of genes, identified by procedures described herein, to tumor growth, metabolism or metastasis.

A number of options can be utilized to deliver the dsRNA into a cell or population of cells such as in a cell culture, tissue or embryo. For instance, RNA can be directly introduced intracellularly. Various physical methods are generally utilized in such instances, such as administration by microinjection (see, e.g., Zernicka-Goetz, et al. (1997) Development 124:1133-1137; and Wianny, et al. (1998) Chromosoma 107: 430-439). Other options for cellular delivery include permeabilizing the cell membrane and electroporation in the presence of the dsRNA, liposome-mediated transfection, or transfection using chemicals such as calcium phosphate. A number of established gene therapy techniques can also be utilized to introduce the dsRNA into a

cell. By introducing a viral construct within a viral particle, for instance, one can achieve efficient introduction of an expression construct into the cell and transcription of the RNA encoded by the construct.

A number of options are available to detect interference of candidate gene expression (i.e., to detect candidate gene silencing). In general, inhibition in expression is detected by detecting a decrease in the level of the protein encoded by the candidate gene, determining the level of mRNA transcribed from the gene and/or detecting a change in phenotype associated with candidate gene expression.

TBT GENES AND POLYPEPTIDES

ARP-2 (SEQ ID NO:7) is a 64 kDa, single chain, acidic, angiopoietin-like protein that includes multiple functional domains, such as a hydrophobic signal sequence from amino acids 1-21, a coiled-coil domain at the amino terminal end from approximately amino acid sequences 22-274, and a fibrinogen-like domain, from approximately about residues 275 through 493. Two major isoforms have been observed, one 2.4 Kb in size and the other about 4 Kb. Epitopes of interest include the fibrinogen region, the coiled coil domain, extracellular region, *etc.* The fibrinogen domain of human ARP-2 is hypothesized to interact with one or more an unknown receptor for the purposes of angiogenesis. The interaction of ARP-2 to these molecules may be through either of the aforementioned structural motifs.

SPARC (SEQ ID NO:8) is an abundant 33 kDa, single chain, acidic, extracellular calcium binding protein that contains a flexible N-terminal acidic domain I (~50 amino acids), a follistatin-like (FS) domain (~75 residues), and a C-terminal extracellular calcium-binding (EC) domain with a pair of EF-hand loops (~150 residues). The N-terminal domain shows a low affinity Ca²⁺ binding site, a transglutaminase cross linking site, and inhibits cell spreading in cell culture assays. Calcium-dependent binding of SPARC to the triple helix of several fibrillar collagen types and basement membrane collagen type IV has been mapped to the EC domain. Two isoforms have been described, bone SPARC with a molecular weight of 31,000 kDa and platelet SPARC with a molecular weight of 33,000 kDa. The EC domain of human SPARC is known to interact with the collagens I, III, IV and V, and to bind to vitronectin, all of which are components of the extracellular matrix surrounding gliomas. The binding of SPARC to these molecules may play a significant role in the oncogenesis and growth of neoplastic cells in the brain.

c-MET (SEQ ID NO:9) is a type I membrane protein heterodimer. Two different receptor variants originate by post-translational processing of a common single-chain precursor of 170 kDa. Isoform p190MET is formed of a 50 kDa α -chain and a 145 kDa β -chain that are disulfide linked, and isoform p140Met is composed of a 50 kDa α -chain and an 85 kDa β -chain, lacking the cytoplasmic kinase domain. This 85 kDa β chain is likely a trans-membrane glycoprotein that is bound to the cell surface. Truncated forms of c-MET containing the 50 kDa α -chain and a carboxyl-

terminally truncated 75 kDa β sub-unit have also been described. The 75 kDa form arises by post-translational proteolytic processing, lacks the trans-membrane domain, and is secreted from the cell. Both the secreted and membrane bound forms are of interest. The amino acid sequence of full length c-MET consists of 1408 amino acids (Park et al. (1987) P.N.A.S. **84**:6379-6383) and 1390 amino acid (Prat *et al.* (1991) *Mol. Cell. Biol.* **11**:5954-5962). The signal sequence is comprised of N-terminal amino acids 1-24; the α chain (amino acids 24-306) makes up the extracellular domain; and the β chain extracellular domain amino acids 306 to 932. The transmembrane segment is amino acids 933 to 955, and the intracellular domain is composed of amino acids 956 to 1390.

BEHAB (SQ ID NO:12) exists in two isoforms: a full-length isoform that is secreted into the extracellular matrix and a shorter isoform that has a hydrophobic carboxy terminus, which predicts a glycosphosphatidylinositol (GPI) anchor. BEHAB contains an N-terminal hyaluronan (HA)-binding domain, which comprises an immunoglobulin-like loop and two proteoglycan tandem repeats, a C-terminal epidermal growth factor (EGF)-like repeat, a C-type lectin-like domain, and a complement regulatory protein (CRP)-like domain. The central region of the protein contains sites for glycosylation and proteolytic cleavage (between glu395-Ser396 of the mature protein, after signal peptide cleavage) by a metallo-protease. The complete cDNA of the secreted isoform is 2878 bp encoding 912 amino acids of 99 kDa. The GPI isoform is 2558 bp. The GPI-linked form is generated by a 'no splice' event, with the transcript reading through an exon/intron junction thereby extending the open reading frame to a stop codon 74 nucleotides further downstream. Up-regulation of BEHAB may be a crucial step in returning the unmalleable mature extracellular matrix to a more immature matrix, permissive for cell growth, thereby promoting the progression of primary brain tumors.

CD-44 (SEQ ID NO:10; 11) is a proteoglycan that is expressed as two major splice variants. CD-44E is a 150 kDa protein isolated from epithelial cells. CD-44E has a C-terminal cytoplasmic tail, a hydrophobic transmembrane domain of 23 amino acids, and an N-terminal extracellular region of 248 amino acids. The extracellular domain is O-glycosylated and also binds chondroitin sulfate. In addition, CD-44E has two of the three immunodominant epitope clusters of native gp90Hermes. CD-44E contains an additional 132 amino acids in the extracellular region. and CD-44H is a 90kDa protein isolated from hematopoietic cells. In addition, CD-44R1 and CD-44R2 are 2 isoforms expressed by hematopoietic cells. The complete cDNA sequence of the 90 kDa CD-44H isoform consist of 1795 bps, encoding a 341 amino acid protein. CD-44H protein has an overall primary structure of 90 kDa, which consist of 341 amino acids. The N-terminus is located outside of the cell and the extracellular domain consist of 248 amino acids. The C-terminus is located inside of the cell and the intracellular domain consist of 72 amino acids, while the transmembrane region consist of 21 amino acids. The CD-44 gene contains 20 exons, of which exons 1-5, 15-17 and 19 encode the CD44H isoform. The intervening exons 6, 6a, 7-14 (also designated v1-v10) are alternatively spliced to generate the variant isoforms with an insertion at the membrane proximal region of the

extracellular domain between amino acids 202 and 203. CD-44 is one of the principal receptors for hyaluronic acid. Within the normal CNS, the CD-44 protein has been localized to astrocytes in the white matter. CD-44H has been shown to be the predominant isoform in normal brain and neuroectoderm-derived tumors. Hence, the up-regulation of CD-44 may be a crucial step in brain
 5 tumor invasiveness and migration.

Tetraspanin (TSPAN3) (SEQ ID NO:13) is a 253 amino acid membrane bound protein. TSPAN3 contains four transmembrane domains, comprising amino acids 12-32, 51-71, 86-106, and 213-233. The protein has two extracellular domains, amino acids 33-50 and 107-212, and three cytoplasmic domains, amino acids 1-11, 72-85, and 234-235. The cysteine residues at positions 147,
 10 148, and 197 in the second extracellular domain are highly conserved among the tetraspanin family and are thought to be essential for proper tetraspanin function.

VIPR-2 (SEQ ID NO:14) is a seven transmembrane spanning G-protein receptor. The complete VIPR-2 protein is encoded by 13 exons. The initiator codon of the predicted 438 amino acid-encoding open reading frame is located in exon 1 and the termination signal is located in exon
 15 13. The 5' untranslated region extends 187 bp upstream of the initiator codon and is extremely GC-rich (80%). The polyadenylation signal is located 2416 bp downstream of the stop codon. Intron sizes range from 68 bp (intron 11) to 45 bp (intron 4), the entire human gene spans 117 kb, while the cDNA sequence spans 1317 bp. Recent studies have also isolated two VIP-2 receptor mRNAs of 4.6 kb and 2.3 kb in size. The receptor is a seven membrane spanning protein where the first 22 amino
 20 acids constitute a signal sequence, and the remaining amino acids constitute two membrane spanning regions between amino acids 127 to 148 and 158 to 178, two more membrane spanning domains between amino acids 202 to 227 and 238 to 261, another between 278 to 303, and two final membrane spanning regions between 327 to 347 and 359 to 380, with three potential N-linked glycosylation sites found in the amino terminal extracellular domain at residues 57, 87 and 91. The
 25 extracellular domain of human VIPR-2 binds PACAP-27, PACAP-38, VIP and secretin.

PTN, or OSF-1 (SEQ ID NO:16) is a 18 kDa, single chain, secreted protein with 10 conserved disulfide linked cysteine residues. The gene sequence of human PTN consists of five exons and four introns. While exon 1 does not encode an amino acid sequence, exon 2 encodes the hydrophobic signal sequence of 32 amino acids, exons 3 and 4 code for the amino terminal and the
 30 ten cysteine residues, and exon 5 codes for the highly basic C-terminal domains. The mature protein consist of 136 amino acids encoded by exons 2 to 5. PTN has been shown to bind to the extracellular domain of RPTP beta and zeta. This binding inactivates the catalytic activity of RPTP, and PTN binds all the three major isoforms pf RPTP beta and zeta. PTN has also been shown to interact with syndecan-3.

OPN (SEQ ID NO:15) is an abundant 34 kDa, single chain, phosphorylated glycoprotein, with a presumed site for cell attachment at residues 144-148. Three isoforms have been identified to be generated by post transcriptional modification, such as alternative splicing, OPN-A, OPN-B, and
 35

OPN-C. OPN-A and OPN-B differ by the addition of 14 amino acids at residue 58 of the protein. Amino acids 58-71 are absent in OPN-B, and amino acids 31-57 are absent in OPN-C. OPN is a negatively charged, highly hydrophilic secreted protein. The cDNA sequence of human bone OPN (OPN-A) has an overall structure of approximately 34 kDA that consist of 298 amino acids, which is 14 amino acids less than the cDNA sequence of OPN isolated from human osteosarcoma (OPN-B). The cDNA transcript for OPN-A is 1.5 kb with an open reading frame of 900 nucleotides, of which the first 16 amino acids are hydrophobic in nature and constitute a signal sequence for the secreted protein. The OPN gene contains 7 exons that are alternatively spliced to generate the variant isoforms, the most common variant being the addition of a 42 bp (14 amino acid) sequence located at base 280 of OPN-A. The cell attachment sequence of human OPN (amino acids 144 to 148) is believed to interact with various cell surface proteins (such as CD-44) to affect cell adhesion, and a highly acidic stretch composed almost exclusively of aspartic acid residues (amino acids 72 to 81) is believed to be the mineral binding site within the protein.

PTP ζ exists in several splice variants, including two membrane bound variants (full length: PTP ζ - α , and shorter version PTP ζ - β) and one secreted form (Phosphacan). Isoform PTP ζ - α is the full length isoform, which contains the primary amino acid sequence aa 25-2314 (aa 1-24 are a signal polypeptide) as shown in SEQ ID NO:1 and SEQ ID NO:2. This full length long form of PTP ζ is a type I membrane protein. After the signal peptide it contains a carbonic anhydrase like (CAH) and a fibronectin type III like (FN3) domain, followed by a long cysteine free spacer (S) domain. This follows a 860 amino acid long insert domain, which can be glycosylated. After a single transmembrane segment, in the intracellular region it has 2 phosphatase domains, but only the membrane-proximal PTPase domain is catalytically active (Krueger 1992).

In isoform PTP ζ - β , aa 755-1614 (corresponding to SEQ ID NO:2 numbering) are missing. Isoform PTP ζ -S (phosphacan), is a secreted isoform, which comprises the extracellular domains of PTP ζ - α . Applicants provide herein two additional novel splice variants, PTP ζ SM1 and PTP ζ SM2, which are described in detail below. Applicants verified the location of the novel sequences by comparison of the known splice variant sequences and the novel sequences with a publicly available genomic sequence database.

The amino acid sequence of full length PTP ζ consists of 2307 amino acids (See U.S. Patent Nos. 5,604,094, and 6,160,090, fully incorporated herein by reference), or 2314 amino acids (Krueger *et al.* (1992) *P.N.A.S.* 89:7417-7421). Amino acids 1-24 of SEQ ID NO. 2 are a signal sequence which directs the proper placement of the transmembrane protein. The extracellular domain of the mature PTP ζ protein spans amino acids 25-1635 of SEQ ID NO. 2 in the long and secreted forms, and amino acids 25-754, 1615-1635 in the short isoform. The transmembrane region of the protein spans amino acids 1636-1661 of SEQ ID NO. 2, and the balance of the protein forms the cytoplasmic domain, amino acids 1662-2314.

The extracellular domain of human PTP ζ is known to bind to tenascin-C, tenascin-R, pleiotrophin (NM_002825), midkine (NM_002391), FGF-2 (XM_00366), Nr-CAM (NM_005010), L1/Ng-CAM, contactin (NM_001843), N-CAM (XM_006332), and axonin-1 (NM_005076.) The first 5 of these molecules are either components of the extracellular matrix in gliomas or are soluble factors known to be present in gliomas, and the latter 4 are neuronal surface molecules. The binding of PTP ζ to these molecules may play a significant role in the oncogenesis and growth of neoplastic cells in the brain

The protein PTP ζ SM1 (SEQ ID NO:3, 4) comprises the amino acids encoded by the first nine exons of PTP ζ - α , with three unique additional carboxy terminal amino acids. These are encoded by additional 3' mRNA sequence (nucleotides 1262-1272) from the intron of the gene between exons nine and ten. The PTP ζ SM1 clone was isolated from a human fetal brain cDNA library, and has been shown to be expressed in several human glioblastoma cell lines. Expression of the SM1 splice variant has also been confirmed in primary brain tumor samples. The protein comprises only extracellular domains of PTP ζ , and is expected to be secreted by the cell. Thus, PTP ζ SM1 may serve a cell signaling or messenger function, and may bind to a receptor on the surface of cells which are associated with or part of central nervous system tissues. The PTP ζ SM1 protein mainly comprises the carbonic anhydrase-like domain.

A comparison was performed of the carbonic anhydrase domain of PTP ζ SM1 and other human carbonic anhydrase domains (carbonic anhydrase III, carbonic anhydrase I, and carbonic anhydrase VIX [e]). Based on alignment with these catalytically active carbonic anhydrases, it seems unlikely that the CA domain could function as a carbonic anhydrase enzyme. Two of the three histidines implicated in binding of the catalytic zinc are missing from the CA domain of the receptor. In active enzymes there is a conserved HxHWG{18,20}ELH motif (the three histidines bind zinc), however, in the receptor this is modified to TFHWG{18,20}EMQ; i.e. two of the three critical zinc atoms would be missing. For comparison, it has been found that a carbonic anhydrase related protein (CAH 8) that lacks just one of these histidines also lacks catalytic activity.

The protein PTP ζ SM2 (SEQ ID NO. 5, 6) comprises the amino acids encoded by all exons of PTP ζ - α , plus a 116 nucleotide "extra" exon, in the correct reading frame, between exons 23 and 24 (nucleotides 6229-6345 of SEQ ID NO. 3). This extra exon, designated exon 23a, contains a portion of the intron sequence between exons 23 and 24 of the PTP ζ gene. PTP ζ SM2 expression has been verified in several human glioblastoma cell lines, and has also been confirmed in primary brain tumor samples. As PTP ζ SM2 comprises all the domains of PTP ζ α , the protein is expected to be membrane-bound. The extra exon lies within the cytoplasmic domain of the protein, and thus may alter the protein tyrosine phosphatase function of PTP ζ SM2.

Because PTP ζ SM1 and PTP ζ SM2 have been shown to be expressed in glioblastoma cell lines and primary tumors, the level of the expression of these splice variants may be useful for

staging or characterizing glioblastoma cells. Such cells may be extracted, for instance, from a primary tumor. Thus, the invention provides for the monitoring of the relative expression level of PTP ζ SM1 or PTP ζ SM2, or both, in relation to each other or to one or more of the known PTP ζ splice variants. In one preferred embodiment, the level of expression of PTP ζ SM1 is compared to at least one other splice variant selected from PTP ζ SM2, PTP ζ α , PTP ζ β , and phosphacan. In another preferred embodiment, the level of expression of PTP ζ SM2 is compared to at least one other splice variant selected from PTP ζ SM1, PTP ζ α , PTP ζ β , and phosphacan. Such comparison may be made in either a qualitative or quantitative manner.

NUCLEIC ACIDS

The sequences of T_{BT} genes find use in diagnostic and therapeutic methods, for the recombinant production of the encoded polypeptide, and the like. The nucleic acids of the invention include nucleic acids having a high degree of sequence similarity or sequence identity to one of the sequences provided in Table 1. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM NaCl/0.9 mM Na citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, U.S. patent 5,707,829. Nucleic acids that are substantially identical to the provided nucleic acid sequence, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to one of the sequences provided in Table 1 under stringent hybridization conditions. Further specific guidance regarding the preparation of nucleic acids is provided by Fleury *et al.* (1997) Nature Genetics 15:269-272; Tartaglia *et al.*, PCT Publication No. WO 96/05861; and Chen *et al.*, PCT Publication No. WO 00/06087, each of which is incorporated herein in its entirety.

The genes listed in Table 1 may be obtained using various methods well known to those skilled in the art, including but not limited to the use of appropriate probes to detect the genes within an appropriate cDNA or genomic DNA library, antibody screening of expression libraries to detect cloned DNA fragments with shared structural features, direct chemical synthesis, and amplification protocols. Libraries are preferably prepared from cells or tissues of normal brains or brain tumors. Cloning methods are described in Berger and Kimmel, *Guide to Molecular Cloning Techniques*, Methods in Enzymology, 152, Academic Press, Inc. San Diego, CA; Sambrook, *et al.* (1989) *Molecular Cloning – A Laboratory Manual* (2nd ed) Vols. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY; and *Current Protocols* (1994), a joint venture between Greene Publishing Associates, Inc. and John Wiley and Sons, Inc.

The sequence obtained from clones containing partial coding sequences or non-coding sequences can be used to obtain the entire coding region by using the RACE method (Chenchik *et al.* (1995) *CLONTECHniques* (X) 1: 5-8). Oligonucleotides can be designed from the partial clone's analyzed sequence and subsequently utilized to amplify a reverse transcribed mRNA encoding the entire coding sequence. Alternatively, probes can be used to screen cDNA libraries prepared from an

appropriate cell or cell line in which the gene is transcribed. Once the target nucleic acid is identified, it can be isolated and cloned using well-known amplification techniques. Such techniques include, the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q β -replicase amplification, the self-sustained sequence replication system (SSR) and the transcription based amplification system (TAS). Such methods include, those described, for example, in U.S. Patent No. 4,683,202 to Mullis et al.; PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990); Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86: 1173; Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87: 1874; Lomell et al. (1989) J. Clin. Chem. 35: 1826; Landegren et al. (1988) Science 241: 1077-1080; Van Brunt (1990) Biotechnology 8: 291-294; Wu and Wallace (1989) Gene 4: 560; and Barringer et al. (1990) Gene 89: 117.

As an alternative to cloning a nucleic acid, a suitable nucleic acid can be chemically synthesized. Direct chemical synthesis methods include, for example, the phosphotriester method of Narang et al. (1979) Meth. Enzymol. 68: 90-99; the phosphodiester method of Brown et al. (1979) Meth. Enzymol. 68: 109-151; the diethylphosphoramidite method of Beaucage et al. (1981) Tetra. Lett., 22: 1859-1862; and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. While chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences can be obtained by the ligation of shorter sequences. Alternatively, subsequences may be cloned and the appropriate subsequences cleaved using appropriate restriction enzymes.

The nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof. The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression, and are useful for investigating the up-regulation of expression in tumor cells.

Probes specific to the nucleic acid of the invention can be generated using the nucleic acid sequence disclosed in Table 1. The probes are preferably at least about 18 nt, 25 nt, 50 nt or more of the corresponding contiguous sequence of one of the sequences provided in Table 1, and are usually less than about 2, 1, or 0.5 kb in length. Preferably, probes are designed based on a contiguous sequence that remains unmasked following application of a masking program for masking low complexity. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag.

The nucleic acids of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the nucleic acids, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant," *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The nucleic acids of the invention can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the nucleic acids can be regulated by their own or by other regulatory sequences known in the art. The nucleic acids of the invention can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other. For hybridization probes, it may be desirable to use nucleic acid analogs, in order to improve the stability and binding affinity. The term "nucleic acid" shall be understood to encompass such analogs.

POLYPEPTIDES

Polypeptides encoded by T_{BT} genes are of interest for screening methods, as reagents to raise antibodies, as therapeutics, and the like. Such polypeptides can be produced through isolation from natural sources, recombinant methods and chemical synthesis. In addition, functionally equivalent

polypeptides may find use, where the equivalent polypeptide may contain deletions, additions or substitutions of amino acid residues that result in a silent change, thus producing a functionally equivalent differentially expressed on pathway gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the
5 amphipathic nature of the residues involved. "Functionally equivalent", as used herein, refers to a protein capable of exhibiting a substantially similar *in vivo* activity as the polypeptide encoded by an ischemia associated gene, as provided in Table 1.

The polypeptides may be produced by recombinant DNA technology using techniques well known in the art. Methods which are well known to those skilled in the art can be used to construct
10 expression vectors containing coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. Alternatively, RNA capable of encoding the polypeptides of interest may be chemically synthesized.

Typically, the coding sequence is placed under the control of a promoter that is functional in
15 the desired host cell to produce relatively large quantities of the gene product. An extremely wide variety of promoters are well-known, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include
20 one or more of these control sequences are termed "expression cassettes." Expression can be achieved in prokaryotic and eukaryotic cells utilizing promoters and other regulatory agents appropriate for the particular host cell. Exemplary host cells include, but are not limited to, *E. coli*, other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines.

In mammalian host cells, a number of viral-based expression systems may be used, including
25 retrovirus, lentivirus, adenovirus, adeno-associated virus, and the like. In cases where an adenovirus is used as an expression vector, the coding sequence of interest can be ligated to an adenovirus transcription/translation control complex, *e.g.*, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination.
30 Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing differentially expressed or pathway gene protein in infected hosts.

Specific initiation signals may also be required for efficient translation of the genes. These signals include the ATG initiation codon and adjacent sequences. In cases where a complete gene,
35 including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the gene coding sequence is inserted, exogenous translational control signals must be

provided. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, *etc.*

5 In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign
10 protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, *etc.*

For long-term, high-yield production of recombinant proteins, stable expression is preferred.
15 For example, cell lines that stably express the differentially expressed or pathway gene protein may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements, and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable
20 marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines that express the target protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the T_{BT} protein. A number of selection systems
25 may be used, including but not limited to the herpes simplex virus thymidine kinase, hypoxanthine-guanine phosphoribosyltransferase, and adenine phosphoribosyltransferase genes. Antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate; gpt, which confers resistance to mycophenolic acid; neo, which confers resistance to the aminoglycoside G-418; and hygromycin.

30 The polypeptide may be labeled, either directly or indirectly. Any of a variety of suitable labeling systems may be used, including but not limited to, radioisotopes such as ¹²⁵I; enzyme labeling systems that generate a detectable colorimetric signal or light when exposed to substrate; and fluorescent labels. Indirect labeling involves the use of a protein, such as a labeled antibody, that specifically binds to the polypeptide of interest. Such antibodies include but are not limited to
35 polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

Once expressed, the recombinant polypeptides can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, ion exchange and/or size exclusivity chromatography, gel electrophoresis and the like (see, generally, R. Scopes, Protein Purification, Springer--Verlag, N.Y. (1982), Deutscher, Methods in Enzymology Vol. 182: 5 Guide to Protein Purification., Academic Press, Inc. N.Y. (1990)).

As an option to recombinant methods, polypeptides and oligopeptides can be chemically synthesized. Such methods typically include solid-state approaches, but can also utilize solution based chemistries and combinations or combinations of solid-state and solution approaches. Examples of solid-state methodologies for synthesizing proteins are described by Merrifield (1964) J. 10 Am. Chem. Soc. 85:2149; and Houghton (1985) Proc. Natl. Acad. Sci., 82:5132. Fragments of a T_{BT} protein can be synthesized and then joined together. Methods for conducting such reactions are described by Grant (1992) Synthetic Peptides: A User Guide, W.H. Freeman and Co., N.Y.; and in "Principles of Peptide Synthesis," (Bodansky and Trost, ed.), Springer-Verlag, Inc. N.Y., (1993).

For various purposes, for example as an immunogen, the entire T_{BT} polypeptide or a 15 fragment derived therefrom may be used. Preferably, one or more 8-30 amino acid peptide portions, *e.g.* of an extracellular domain may be utilized, with peptides in the range of 10-20 being a more economical choice. Custom-synthesized peptides in this range are available from a multitude of vendors, and can be order conjugated to KLH or BSA. Alternatively, peptides in excess of 30 amino acids may be synthesized by solid-phase methods, or may be recombinantly produced in a suitable 20 recombinant protein production system. In order to ensure proper protein glycosylation and processing, an animal cell system (*e.g.*, Sf9 or other insect cells, CHO or other mammalian cells) is preferred.

SPECIFIC BINDING MEMBERS

25 The term "specific binding member" or "binding member" as used herein refers to a member of a specific binding pair, *i.e.* two molecules, usually two different molecules, where one of the molecules (*i.e.*, first specific binding member) through chemical or physical means specifically binds to the other molecule (*i.e.*, second specific binding member). The complementary members of a specific binding pair are sometimes referred to as a ligand and receptor; or receptor and counter- 30 receptor. For the purposes of the present invention, the two binding members may be known to associate with each other, for example where an assay is directed at detecting compounds that interfere with the association of a known binding pair. Alternatively, candidate compounds suspected of being a binding partner to a compound of interest may be used.

Specific binding pairs of interest include carbohydrates and lectins; complementary 35 nucleotide sequences; peptide ligands and receptor; effector and receptor molecules; hormones and hormone binding protein; enzyme cofactors and enzymes; enzyme inhibitors and enzymes; lipid and lipid-binding protein; etc. The specific binding pairs may include analogs, derivatives and fragments

of the original specific binding member. For example, a receptor and ligand pair may include peptide fragments, chemically synthesized peptidomimetics, labeled protein, derivatized protein, etc.

In a preferred embodiment, the specific binding member is an antibody. The term “antibody” or “antibody moiety” is intended to include any polypeptide chain-containing molecular structure with a specific shape that fits to and recognizes an epitope, where one or more non-covalent binding interactions stabilize the complex between the molecular structure and the epitope. Antibodies that bind specifically to one of the brain tumor protein targets are referred to as anti-brain tumor protein target antibodies, or α (TBT). The specific or selective fit of a given structure and its specific epitope is sometimes referred to as a “lock and key” fit. The archetypal antibody molecule is the immunoglobulin, and all types of immunoglobulins, IgG, IgM, IgA, IgE, IgD, etc., from all sources, e.g. human, rodent, rabbit, cow, sheep, pig, dog, other mammal, chicken, other avians, etc., are considered to be “antibodies.” Antibodies utilized in the present invention may be polyclonal antibodies, although monoclonal antibodies are preferred because they may be reproduced by cell culture or recombinantly, and can be modified to reduce their antigenicity.

Polyclonal antibodies can be raised by a standard protocol by injecting a production animal with an antigenic composition, formulated as described above. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an T_{BT} antigen comprising an antigenic portion of the brain tumor protein targets’ polypeptide is initially injected into any of a wide variety of mammals (e.g., mice, rats, rabbits, sheep or goats). When utilizing an entire protein, or a larger section of the protein, antibodies may be raised by immunizing the production animal with the protein and a suitable adjuvant (e.g., Freund’s, Freund’s complete, oil-in-water emulsions, etc.) When a smaller peptide is utilized, it is advantageous to conjugate the peptide with a larger molecule to make an immunostimulatory conjugate. Commonly utilized conjugate proteins that are commercially available for such use include bovine serum albumin (BSA) and keyhole limpet hemocyanin (KLH). In order to raise antibodies to particular epitopes, peptides derived from the full sequence may be utilized. Alternatively, in order to generate antibodies to relatively short peptide portions of the brain tumor protein target, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as ovalbumin, BSA or KLH. The peptide-conjugate is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Alternatively, for monoclonal antibodies, hybridomas may be formed by isolating the stimulated immune cells, such as those from the spleen of the inoculated animal. These cells are then fused to immortalized cells, such as myeloma cells or transformed cells, which are capable of replicating indefinitely in cell culture, thereby producing an immortal, immunoglobulin-secreting cell

line. The immortal cell line utilized is preferably selected to be deficient in enzymes necessary for the utilization of certain nutrients. Many such cell lines (such as myelomas) are known to those skilled in the art, and include, for example: thymidine kinase (TK) or hypoxanthine-guanine phosphoriboxyl transferase (HGPRT). These deficiencies allow selection for fused cells according to
5 their ability to grow on, for example, hypoxanthine aminopterinthymidine medium (HAT).

Preferably, the immortal fusion partners utilized are derived from a line that does not secrete immunoglobulin. The resulting fused cells, or hybridomas, are cultured under conditions that allow for the survival of fused, but not unfused, cells and the resulting colonies screened for the production of the desired monoclonal antibodies. Colonies producing such antibodies are cloned, expanded, and
10 grown so as to produce large quantities of antibody, see Kohler and Milstein, 1975 Nature 256:495 (the disclosures of which are hereby incorporated by reference).

Large quantities of monoclonal antibodies from the secreting hybridomas may then be produced by injecting the clones into the peritoneal cavity of mice and harvesting the ascites fluid therefrom. The mice, preferably primed with pristane, or some other tumor-promoter, and
15 immunosuppressed chemically or by irradiation, may be any of various suitable strains known to those in the art. The ascites fluid is harvested from the mice and the monoclonal antibody purified therefrom, for example, by CM Sepharose column or other chromatographic means. Alternatively, the hybridomas may be cultured *in vitro* or as suspension cultures. Batch, continuous culture, or other suitable culture processes may be utilized. Monoclonal antibodies are then recovered from the
20 culture medium or supernatant.

Several monoclonal antibodies against various isoforms of the brain tumor protein targets are currently available from commercial sources. For instance, a non-exclusive list of available commercial antibodies includes: for SPARC / Osteonectin, from *Zymed*, mouse anti-bovine MAb, Cat# 33-5500; for c-MET, from *Zymed*, rabbit anti-human polyclonal, and from *RDI*, rabbit anti-
25 human MAb; for CD44, from *RDI*, mouse anti-human MAb, and from *Lab vision*, mouse anti-human MAb, known to block binding of hyaluronic acid to its receptor CD44, "CD44/H-CAM Ab-2"; for Brevican/BEHAB, from *BD Transduction Lab.*, a mouse anti-human MAb; for VIP 2 receptor, from *Exalpha*, mouse anti-rat MAb; for Laminin receptor from *Lab vision*, mouse anti-human MAb "laminin receptor Ab-1"; for Osteopontin, from *Chemicon*, rat anti-human MAb, raised against rh-
30 Osteopontin. "MAB3057"; for Pleiotrophin, from *R&D* goat anti-human polyclonal recognizes rh-Pleiotrophin. "BAF252", and from *Oncogene* goat anti-human polyclonal detects rh-Pleiotrophin. "PC187L".; for PTP ζ - α and PTP ζ - β , from *BD Transduction Labs*, mouse anti-human MAb and from *Chemicon*, mouse anti-human MAb, which recognizes both of the transmembrane isoforms, and also recognizes the soluble isoform (phosphacan, PTP ζ -S). These antibodies are suitable for use in
35 the compositions of the present invention, especially in Fab fragment form (which eliminates significant portions of the antigenic constant heavy and light chain regions). However, it is preferred that such antibodies be humanized or chimerized according to one of the procedures outlined below.

In addition, the antibodies or antigen binding fragments may be produced by genetic engineering. In this technique, as with the standard hybridoma procedure, antibody-producing cells are sensitized to the desired antigen or immunogen. The messenger RNA isolated from the immune spleen cells or hybridomas is used as a template to make cDNA using PCR amplification. A library
5 of vectors, each containing one heavy chain gene and one light chain gene retaining the initial antigen specificity, is produced by insertion of appropriate sections of the amplified immunoglobulin cDNA into the expression vectors. A combinatorial library is constructed by combining the heavy chain gene library with the light chain gene library. This results in a library of clones which co-express a heavy and light chain (resembling the Fab fragment or antigen binding fragment of an
10 antibody molecule). The vectors that carry these genes are co-transfected into a host (e.g. bacteria, insect cells, mammalian cells, or other suitable protein production host cell.). When antibody gene synthesis is induced in the transfected host, the heavy and light chain proteins self-assemble to produce active antibodies that can be detected by screening with the antigen or immunogen.

Preferably, recombinant antibodies are produced in a recombinant protein production system
15 which correctly glycosylates and processes the immunoglobulin chains, such as insect or mammalian cells. An advantage to using insect cells, which utilize recombinant baculoviruses for the production of antibodies, is that the baculovirus system allows production of mutant antibodies much more rapidly than stably transfected mammalian cell lines. In addition, insect cells have been shown to correctly process and glycosylate eukaryotic proteins, which prokaryotic cells do not. Finally, the
20 baculovirus expression of foreign protein has been shown to constitute as much as 50-75% of the total cellular protein late in viral infection, making this system an excellent means of producing milligram quantities of the recombinant antibodies.

Antibodies with a reduced propensity to induce a violent or detrimental immune response in humans (such as anaphylactic shock), and which also exhibit a reduced propensity for priming an
25 immune response which would prevent repeated dosage with the antibody therapeutic or imaging agent are preferred for use in the invention. Even though the brain is relatively isolated behind the blood brain barrier, an immune response still can occur in the form of increased leukocyte infiltration, and inflammation. Although some increased immune response against the tumor is desirable, the concurrent binding and inactivation of the therapeutic or imaging agent generally
30 outweighs this benefit. Thus, humanized, chimeric, or xenogenic human antibodies, which produce less of an immune response when administered to humans, are preferred for use in the present invention.

Chimeric antibodies may be made by recombinant means by combining the murine variable light and heavy chain regions (VK and VH), obtained from a murine (or other animal-derived)
35 hybridoma clone, with the human constant light and heavy chain regions, in order to produce an antibody with predominantly human domains. The production of such chimeric antibodies is well known in the art, and may be achieved by standard means (as described, e.g., in U.S. Patent No.

5,624,659, incorporated fully herein by reference). Humanized antibodies are engineered to contain even more human-like immunoglobulin domains, and incorporate only the complementarity-determining regions of the animal-derived antibody. This is accomplished by carefully examining the sequence of the hyper-variable loops of the variable regions of the monoclonal antibody, and fitting them to the structure of the human antibody chains. Although facially complex, the process is straightforward in practice. See, e.g., U.S. Patent No. 6,187,287, incorporated fully herein by reference.

Alternatively, polyclonal or monoclonal antibodies may be produced from animals that have been genetically altered to produce human immunoglobulins. Techniques for generating such animals, and deriving antibodies therefrom, are described in U.S. Patents No. 6,162,963 and 6,150,584, incorporated fully herein by reference.

Alternatively, single chain antibodies (Fv, as described below) can be produced from phage libraries containing human variable regions. See U.S. Patent No. 6,174,708. Intrathecal administration of single-chain immunotoxin, LMB-7 [B3(Fv)- PE38], has been shown to cure of carcinomatous meningitis in a rat model. *Proc Natl. Acad. Sci U S A* 92, 2765-9, all of which are incorporated by reference fully herein.

In addition to entire immunoglobulins (or their recombinant counterparts), immunoglobulin fragments comprising the epitope binding site (e.g., Fab', F(ab')₂, or other fragments) are useful as antibody moieties in the present invention. Such antibody fragments may be generated from whole immunoglobulins by ficin, pepsin, papain, or other protease cleavage. "Fragment," or minimal immunoglobulins may be designed utilizing recombinant immunoglobulin techniques. For instance "Fv" immunoglobulins for use in the present invention may be produced by linking a variable light chain region to a variable heavy chain region via a peptide linker (e.g., poly-glycine or another sequence which does not form an alpha helix or beta sheet motif).

Fv fragments are heterodimers of the variable heavy chain domain (V_H) and the variable light chain domain (V_L). The heterodimers of heavy and light chain domains that occur in whole IgG, for example, are connected by a disulfide bond. Recombinant Fvs in which V_H and V_L are connected by a peptide linker are typically stable, see, for example, Huston et al., *Proc. Natl. Acad. Sci. USA* 85:5879-5883 (1988) and Bird et al., *Science* 242:423-426 (1988), both fully incorporated herein, by reference. These are single chain Fvs which have been found to retain specificity and affinity and have been shown to be useful for imaging tumors and to make recombinant immunotoxins for tumor therapy. However, researchers have found that some of the single chain Fvs have a reduced affinity for antigen and the peptide linker can interfere with binding. Improved Fv's have been also been made which comprise stabilizing disulfide bonds between the V_H and V_L regions, as described in U.S. Patent No. 6,147,203, incorporated fully herein by reference. Any of these minimal antibodies may be utilized in the present invention, and those which are humanized to avoid HAMA reactions are preferred for use in embodiments of the invention.

In addition, derivatized immunoglobulins with added chemical linkers, detectable moieties, such as fluorescent dyes, enzymes, substrates, chemiluminescent moieties and the like, or specific binding moieties, such as streptavidin, avidin, or biotin, and the like may be utilized in the methods and compositions of the present invention. For convenience, the term “antibody” or “antibody moiety” will be used throughout to generally refer to molecules which specifically bind to an epitope of the brain tumor protein targets, although the term will encompass all immunoglobulins, derivatives, fragments, recombinant or engineered immunoglobulins, and modified immunoglobulins, as described above.

Candidate anti-T_{BT} antibodies can be tested for anti-T_{BT} activity by any suitable standard means. As a first screen, the antibodies may be tested for binding against the immunogen, or against the entire brain tumor protein target extracellular domain or protein. As a second screen, anti-T_{BT} candidates may be tested for binding to an appropriate tumor cell line, or to primary tumor tissue samples. For these screens, the anti-T_{BT} candidate antibody may be labeled for detection. After selective binding to the brain tumor protein target is established, the candidate antibody, or an antibody conjugate produced as described below, may be tested for appropriate activity (i.e., the ability to decrease tumor cell growth and/or to aid in visualizing tumor cells) in an *in vivo* model, such as an appropriate tumor cell line, or in a mouse or rat human brain tumor model, as described below.

Antibodies that alter the biological activity of a T_{BT} protein may be assayed in functional formats, such as glioblastoma cell culture or mouse/rat CNS tumor model studies. In glioblastoma cell models of activity, expression of the protein is first verified in the particular cell strain to be used. If necessary, the cell line may be stably transfected with a coding sequence of the protein under the control of an appropriate constituent promoter, in order to express the protein at a level comparable to that found in primary tumors. The ability of the glioblastoma cells to survive in the presence of the candidate function-altering anti-protein antibody is then determined. In addition to cell-survival assays, cell migration assays may be utilized to determine the effect of the candidate antibody therapeutic agent on the tumor-like behavior of the cells. Alternatively, if the brain tumor protein target is involved in angiogenesis, assays may be utilized to determine the ability of the candidate antibody therapeutic to inhibit vascular neogenesis, an important function in tumor biology.

Similarly, *in vivo* models for human brain tumors, particularly nude mice/SCID mice model or rat models, have been described, for example see Antunes et al. (2000). *J Histochem Cytochem* 48, 847-58; Price et al. (1999) *Clin Cancer Res* 5, 845-54; and Senner et al. (2000). *Acta Neuropathol* (Berl) 99, 603-8. Once correct expression of the protein in the tumor model is verified, the effect of the candidate anti-protein antibodies on the tumor masses in these models can be evaluated, wherein the ability of the anti-protein antibody candidates to alter protein activity is indicated by a decrease in tumor growth or a reduction in the tumor mass. Thus, antibodies that exhibit the appropriate anti-

tumor effect may be selected without direct knowledge of the particular biomolecular role of the protein in oncogenesis.

Arrays

5 Arrays provide a high throughput technique that can assay a large number of polynucleotides in a sample. In one aspect of the invention, an array is constructed comprising one or more of the T_{BT} genes, proteins or antibodies, preferably comprising all of these sequences, which array may further comprise other sequences known to be up- or down-regulated in tumor cells. This technology can be used as a tool to test for differential expression. Arrays can be created by spotting
10 polynucleotide probes onto a substrate (*e.g.*, glass, nitrocellulose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Techniques for constructing arrays and methods of using these arrays are described in, for example, Schena *et al.* (1996) Proc Natl Acad Sci U S A. **93**(20):10614-9; Schena *et al.* (1995) Science **270**(5235):467-70; Shalon *et al.* (1996)
15 Genome Res. **6**(7):639-45, USPN 5,807,522, EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5,593,839; USPN 5,578,832; EP 728 520; USPN 5,599,695; EP 721 016; USPN 5,556,752; WO 95/22058; and USPN 5,631,734.

The probes utilized in the arrays can be of varying types and can include, for example, synthesized probes of relatively short length (*e.g.*, a 20-mer or a 25-mer), cDNA (full length or
20 fragments of gene), amplified DNA, fragments of DNA (generated by restriction enzymes, for example) and reverse transcribed DNA. Both custom and generic arrays can be utilized in detecting differential expression levels. Custom arrays can be prepared using probes that hybridize to particular preselected subsequences of mRNA gene sequences or amplification products prepared from them.

25 Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of T_{BT} genes, where expression is compared between a test cell and control cell. Exemplary uses of arrays are further described in, for example, Pappalarado *et al.* (1998) Sem. Radiation Oncol. **8**:217; and Ramsay. (1998) Nature Biotechnol. **16**:40. Furthermore, many variations on methods of detection
30 using arrays are well within the skill in the art and within the scope of the present invention. For example, rather than immobilizing the probe to a solid support, the test sample can be immobilized on a solid support which is then contacted with the probe. Additional discussion regarding the use of microarrays in expression analysis can be found, for example, in Duggan, *et al.*, *Nature Genetics Supplement* 21:10-14 (1999); Bowtell, *Nature Genetics Supplement* 21:25-32 (1999); Brown and
35 Botstein, *Nature Genetics Supplement* 21:33-37 (1999); Cole *et al.*, *Nature Genetics Supplement* 21:38-41 (1999); Debouck and Goodfellow, *Nature Genetics Supplement* 21:48-50 (1999); Bassett,

Jr., et al., Nature Genetics Supplement 21:51-55 (1999); and Chakravarti, Nature Genetics Supplement 21:56-60 (1999).

For detecting expression levels, usually nucleic acids are obtained from a test sample, and either directly labeled, or reversed transcribed into labeled cDNA. The test sample containing the labeled nucleic acids is then contacted with the array. After allowing a period sufficient for any labeled nucleic acid present in the sample to hybridize to the probes, the array is typically subjected to one or more high stringency washes to remove unbound nucleic acids and to minimize nonspecific binding to the nucleic acid probes of the arrays. Binding of labeled sequences is detected using any of a variety of commercially available scanners and accompanying software programs.

For example, if the nucleic acids from the sample are labeled with fluorescent labels, hybridization intensity can be determined by, for example, a scanning confocal microscope in photon counting mode. Appropriate scanning devices are described by *e.g.*, U.S. 5,578,832 to Trulson *et al.*, and U.S. 5,631,734 to Stern *et al.* and are available from Affymetrix, Inc., under the GeneChipTM label. Some types of label provide a signal that can be amplified by enzymatic methods (*see* Broude, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 91, 3072-3076 (1994)). A variety of other labels are also suitable including, for example, radioisotopes, chromophores, magnetic particles and electron dense particles.

Those locations on the probe array that are hybridized to labeled nucleic acid are detected using a reader, such as described by U.S. Patent No. 5,143,854, WO 90/15070, and U.S. 5,578,832. For customized arrays, the hybridization pattern can then be analyzed to determine the presence and/or relative amounts or absolute amounts of known mRNA species in samples being analyzed as described in *e.g.*, WO 97/10365.

Diagnostic and Prognostic Methods

The differential expression of T_{BT} genes and/or gene products in tumors indicates that these can serve as markers for diagnosis, for imaging, as well as for therapeutic applications. In general, such diagnostic methods involve detecting an elevated level of expression of T_{BT} gene transcripts or gene products in the cells or tissue of an individual or a sample therefrom. A variety of different assays can be utilized to detect an increase in gene expression, including both methods that detect gene transcript and protein levels. More specifically, the diagnostic and prognostic methods disclosed herein involve obtaining a sample from an individual and determining at least qualitatively, and preferably quantitatively, the level of a T_{BT} gene product expression in the sample. Usually this determined value or test value is compared against some type of reference or baseline value.

Nucleic acids or binding members such as antibodies that are specific for polypeptides derived from the sequence of one of the sequences provided in Table 1 are used to screen patient samples for increased expression of the corresponding mRNA or protein, or for the presence of amplified DNA in the cell. Samples can be obtained from a variety of sources. Samples are typically obtained from a human subject. However, the methods can also be utilized with samples obtained

from various other mammals, such as primates, *e.g.* apes and chimpanzees, mice, cats, rats, and other animals. Such samples are referred to as a patient sample.

Samples can be obtained from the tissues or fluids of an individual, as well as from cell cultures or tissue homogenates. For example, samples can be obtained from spinal fluid, or tumor biopsy samples. Also included in the term are derivatives and fractions of such cells and fluids. Samples can also be derived from *in vitro* cell cultures, including the growth medium, recombinant cells and cell components. Diagnostic samples are collected from an individual that has, or is suspected to have, a brain tumor. The presence of specific markers is useful in identifying and staging the tumor.

Nucleic Acid Screening Methods

Some of the diagnostic and prognostic methods that involve the detection of a T_{BT} gene transcript begin with the lysis of cells and subsequent purification of nucleic acids from other cellular material, particularly mRNA transcripts. A nucleic acid derived from an mRNA transcript refers to a nucleic acid for whose synthesis the mRNA transcript, or a subsequence thereof, has ultimately served as a template. Thus, a cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA transcribed from the amplified DNA, are all derived from the mRNA transcript and detection of such derived products is indicative of the presence and/or abundance of the original transcript in a sample.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.* upregulated or downregulated expression. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki *et al.* (1985) Science **239**:487, and a review of techniques may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2-14.33.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, *e.g.* ALEXA dyes (available from Molecular Probes, Inc.); fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2,7-dimethoxy-4,5-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2,4,7,4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N,N-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, *e.g.* ³²P, ³⁵S, ³H; *etc.* The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, *e.g.* amplified, labeled, cloned fragment, *etc.* is analyzed by one of a number of methods known in the art. Probes may be hybridized to northern or dot blots, or liquid hybridization reactions performed. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to a wild-type sequence. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

In situ hybridization methods are hybridization methods in which the cells are not lysed prior to hybridization. Because the method is performed *in situ*, it has the advantage that it is not necessary to prepare RNA from the cells. The method usually involves initially fixing test cells to a support (*e.g.* the flat surface of a microscope slide or the walls of a microtiter well) and then permeabilizing the cells with an appropriate permeabilizing solution. A solution containing labeled probes is then contacted with the cells and the probes allowed to hybridize. Excess probe is digested, washed away and the amount of hybridized probe measured. This approach is described in greater detail by Nucleic Acid Hybridization: A Practical Approach (Hames, et al., eds., 1987).

A variety of so-called "real time amplification" methods or "real time quantitative PCR" methods can also be utilized to determine the quantity of mRNA present in a sample. Such methods involve measuring the amount of amplification product formed during an amplification process. Fluorogenic nuclease assays are one specific example of a real time quantitation method that can be used to detect and quantitate transcripts. In general such assays continuously measure PCR product accumulation using a dual-labeled fluorogenic oligonucleotide probe -- an approach frequently referred to in the literature simply as the "TaqMan" method. Additional details regarding the theory and operation of fluorogenic methods for making real time determinations of the concentration of amplification products are described, for example, in U.S. Pat Nos. 5,210,015 to Gelfand, 5,538,848 to Livak, et al., and 5,863,736 to Haaland, each of which is incorporated by reference in its entirety.

Polypeptide Screening Methods

Screening for expression of the subject sequences may be based on the functional or antigenic characteristics of the protein. Various immunoassays designed to detect polymorphisms in proteins encoded by the sequences corresponding to the sequences provided in Table 1 may be used in screening. Detection may utilize staining of cells or histological sections, performed in accordance with conventional methods, using antibodies or other specific binding members that specifically bind to the T_{BT} polypeptides. The antibodies or other specific binding members of interest are added to a cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemilumescers, or other labels for direct detection. Alternatively, a second stage antibody or

reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by
5 various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.*

An alternative method for diagnosis depends on the *in vitro* detection of binding between antibodies and the polypeptide corresponding to a sequence of Table 1 in a lysate. Measuring the concentration of the target protein in a sample or fraction thereof may be accomplished by a variety
10 of specific assays. A conventional sandwich type assay may be used. For example, a sandwich assay may first attach specific antibodies to an insoluble surface or support. The particular manner of binding is not crucial so long as it is compatible with the reagents and overall methods of the invention. They may be bound to the plates covalently or non-covalently, preferably non-covalently.

The insoluble supports may be any compositions to which polypeptides can be bound, which
15 is readily separated from soluble material, and which is otherwise compatible with the overall method. The surface of such supports may be solid or porous and of any convenient shape. Examples of suitable insoluble supports to which the receptor is bound include beads, *e.g.* magnetic beads, membranes and microtiter plates. These are typically made of glass, plastic (*e.g.* polystyrene), polysaccharides, nylon or nitrocellulose. Microtiter plates are especially convenient because a large
20 number of assays can be carried out simultaneously, using small amounts of reagents and samples.

Patient sample lysates are then added to separately assayable supports (for example, separate wells of a microtiter plate) containing antibodies. Preferably, a series of standards, containing known concentrations of the test protein is assayed in parallel with the samples or aliquots thereof to serve as controls. Preferably, each sample and standard will be added to multiple wells so that mean values
25 can be obtained for each. The incubation time should be sufficient for binding. After incubation, the insoluble support is generally washed of non-bound components. After washing, a solution containing a second antibody is applied. The antibody will bind to one of the proteins of interest with sufficient specificity such that it can be distinguished from other components present. The second antibodies may be labeled to facilitate direct, or indirect quantification of binding. In a
30 preferred embodiment, the antibodies are labeled with a covalently bound enzyme capable of providing a detectable product signal after addition of suitable substrate. Examples of suitable enzymes for use in conjugates include horseradish peroxidase, alkaline phosphatase, malate dehydrogenase and the like. Where not commercially available, such antibody-enzyme conjugates are readily produced by techniques known to those skilled in the art. The incubation time should be
35 sufficient for the labeled ligand to bind available molecules.

After the second binding step, the insoluble support is again washed free of non-specifically bound material, leaving the specific complex formed between the target protein and the specific

binding member. The signal produced by the bound conjugate is detected by conventional means. Where an enzyme conjugate is used, an appropriate enzyme substrate is provided so a detectable product is formed.

Other immunoassays are known in the art and may find use as diagnostics. Ouchterlony
5 plates provide a simple determination of antibody binding. Western blots may be performed on protein gels or protein spots on filters, using a detection system specific for the targeted polypeptide, conveniently using a labeling method as described for the sandwich assay.

In some cases, a competitive assay will be used. In addition to the patient sample, a competitor to the targeted protein is added to the reaction mix. The competitor and the target
10 compete for binding to the specific binding partner. Usually, the competitor molecule will be labeled and detected as previously described, where the amount of competitor binding will be proportional to the amount of target protein present. The concentration of competitor molecule will be from about 10 times the maximum anticipated protein concentration to about equal concentration in order to make the most sensitive and linear range of detection.

15

Imaging in vivo

In some embodiments, the methods are adapted for imaging use *in vivo*, e.g., to locate or identify sites where tumor cells are present. In these embodiments, a detectably-labeled moiety, e.g., an antibody, which is specific for the T_{BT} polypeptide is administered to an individual (e.g., by
20 injection), and labeled cells are located using standard imaging techniques, including, but not limited to, magnetic resonance imaging, computed tomography scanning, and the like.

For diagnostic *in vivo* imaging, the type of detection instrument available is a major factor in selecting a given radionuclide. The radionuclide chosen must have a type of decay that is detectable by a given type of instrument. In general, any conventional method for visualizing diagnostic
25 imaging can be utilized in accordance with this invention. Another important factor in selecting a radionuclide for *in vivo* diagnosis is that its half-life be long enough that it is still detectable at the time of maximum uptake by the target tissue, but short enough that deleterious radiation of the host is minimized. A currently used method for labeling with ^{99m}Tc is the reduction of pertechnetate ion in the presence of a chelating precursor to form the labile ^{99m}Tc-precursor complex, which, in turn,
30 reacts with the metal binding group of a bifunctionally modified chemotactic peptide to form a ^{99m}Tc-chemotactic peptide conjugate.

The detectably labeled T_{BT} specific antibody is used in conjunction with imaging techniques, in order to analyze the expression of the target. In one embodiment, the imaging method is one of PET or SPECT, which are imaging techniques in which a radionuclide is synthetically or locally
35 administered to a patient. The subsequent uptake of the radiotracer is measured over time and used to obtain information about the targeted tissue. Because of the high-energy (γ-ray) emissions of the

specific isotopes employed and the sensitivity and sophistication of the instruments used to detect them, the two-dimensional distribution of radioactivity may be inferred from outside of the body.

Among the most commonly used positron-emitting nuclides in PET are included ^{11}C , ^{13}N , ^{15}O , and ^{18}F . Isotopes that decay by electron capture and/or β^+ emission are used in SPECT, and
5 include ^{123}I and $^{99\text{m}}\text{Tc}$.

Therapeutic/Prophylactic Treatment Methods

Agents that modulate activity of T_{BT} genes or proteins provide a point of therapeutic or prophylactic intervention, particularly agents that inhibit or upregulate activity of the polypeptide, or
10 expression of the gene. Numerous agents are useful in modulating this activity, including agents that directly modulate expression, *e.g.* expression vectors, antisense specific for the targeted polypeptide; and agents that act on the protein, *e.g.* specific antibodies and analogs thereof, small organic molecules that block catalytic activity, *etc.*

Methods can be designed to selectively deliver nucleic acids to certain cells. Examples of
15 such cells include, neurons, microglia, astrocytes, endothelial cells, oligodendrocytes, *etc.* Certain treatment methods are designed to selectively express an expression vector to neuron cells and/or target the nucleic acid for delivery to nerve cells (microglia, astrocytes, endothelial cells, oligodendrocytes) One technique for achieving selective expression in nerve cells is to operably link the coding sequence to a promoter that is primarily active in nerve cells. Examples of such
20 promoters include, but are not limited to, prion protein promoter, calcium-calmodulin dependent protein kinase promoter. Alternatively, or in addition, the nucleic acid can be administered with an agent that targets the nucleic acid to nerve cells. For instance, the nucleic acid can be administered with an antibody that specifically binds to a cell-surface antigen on the nerve cells or a ligand for a receptor on neuronal cells.

25 When liposomes are utilized, substrates that bind to a cell-surface membrane protein associated with endocytosis can be attached to the liposome to target the liposome to nerve cells and to facilitate uptake. Examples of proteins that can be attached include capsid proteins or fragments thereof that bind to nerve cells, antibodies that specifically bind to cell-surface proteins on nerve cells that undergo internalization in cycling and proteins that target intracellular localizations within nerve
30 cells (see, *e.g.*, Wu et al. (1987) *J. Biol. Chem.* 262:4429-4432; and Wagner, et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:3410-3414). Gene marking and gene therapy protocols are reviewed by Anderson et al. (1992) *Science* 256:808-813. Various other delivery options can also be utilized. For instance, a nucleic acid containing a sequence of interest can be injected directly into the cerebrospinal fluid. Alternatively, such nucleic acids can be administered by intraventricular
35 injections.

Antisense molecules can be used to down-regulate expression in cells. The antisense reagent may be antisense oligonucleotides (ODN), particularly synthetic ODN having chemical

modifications from native nucleic acids, or nucleic acid constructs that express such antisense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, e.g. by reducing the amount of mRNA available for translation,
5 through activation of RNase H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences.

Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such that an antisense strand is produced as an RNA molecule. Alternatively, the antisense molecule is a synthetic
10 oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and
15 selective inhibitors of gene expression (see Wagner *et al.* (1996) Nature Biotechnology 14:840-844).

A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene *in vitro* or in an animal model. A combination of sequences may also
20 be used, where several regions of the mRNA sequence are selected for antisense complementation.

Antisense oligonucleotides may be chemically synthesized by methods known in the art (see Wagner *et al.* (1993) *supra.* and Milligan *et al.*, *supra.*) Preferred oligonucleotides are chemically modified from the native phosphodiester structure, in order to increase their intracellular stability and binding affinity. A number of such modifications have been described in the literature, which alter
25 the chemistry of the backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH₂-5'-O-phosphonate and 3'-
30 NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage. Sugar modifications are also used to enhance stability and affinity. The alpha.-anomer of deoxyribose may be used, where the base is inverted with respect to the natural .beta.-anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity. Modification of the
35 heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-

propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

COMPOUND SCREENING

5 Compound screening may be performed using an *in vitro* model, a genetically altered cell or animal, or purified protein corresponding to any one of the provided T_{BT} genes. One can identify ligands or substrates that bind to, modulate or mimic the action of the encoded polypeptide.

 The polypeptides include those encoded by T_{BT} genes, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed nucleic acids, and
10 variants thereof. Variant polypeptides can include amino acid (aa) substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine
15 residues that are not necessary for function. Variants can be designed so as to retain or have enhanced biological activity of a particular region of the protein (*e.g.*, a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus
20 sequence). Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50
25 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 500 aa in length, where the fragment will have a contiguous stretch of amino acids that is identical to a polypeptide encoded by brain tumor associated genes, or a homolog thereof.

 Transgenic animals or cells derived therefrom are also used in compound screening. Transgenic animals may be made through homologous recombination, where the normal locus
25 corresponding to a T_{BT} gene is altered. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. A series of small deletions and/or substitutions may be made in the coding sequence to determine the role of different exons in enzymatic activity, oncogenesis, signal transduction, *etc.* Specific constructs of interest include antisense sequences that block expression of
30 the targeted gene and expression of dominant negative mutations. A detectable marker, such as lac Z may be introduced into the locus of interest, where up-regulation of expression will result in an easily detected change in phenotype. One may also provide for expression of the target gene or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development. By providing expression of the target protein in cells in which it is not normally produced, one can
35 induce changes in cell behavior.

 Compound screening identifies agents that modulate function of the T_{BT} polypeptides. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide

variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. Knowledge of the 3-dimensional structure of the encoded protein, derived from crystallization of purified recombinant protein, could lead to the rational design of small drugs that specifically inhibit activity. These drugs may be directed at specific domains.

The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of altering or mimicking the physiological function of a T_{BT} polypeptide. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs. Test agents can be obtained from libraries, such as natural product libraries or combinatorial libraries, for example. A number of different types of combinatorial libraries and methods for preparing such libraries have been described, including for example, PCT publications WO 93/06121, WO 95/12608, WO 95/35503, WO 94/08051 and WO 95/30642, each of which is incorporated herein by reference.

Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemilumescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and

streptavidin, digoxin and antidigoxin, *etc.* For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents
5 like salts, neutral proteins, *e.g.* albumin, detergents, etc that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40° C.
10 Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

Preliminary screens can be conducted by screening for compounds capable of binding to a T_{BT} polypeptide, as at least some of the compounds so identified are likely inhibitors. The binding assays usually involve contacting a T_{BT} polypeptide with one or more test compounds and allowing
15 sufficient time for the protein and test compounds to form a binding complex. Any binding complexes formed can be detected using any of a number of established analytical techniques. Protein binding assays include, but are not limited to, methods that measure co-precipitation, co-migration on non-denaturing SDS-polyacrylamide gels, and co-migration on Western blots (see, *e.g.*, Bennet, J.P. and Yamamura, H.I. (1985) "Neurotransmitter, Hormone or Drug Receptor Binding
20 Methods," in *Neurotransmitter Receptor Binding* (Yamamura, H. I., et al., eds.), pp. 61-89.

Certain screening methods involve screening for a compound that modulates the expression of a T_{BT} gene. Such methods generally involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing a T_{BT} gene and then detecting and an increase in expression. Some assays are performed with tumor cells that express endogenous T_{BT}
25 genes. Other expression assays are conducted with non-neuronal cells that express an exogenous T_{BT} gene.

The level of expression or activity can be compared to a baseline value. As indicated above, the baseline value can be a value for a control sample or a statistical value that is representative of expression levels for a control population. Expression levels can also be determined for cells that do
30 not express T_{BT} gene, as a negative control. Such cells generally are otherwise substantially genetically the same as the test cells. Various controls can be conducted to ensure that an observed activity is authentic including running parallel reactions with cells that lack the reporter construct or by not contacting a cell harboring the reporter construct with test compound. Compounds can also be further validated as described below.

35 Compounds that are initially identified by any of the foregoing screening methods can be further tested to validate the apparent activity. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model

for humans and then determining if T_{BT} gene is in fact upregulated. The animal models utilized in validation studies generally are mammals. Specific examples of suitable animals include, but are not limited to, primates, mice, and rats.

Active test agents identified by the screening methods described herein that inhibit T_{BT} polypeptide activity and/or tumor growth can serve as lead compounds for the synthesis of analog compounds. Typically, the analog compounds are synthesized to have an electronic configuration and a molecular conformation similar to that of the lead compound. Identification of analog compounds can be performed through use of techniques such as self-consistent field (SCF) analysis, configuration interaction (CI) analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are available. See, e.g., Rein et al., (1989) Computer-Assisted Modeling of Receptor-Ligand Interactions (Alan Liss, New York).

Antibody Conjugates

The anti-T_{BT} antibodies for use in the present invention may have utility without conjugation when the native activity of the brain tumor protein target is altered in the tumor cell. Such antibodies, which may be selected as described above, may be utilized without as a therapeutic agent. In another embodiment of the invention, T_{BT} specific antibodies, which may or may not alter the activity of the target polypeptide, are conjugated to cytotoxic or imaging agents, which add functionality to the antibody.

The anti-T_{BT} antibodies can be coupled or conjugated to one or more therapeutic cytotoxic or imaging moieties. As used herein, "cytotoxic moiety" is a moiety that inhibits cell growth or promotes cell death when proximate to or absorbed by the cell. Suitable cytotoxic moieties in this regard include radioactive isotopes (radionuclides), chemotoxic agents such as differentiation inducers and small chemotoxic drugs, toxin proteins, and derivatives thereof. "Imaging moiety" (I) is a moiety that can be utilized to increase contrast between a tumor and the surrounding healthy tissue in a visualization technique (e.g., radiography, positron-emission tomography, magnetic resonance imaging, direct or indirect visual inspection). Thus, suitable imaging moieties include radiography moieties (e.g. heavy metals and radiation emitting moieties), positron emitting moieties, magnetic resonance contrast moieties, and optically visible moieties (e.g., fluorescent or visible-spectrum dyes, visible particles, etc.). It will be appreciated by one of ordinary skill that some overlap exists between therapeutic and imaging moieties. For instance ²¹²Pb and ²¹²Bi are both useful radioisotopes for therapeutic compositions, but are also electron-dense, and thus provide contrast for X-ray radiographic imaging techniques, and can also be utilized in scintillation imaging techniques.

In general, therapeutic or imaging agents may be conjugated to the anti-T_{BT} moiety by any suitable technique, with appropriate consideration of the need for pharmacokinetic stability and reduced overall toxicity to the patient. A therapeutic agent may be coupled to a suitable antibody moiety either directly or indirectly (e.g. via a linker group). A direct reaction between an agent and an

antibody is possible when each possesses a functional group capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (e.g., a halide). Alternatively, a suitable chemical linker group may be used. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on a moiety or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of moieties, or functional groups on moieties, which otherwise would not be possible.

Suitable linkage chemistries include maleimidyl linkers and alkyl halide linkers (which react with a sulfhydryl on the antibody moiety) and succinimidyl linkers (which react with a primary amine on the antibody moiety). Several primary amine and sulfhydryl groups are present on immunoglobulins, and additional groups may be designed into recombinant immunoglobulin molecules. It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, Ill.), may be employed as a linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, e.g., U.S. Patent No. 4,671,958. As an alternative coupling method, cytotoxic or imaging moieties may be coupled to the anti-T_{BT} antibody moiety through a an oxidized carbohydrate group at a glycosylation site, as described in U.S. Patents No. 5,057,313 and 5,156,840. Yet another alternative method of coupling the antibody moiety to the cytotoxic or imaging moiety is by the use of a non-covalent binding pair, such as streptavidin/biotin, or avidin/biotin. In these embodiments, one member of the pair is covalently coupled to the antibody moiety and the other member of the binding pair is covalently coupled to the cytotoxic or imaging moiety.

Where a cytotoxic moiety is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group that is cleavable during or upon internalization into a cell, or which is gradually cleavable over time in the extracellular environment. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of a cytotoxic moiety agent from these linker groups include cleavage by reduction of a disulfide bond (e.g., U.S. Patent No. 4,489,710), by irradiation of a photolabile bond (e.g., U.S. Patent No. 4,625,014), by hydrolysis of derivatized amino acid side chains (e.g., U.S. Patent No. 4,638,045), by serum complement-mediated hydrolysis (e.g., U.S. Patent No. 4,671,958), and acid-catalyzed hydrolysis (e.g., U.S. Patent No. 4,569,789).

Two or more cytotoxic and/or imaging moieties may be conjugated to an antibody, where the conjugated moieties are the same or different. By poly-derivatizing the anti-T_{BT} antibody, several cytotoxic strategies can be simultaneously implemented; an antibody may be made useful as a

contrasting agent for several visualization techniques; or a therapeutic antibody may be labeled for tracking by a visualization technique. Immunoconjugates with more than one moiety may be prepared in a variety of ways. For example, more than one moiety may be coupled directly to an antibody molecule, or linkers which provide multiple sites for attachment (e.g., dendrimers) can be used. Alternatively, a carrier with the capacity to hold more than one cytotoxic or imaging moiety can be used.

A carrier may bear the cytotoxic or imaging moiety in a variety of ways, including covalent bonding either directly or via a linker group, and non-covalent associations. Suitable covalent-bond carriers include proteins such as albumins (e.g., U.S. Patent No. 4,507,234), peptides, and polysaccharides such as aminodextran (e.g., U.S. Patent No. 4,699,784), each of which have multiple sites for the attachment of moieties. A carrier may also bear an agent by non-covalent associations, such as non-covalent bonding or by encapsulation, such as within a liposome vesicle (e.g., U.S. Patents Nos. 4,429,008 and 4,873,088). Encapsulation carriers are especially useful for imaging moiety conjugation to anti-T_{BT} antibody moieties for use in the invention, as a sufficient amount of the imaging moiety (dye, magnetic resonance contrast reagent, etc.) for detection may be more easily associated with the antibody moiety. In addition, encapsulation carriers are also useful in chemotoxic therapeutic embodiments, as they can allow the therapeutic compositions to gradually release a chemotoxic moiety over time while concentrating it in the vicinity of the tumor cells.

Carriers and linkers specific for radionuclide agents (both for use as cytotoxic moieties or positron-emission imaging moieties) include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis. Such chelation carriers are also useful for magnetic spin contrast ions for use in magnetic resonance imaging tumor visualization methods, and for the chelation of heavy metal ions for use in radiographic visualization methods.

Preferred radionuclides for use as cytotoxic moieties are radionuclides that are suitable for pharmacological administration. Such radionuclides include ¹²³I, ¹²⁵I, ¹³¹I, ⁹⁰Y, ²¹¹At, ⁶⁷Cu, ¹⁸⁶Re, ¹⁸⁸Re, ²¹²Pb, and ²¹²Bi. Iodine and astatine isotopes are more preferred radionuclides for use in the therapeutic compositions of the present invention, as a large body of literature has been accumulated regarding their use. ¹³¹I is particularly preferred, as are other β-radiation emitting nuclides, which have an effective range of several millimeters. ¹²³I, ¹²⁵I, ¹³¹I, or ²¹¹At may be conjugated to antibody moieties for use in the compositions and methods utilizing any of several known conjugation reagents, including Iodogen, *N*-succinimidyl 3-[²¹¹At]astatobenzoate, *N*-succinimidyl 3-[¹³¹I]iodobenzoate (SIB), and , *N*-succinimidyl 5-[¹³¹I]iodob-3-pyridinecarboxylate (SIPC). Any iodine isotope may be utilized in the recited iodo-reagents. Radionuclides can be conjugated to anti-

T_{BT} antibody moieties by suitable chelation agents known to those of skill in the nuclear medicine arts.

Preferred chemotoxic agents include small-molecule drugs such as carboplatin, cisplatin, vincristine, taxanes such as paclitaxel and docetaxel, hydroxyurea, gemcitabine, vinorelbine, irinotecan, tirapazamine, matrilysin, methotrexate, pyrimidine and purine analogs, and other suitable small toxins known in the art. Preferred chemotoxin differentiation inducers include phorbol esters and butyric acid. Chemotoxic moieties may be directly conjugated to the anti-T_{BT} antibody moiety via a chemical linker, or may be encapsulated in a carrier, which is in turn coupled to the anti-T_{BT} antibody moiety.

Chemotherapy is helpful in controlling high-grade gliomas. A common combination of chemotherapeutics is "PCV", which refers to the three drugs: Procarbazine, CCNU, and Vincristine. Temozolomide (Temodar) is approved by the FDA for treatment of anaplastic astrocytoma, and this drug is now widely used for high-grade gliomas. Neupogen may be administered to patients whose white blood counts fall to very low levels after chemotherapy.

Preferred toxin proteins for use as cytotoxic moieties include ricins A and B, abrin, diphtheria toxin, bryodin 1 and 2, momordin, trichokirin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, pokeweed antiviral protein, and other toxin proteins known in the medicinal biochemistry arts. The nontoxic ricin B chain is the moiety that binds to cells while the A chain is the toxic portion that inactivates protein synthesis- but only after delivery to the cytoplasm by the disulfide-linked B chain which binds to galactose-terminal membrane proteins. Abrin, diphtheria toxin, and Pseudomonas exotoxins all have similar 2-chain components; with one chain mediating cell membrane binding and entry and the toxic enzymatic A chain. Cholera has a pentameric binding subunit coupled to the toxic A chain. As these toxin agents may elicit undesirable immune responses in the patient, especially if injected intravascularly, it is preferred that they be encapsulated in a carrier for coupling to the anti-T_{BT} antibody moiety.

Preferred radiographic moieties for use as imaging moieties in the present invention include compounds and chelates with relatively large atoms, such as gold, iridium, technetium, barium, thallium, iodine, and their isotopes. It is preferred that less toxic radiographic imaging moieties, such as iodine or iodine isotopes, be utilized in the compositions and methods of the invention. Examples of such compositions which may be utilized for x-ray radiography are described in U.S. Patent No. 5,709,846, incorporated fully herein by reference. Such moieties may be conjugated to the anti-T_{BT} antibody moiety through an acceptable chemical linker or chelation carrier. In addition, radionuclides which emit radiation capable of penetrating the skull may be useful for scintillation imaging techniques. Suitable radionuclides for conjugation include ⁹⁹Tc, ¹¹¹In, and ⁶⁷Ga. Positron emitting moieties for use in the present invention include ¹⁸F, which can be easily conjugated by a fluorination reaction with the anti-T_{BT} antibody moiety according to the method described in U.S. Patent No. 6,187,284.

Preferred magnetic resonance contrast moieties include chelates of chromium(III), manganese(II), iron(II), nickel(II), copper(II), praseodymium(III), neodymium(III), samarium(III) and ytterbium(III) ion. Because of their very strong magnetic moment, the gadolinium(III), terbium(III), dysprosium(III), holmium(III), erbium(III), and iron(III) ions are especially preferred.

5 Examples of such chelates, suitable for magnetic resonance spin imaging, are described in U.S. Patent No. 5,733,522, incorporated fully herein by reference. Nuclear spin contrast chelates may be conjugated to the anti-T_{BT} antibody moieties through a suitable chemical linker.

Optically visible moieties for use as imaging moieties include fluorescent dyes, or visible-spectrum dyes, visible particles, and other visible labeling moieties. Fluorescent dyes such as
10 ALEXA dyes, fluorescein, coumarin, rhodamine, bodipy Texas red, and cyanine dyes, are useful when sufficient excitation energy can be provided to the site to be inspected visually. Endoscopic visualization procedures may be more compatible with the use of such labels. For many procedures where imaging agents are useful, such as during an operation to resect a brain tumor, visible spectrum dyes are preferred. Acceptable dyes include FDA-approved food dyes and colors, which
15 are non-toxic, although pharmaceutically acceptable dyes which have been approved for internal administration are preferred. In preferred embodiments, such dyes are encapsulated in carrier moieties, which are in turn conjugated to the anti-T_{BT} antibody. Alternatively, visible particles, such as colloidal gold particles or latex particles, may be coupled to the anti-T_{BT} antibody moiety via a suitable chemical linker.

20

Pharmaceutical Formulations

One strategy for drug delivery through the blood brain barrier (BBB) entails disruption of the BBB, either by osmotic means such as mannitol or leukotrienes, or biochemically by the use of vasoactive substances such as bradykinin. The potential for using BBB opening to target specific
25 agents to brain tumors is also an option. A BBB disrupting agent can be co-administered with the therapeutic or imaging compositions of the invention when the compositions are administered by intravascular injection. Other strategies to go through the BBB may entail the use of endogenous transport systems, including carrier-mediated transporters such as glucose and amino acid carriers, receptor-mediated transcytosis for insulin or transferrin, and active efflux transporters such as p-
30 glycoprotein. Active transport moieties may also be conjugated to the therapeutic or imaging compounds for use in the invention to facilitate transport across the epithelial wall of the blood vessel. Alternatively, drug delivery behind the BBB is by intrathecal delivery of therapeutics or imaging agents directly to the cranium, as through an Ommaya reservoir.

Pharmaceutical compositions can include, depending on the formulation desired,
35 pharmaceutically-acceptable, non-toxic carriers of diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are

distilled water, buffered water, physiological saline, PBS, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation can include other carriers, adjuvants, or non-toxic, nontherapeutic, nonimmunogenic stabilizers, excipients and the like. The compositions can also include additional substances to approximate physiological conditions,
5 such as pH adjusting and buffering agents, toxicity adjusting agents, wetting agents and detergents.

The composition can also include any of a variety of stabilizing agents, such as an antioxidant for example. When the pharmaceutical composition includes a polypeptide, the polypeptide can be complexed with various well-known compounds that enhance the *in vivo* stability of the polypeptide, or otherwise enhance its pharmacological properties (*e.g.*, increase the half-life of
10 the polypeptide, reduce its toxicity, enhance solubility or uptake). Examples of such modifications or complexing agents include sulfate, gluconate, citrate and phosphate. The polypeptides of a composition can also be complexed with molecules that enhance their *in vivo* attributes. Such molecules include, for example, carbohydrates, polyamines, amino acids, other peptides, ions (*e.g.*, sodium, potassium, calcium, magnesium, manganese), and lipids.

15 Further guidance regarding formulations that are suitable for various types of administration can be found in Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, PA, 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, Science 249:1527-1533 (1990).

The pharmaceutical compositions can be administered for prophylactic and/or therapeutic
20 treatments. Toxicity and therapeutic efficacy of the active ingredient can be determined according to standard pharmaceutical procedures in cell cultures and/or experimental animals, including, for example, determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that
25 exhibit large therapeutic indices are preferred.

The data obtained from cell culture and/or animal studies can be used in formulating a range of dosages for humans. The dosage of the active ingredient typically lines within a range of circulating concentrations that include the ED₅₀ with low toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized.

30 The pharmaceutical compositions described herein can be administered in a variety of different ways. Examples include administering a composition containing a pharmaceutically acceptable carrier via oral, intranasal, rectal, topical, intraperitoneal, intravenous, intramuscular, subcutaneous, subdermal, transdermal, intrathecal, and intracranial methods.

For oral administration, the active ingredient can be administered in solid dosage forms, such
35 as capsules, tablets, and powders, or in liquid dosage forms, such as elixirs, syrups, and suspensions. The active component(s) can be encapsulated in gelatin capsules together with inactive ingredients and powdered carriers, such as glucose, lactose, sucrose, mannitol, starch, cellulose or cellulose

derivatives, magnesium stearate, stearic acid, sodium saccharin, talcum, magnesium carbonate. Examples of additional inactive ingredients that may be added to provide desirable color, taste, stability, buffering capacity, dispersion or other known desirable features are red iron oxide, silica gel, sodium lauryl sulfate, titanium dioxide, and edible white ink. Similar diluents can be used to make compressed tablets. Both tablets and capsules can be manufactured as sustained release products to provide for continuous release of medication over a period of hours. Compressed tablets can be sugar coated or film coated to mask any unpleasant taste and protect the tablet from the atmosphere, or enteric-coated for selective disintegration in the gastrointestinal tract. Liquid dosage forms for oral administration can contain coloring and flavoring to increase patient acceptance.

The active ingredient, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen.

Formulations suitable for parenteral administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives.

The components used to formulate the pharmaceutical compositions are preferably of high purity and are substantially free of potentially harmful contaminants (e.g., at least National Food (NF) grade, generally at least analytical grade, and more typically at least pharmaceutical grade). Moreover, compositions intended for *in vivo* use are usually sterile. To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions for parental administration are also sterile, substantially isotonic and made under GMP conditions.

The compositions of the invention may be administered using any medically appropriate procedure, e.g., intravascular (intravenous, intraarterial, intracapillary) administration, injection into the cerebrospinal fluid, intracavity or direct injection in the tumor. Intrathecal administration may be carried out through the use of an Ommaya reservoir, in accordance with known techniques. (F. Balis et al., Am J. Pediatr. Hematol. Oncol. 11, 74, 76 (1989). For the imaging compositions of the invention, administration via intravascular injection is preferred for pre-operative visualization of the tumor. Post-operative visualization or visualization concurrent with an operation may be through intrathecal or intracavity administration, as through an Ommaya reservoir, or also by intravascular administration.

One method for administration of the therapeutic compositions of the invention is by deposition into the inner cavity of a cystic tumor by any suitable technique, such as by direct injection (aided by stereotaxic positioning of an injection syringe, if necessary) or by placing the tip of an Ommaya reservoir into a cavity, or cyst, for administration. Where the tumor is a solid tumor, the antibody may be administered by first creating a resection cavity in the location of the tumor. This procedure differs from an ordinary craniotomy and tumor resection only in a few minor respects. As tumor resection is a common treatment procedure, and is often indicated to relieve pressure, administration of the therapeutic compositions of the invention can be performed following tumor resection. Following gross total resection in a standard neurosurgical fashion, the cavity is preferable rinsed with saline until all bleeding is stopped by cauterization. Next the pia-arachnoid membrane, surrounding the tumor cavity at the surface, is cauterized to enhance the formation of fibroblastic reaction and scarring in the pia-arachnoid area. The result is the formation of an enclosed, fluid-filled cavity within the brain tissue at the location from where the tumor was removed. After the cyst has been formed, either the tip of an Ommaya reservoir or a micro catheter, which is connected to a pump device and allows the continuous infusion of an antibody solution into the cavity, can be placed into the cavity. See, e.g., U.S. Patent No. 5,558,852, incorporated fully herein by reference.

Alternatively, a convection-enhanced delivery catheter may be implanted directly into the tumor mass, into a natural or surgically created cyst, or into the normal brain mass. Such convection-enhanced pharmaceutical composition delivery devices greatly improve the diffusion of the composition throughout the brain mass. The implanted catheters of these delivery devices utilize high-flow microinfusion (with flow rates in the range of about 0.5 to 15.0 $\mu\text{l}/\text{minute}$), rather than diffusive flow, to deliver the therapeutic or imaging composition to the brain and/or tumor mass. Such devices are described in U.S. Patent No. 5,720,720, incorporated fully herein by reference.

The effective amount of a therapeutic composition to be given to a particular patient will depend on a variety of factors, several of which will be different from patient to patient. A competent clinician will be able to determine an effective amount of a therapeutic agent to administer to a patient to retard the growth and promote the death of tumor cells, or an effective amount of an imaging composition to administer to a patient to facilitate the visualization of a tumor. Dosage of the antibody-conjugate will depend on the treatment of the tumor, route of administration, the nature of the therapeutics, sensitivity of the tumor to the therapeutics, etc. Utilizing LD_{50} animal data, and other information available for the conjugated cytotoxic or imaging moiety, a clinician can determine the maximum safe dose for an individual, depending on the route of administration. For instance, an intravenously administered dose may be more than an intrathecally administered dose, given the greater body of fluid into which the therapeutic composition is being administered. Similarly, compositions which are rapidly cleared from the body may be administered at higher doses, or in repeated doses, in order to maintain a therapeutic concentration. Imaging moieties are typically less

toxic than cytotoxic moieties and may be administered in higher doses in some embodiments. Utilizing ordinary skill, the competent clinician will be able to optimize the dosage of a particular therapeutic or imaging composition in the course of routine clinical trials.

Typically the dosage will be 0.001 to 100 milligrams of conjugate per kilogram subject body weight. Doses in the range of 0.01 to 1 mg per kilogram of patient body weight may be utilized for a radionuclide therapeutic composition which is administered intrathecally. Relatively large doses, in the range of 0.1 to 10 mg per kilogram of patient body weight, may be used for imaging conjugates with a relatively non-toxic imaging moiety. The amount utilized will depend on the sensitivity of the imaging method, and the relative toxicity of the imaging moiety. In a therapeutic example, for example where the therapeutic composition comprises a ^{131}I cytotoxic moiety, the dosage to the patient will typically start at a lower range of 10 mCi, and go up to 100, 300 or even 500 mCi. Stated otherwise, where the therapeutic agent is ^{131}I , the dosage to the patient will typically be from 5,000 Rads to 100,000 Rads (preferably at least 13,000 Rads, or even at least 50,000 Rads). Doses for other radionuclides are typically selected so that the tumoricidal dose will be equivalent to the foregoing range for ^{131}I . Similarly, chemotoxic or toxin protein doses may be scaled accordingly.

The compositions can be administered to the subject in a series of more than one administration. For therapeutic compositions, regular periodic administration (e.g., every 2-3 days) will sometimes be required, or may be desirable to reduce toxicity. For therapeutic compositions which will be utilized in repeated-dose regimens, antibody moieties which do not provoke immune responses are preferred. The imaging antibody conjugate compositions may be administered at an appropriate time before the visualization technique. For example, administration within an hour before direct visual inspection may be appropriate, or administration within twelve hours before an MRI scan may be appropriate. Care should be taken, however, to not allow too much time to pass between administration and visualization, as the imaging compound may eventually be cleared from the patient's system.

In addition to the use of imaging antibody conjugates for simple visualization, these compositions may be utilized as a "dry run" for more toxic cytotoxic antibody conjugates. If the same antibody moiety is utilized for the imaging conjugate as for the therapeutic conjugate, the physician may first use a visualization technique to determine precisely where in the brain the cytotoxic conjugate will concentrate. If a sufficient degree of tissue selectivity is not achieved (e.g., if the tumor cells are too dispersed in the normal tissue, or if the particular brain tumor protein target chosen is not sufficiently overexpressed in the particular patient's tumor cells), then the physician may choose another brain tumor protein target. The provision of numerous brain tumor protein targets by the present invention, along with both imaging and therapeutic agents, allows a high degree of flexibility in designing an effective treatment regimen for the individual patient.

Combination Therapies

Brain tumors tend to be heterogeneous in character, and pervasive throughout the brain tissue. This combination often makes them difficult to treat. In some cases, it may be preferred to use various combinations of therapeutic or imaging agents, in order to more fully target all of the cells exhibiting tumorigenic characteristics. Such combination treatments may be by administering
5 blended antibody therapeutic or imaging compositions, individually prepared as described above, and administering the blended therapeutic to the patient as described. The skilled administering physician will be able to take such factors as combined toxicity, and individual agent efficacy, into account when administering such combined agents. Additionally, those of skill in the art will be able to screen for potential cross-reaction with each other, in order to assure full efficacy of each agent.

10 Alternatively, several individual brain tumor protein target compositions may be administered simultaneously or in succession for a combined therapy. This may be desirable to avoid accumulated toxicity from several antibody conjugate reagents, or to more closely monitor potential adverse reactions to the individual antibody reagents. Thus, cycles such as where a first antibody therapeutic agent is administered on day one, followed by a second on day two, then a period with
15 out administration, followed by re-administration of the antibody therapeutics on different successive days, is comprehended within the present invention.

EXPERIMENTAL

The following examples are put forth so as to provide those of ordinary skill in the art with a
20 complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.*, amounts, temperature, *etc.*) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are
25 parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

30 The present invention has been described in terms of particular embodiments found or proposed by the present inventor to comprise preferred modes for the practice of the invention. It will be appreciated by those of skill in the art that, in light of the present disclosure, numerous modifications and changes can be made in the particular embodiments exemplified without departing from the intended scope of the invention. For example, due to codon redundancy, changes can be
35 made in the underlying DNA sequence without affecting the protein sequence. Moreover, due to biological functional equivalency considerations, changes can be made in protein structure without

affecting the biological action in kind or amount. All such modifications are intended to be included within the scope of the appended claims.

Example 1

5 Identification of Differentially Expressed Sequences

Brain Tumors: Tumor tissue, confirmed as glioblastoma grade IV by neuropathology, from an unknown patient was snap frozen in the operation hall and served as experimental sample. Human whole brain tissue (Clontech Laboratories, Palo Alto, USA) served as control sample. Poly-A⁺ RNA prepared from the cells was converted into double-stranded cDNA (dscDNA) and
 10 normalized. Subtractive hybridization was carried out using the dscDNA from tumors with an excess of dscDNA prepared from the same region of a non-cancerous brain. Differentially expressed gene fragments were cloned into a plasmid vector, and the resulting library was transformed into *E. coli* cells. Inserts of recombinant clones were amplified by the polymerase chain reaction (PCR). The PCR products (fragments of 200-2000 bp in size) were sequenced using an oligonucleotide
 15 complementary to common vector sequences. The resulting sequence information was compared to public databases using the BLAST (blastn) and Smith Waterman algorithms. The differentially expressed sequences thus identified are listed in Table 1.

Example 2:

20 Identification of Two New Splicing Variant Isoforms of PTP ζ :PTP ζ SM1 and SM2

The mRNA nucleotide sequence for PTP ζ SM1 was identified in a human fetal brain phage cDNA library, and the mRNA nucleotide sequence for PTP ζ SM2 was identified by PCR amplification of adult human brain cDNA.

25 For the RT-PCR analyses performed below, total RNA was isolated from either cells (glioblastoma cultured lines) or tissue using Trizol (Gibco Life Technologies, Inc.), following the manufacture's protocol. cDNA was generated from total RNA using the 1st Strand synthesis kit from Gibco Life Technologies, Inc., and an oligo dT₃₀ anchored primer. For each RT-PCR reaction, 1 μ l of cDNA was utilized. The PCR reaction was carried out using an Advantage 2 kit (Clontech) under
 30 standard conditions. The products of the PCR reactions were confirmed via sequencing.

Both clones were verified by RT-PCR analysis of glioblastoma cell lines and primary tumors. For PTP ζ SM1, primers (SEQ ID NO:17) CAGCAGTTGGATGGAAGAGGAC and (SEQ ID NO:18) CACTGAGATT CTGGCACTATTC were used, producing an identifiable 1116 bp product. RT-PCR analysis was performed, confirming expression of the SM1 splice variant in 11 of
 35 17 different glioblastoma cell lines tested, fetal brain, and adult brain, using the unique 3' end and portion of the 3' untranslated region as the hybridization target for the probe. In addition, RT-PCR

analysis was performed on 28 primary brain tumor samples, confirming expression of the PTP ζ SM1 variant in 16 of the 28 tumors.

For PTP ζ SM2, primers (SEQ ID NO:19) AACAAATTCAGGGTCTCACTC and (SEQ ID NO:20) TTGACTGGCTCAGGAGTATAG were used, which produce a 130 bp product when the
5 extra exon 23a is present, and a no product when the exon 23a is absent. RT-PCR analysis was performed, confirming expression in 6 of 17 different glioblastoma cell lines tested. In addition, RT-PCR analysis was performed on 28 primary brain tumor samples, confirming expression of the PTP ζ SM1 variant in 19 of the 28 tumors.

For comparison, RT-PCR analysis was also done for the expression of PTP ζ - α (primers
10 (SEQ ID NO:21) CTGATAATGAGGGCTCCCAAC and CTCTGCACTTCCTGGTAAACTCT) and PTP ζ - β (primers (SEQ ID NO:22) CAGCAGTTGGATGGAAGAGGAC and (SEQ ID NO:23) CTCTGCACTTCCTGGT AAAACTCT) in the 28 brain tumor tissue samples. PTP ζ - α was shown to be expressed in 16 of the 28 samples, and the short form PTP ζ - β was shown to be expressed in 19 of the 28 samples.

15 The nucleotide sequence alignment of the two new splice variants with the reference sequence for PTP ζ - α is shown in Table 2.

TABLE 2

PTP 5'	PTP 3'	PAC 1 5'	PAC 1 3'	Corresponding Exon Key:
1	48	87274	87321	5' UTR PAC 1: RP5-1062J16
70	205	87343	87487	1 BAC: RP11-384A20
205	272	142076	142143	2 PAC 2: RP5-1049N15
		BAC 5'	BAC 3'	
291	451	24001	24161	3 * 88 nt deletion seen in 5' PCR clone from PTP 363-451
450	603	28570	28723	4
602	701	32814	32888	5
698	772	32814	32888	6
766	924	39695	39853	7
922	1075	39995	40148	8
1074	1261	52411	52598	9 * not spliced at 1261 in phage library clones
1260	1387	53910	54037	10
1387	1435	60644	60692	11
1432	2346	66362	67276	5' 12 (end of BAC)
		PAC 2 5'	PAC 2 3'	
2147	4409	1	2263	mid 12
4437	4987	2294	2844	3' 12
4925	5133	8027	8224	13
5131	5224	17505	17598	14
5223	5310	20427	20514	15
5309	5332	23048	23071	16
5329	5428	23234	23333	17
5429	5512	25555	25638	18
5512	5646	27710	27844	19

5572	5602	42925	42955	Duplicate of mid 19 * duplicated regions of exons 19
5646	5768	28408	28530	most of 20 (-12 bp 3') and 26 vary by one aa / two nt
5791	5945	29770	29934	21 (-10 bp 5')
5943	6082	31560	31699	22
6080	6228	33375	33523	23 ~116 nt insert seen b/w exons 23 & 24 in 3' PCR clone: maps to PAC b/w 23 & 24
6225	6322	40379	40476	24 PTP location PAC 2 5' PAC 2 3'
6322	6397	40820	40895	25 6228 36744 36629
6396	6526	42864	42994	26
6457	6487	27770	27800	Duplicate of mid 26
6525	6673	43895	44043	27
6671	6816	47753	47898	28
6816	6952	48708	48844	29

BOUNDARIES DETERMINED FROM HOMO SAPIENS CHROMOSOME 7 WORKING DRAFT (NT_007845.3)Nucleotide location refers to position in full length RPTPZ (accession M93426)

Example 3

Cell Migration Assay For Determining Antibody Activity on Protein Targets

5 Tumor cells are known to migrate more rapidly towards chemoattractants, and the ability to migrate is taken as a measure of tumorigenicity. Chemoattractants generally used are fetal bovine serum, pleiotrophin, bFGF, and VEGF. This assay is used to determine migration capability of a cell in which the target gene has been knocked down or overexpressed.

10 The ChemoTx[®] disposable chemotaxis system (Neuroprobe, Inc., Gaithersburg, MD) is used according to the manufacturer's instructions, with a few modifications. Briefly, glioblastoma cultured cells from cell line G55T2 are prepared by splitting the cells the day before the assay is performed. A ChemoTx[®] chamber with the following specifications is used: Pore size 8 μm , exposed filter area 8 mm^2 , exposed filter area diameter 3.2 mm. The plate configuration is: 30 μl per well, 96 well plate. The membrane type is: Track-etched polycarbonate.

15 In preparation for the assays, the filter membrane is coated in 100ml PBS containing 0.1% acetic acid and 3.5 ml Vitrogen 100 (from Cohesion) at 37°C overnight. About 30 minutes before starting the assay the coated membrane is washed and rinsed with PBS containing 0.1% BSA. Cells are harvested by using the standard technique (trypsin-EDTA). The cells are washed once with DMEM 10% FBS, and then spun at 1000 RPM, for 5 minutes at room temperature. The pellet is resuspended in DMEM without serum, containing 0.1% BSA (serum free medium). The cells are
20 spun and resuspended again in serum free medium, and then spun and resuspended in the amount of serum free medium needed to provide a concentration of 1 mio. cells / ml, or 25,000 cells per 25ul. Just prior to the assay, a suitable amount of the antibody to be tested for anti-target function activity is added to the cell suspension.

For the assay, a standard chemoattractant is used to measure the mobility of the cells. The
25 chemoattractants are diluted in serum free medium. A suitable nonspecific chemoattractant is DMEM with 5% FBS. The chemoattractant solutions and control solutions without chemoattractant are pipetted (29 μl) into the lower plate wells. After placing and securing the filter plate over the lower wells, ensuring contact with the solution in the bottom wells, serial dilutions of the cell suspension are pipetted onto each site on the filter top. The plates are then covered and incubated at
30 37° C, 5% CO₂, for 3-4 hours.

After incubation, the upper filter side is rinsed with PBS and exposed while the upper filter areas are cleaned with wet cotton swabs. The filter is stained using the Diff-Quik[™] (VWR) dye kit, according to the manufacturer's instructions. The migrated cells are counted on the lower filter side using a microscope (magnification 200x), by counting of 5 high power field sections per well.

35

Example 2Endothelial Sprouting assay For Determining Antibody Activity

Cell-sprouting morphology is utilized as an easily visualized assay to determine the inhibitory effect of a candidate antibody on the protein target function for protein targets that stimulate endothelial cell sprouting, such as ARP2. Such assays have been described extensively in the literature (Nehls, V., et al., Histochem. Cell Biol. 104: 459-466 (1995); Koblizek, T. I., et al., Curr. Biol. 8: 529-532 (1988); and Kwak, H.J., et al., FEBS Lett. 448: 249-253). Briefly, endothelial cells from a suitable source, such as HUVECs or PPAECs (porcine pulmonary artery endothelial cells) are grown to confluence on microcarrier (MC) beads (diameter 175 μ m, available from Sigma) and placed into a 2.5 mg/ml fibrinogen gel containing the protein target at an appropriate effective concentration (200ng/ml is a suitable starting concentration, which the skilled practitioner may optimize) and the antibody in an appropriate range of concentrations (this will depend on antibody titer and affinity for the target), and 200 units/ml Trasylol (available from Bayer). Fibrin gels are incubated in M-199 with a daily supplement of the same amount of recombinant protein and antibody, 2.0% heat-inactivated fetal bovine serum, and 200 units/ml Trasylol. After three days, the extent of sprouting is determined using a phase-contrast microscope. A decrease in cell sprouting as compared to controls without antibody indicates a reduction in protein target activity by the antibody.

The foregoing is intended to be illustrative of the embodiments of the present invention, and are not intended to limit the invention in any way. Although the invention has been described with respect to specific modifications, the details thereof are not to be construed as limitations, for it will be apparent that various equivalents, changes and modifications may be resorted to without departing from the spirit and scope thereof and it is understood that such equivalent embodiments are to be included herein. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

What is Claimed is:

1. A method for developing biologically active agents that modulate activity of a brain tumor protein target (T_{BT}) gene or gene product, the method comprising:
combining a candidate biologically active agent with any one of:
 - 5 (a) a polypeptide encoded by any one of the sequences set forth in SEQ ID NO:1, 3, 5 or 7-16;
 - (b) a cell comprising a nucleic acid encoding and expressing a polypeptide encoded by any one of the sequences set forth in SEQ ID NO:1, 3, 5 or 7-16; or
 - (c) a non-human transgenic animal model for brain tumor gene function comprising one of:
 - 10 (i) a knockout of a gene corresponding to any one of the sequences set forth in Table 1; (ii) an exogenous and stably transmitted mammalian gene sequence comprising any one of the sequences set forth in SEQ ID NO:1, 3, 5 or 7-16; and
determining the effect of said agent on brain tumor induced molecular and cellular changes.
- 15 2. The method according to Claim 1, wherein said biologically active agent downregulates or upregulates expression.
3. The method according to Claim 1, wherein said biologically active agent inhibits or increases activity of said polypeptide.
- 20 4. A method for the diagnosis or staging of a brain tumor, the method comprising:
determining the upregulation or downregulation of expression in any one of the sequences set forth in Table 1.
- 25 5. The method according to Claim 4, wherein said brain tumor is an astrocytoma.
6. The method according to Claim 5, wherein said astrocytoma is a glioblastoma.
7. The method according to Claim 4, wherein said determining comprises detecting
30 increased or decreased amounts of mRNA or polypeptide in brain tumor cells.
8. An array of nucleic acids, comprising:
two or more nucleic acids comprising sequences set forth in SEQ ID NO:1, 3, 5 or 7-16.
- 35 9. A method of imaging a brain tumor, the method comprising:

administering to a patient an effective amount of a compound that specifically binds a protein encoded by a sequence set forth in SEQ ID NO:1, 3, 5 or 7-16, wherein said compound is conjugated to an imaging moiety; and

visualizing the imaging moieties of the compound.

5

10. The method of Claim 10 wherein said compound is administered by intrathecal administration.

11. The method of Claim 10 wherein said compound is administered by intravascular
10 administration.

12. The method of Claim 10 wherein the brain tumor is an astrocytoma.

13. The method of Claim 13, wherein said astrocytoma is a glioblastoma.

15

14. The method of Claim 10, wherein said compound is an antibody or antibody
fragment.

15. The method of Claim 10, wherein said imaging moiety is selected from the group
20 consisting of a radiographic moiety, a positron-emitting moiety, an optically visible dye, an optically visible particle, and a magnetic spin contrast moiety.

16. A method for the treatment of a brain tumor, the method comprising:

25 17. A method to treat a brain tumor, the method comprising:

administering a therapeutic amount of a compound that specifically binds a protein encoded
by a sequence set forth in SEQ ID NO:1, 3, 5 or 7-16, wherein said compound is conjugated to one or
more cytotoxic moieties.

30 18. The method of claim 18 wherein said compound is administered by intrathecal
administration.

19. The method of claim 18 wherein said compound is administered by intravascular
administration.

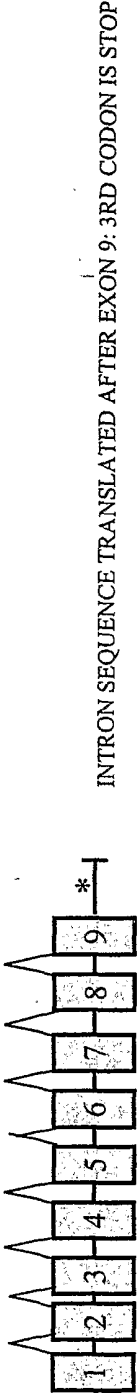
35

20. The method of Claim 18 wherein the brain tumor is an astrocytoma.

21. The method of Claim 21, wherein said astrocytoma is a glioblastoma.
22. The method of Claim 18, wherein said compound is an antibody or antibody fragment.
- 5 23. The method of Claim 18, wherein said cytotoxic moiety is selected from the group consisting of a radioactive moiety, a chemotoxic moiety, and a toxin protein moiety.
24. An isolated nucleic acid encoding a PTP-□ SM1 or SM2 splice variant.
- 10 25. The nucleic acid according to Claim 25, wherein said PTP-□ is human.
26. The nucleic acid according to Claim 25, wherein the sequence of said nucleic acid is set forth in SEQ ID NO:3 or SEQ ID NO:5.
- 15 27. A polypeptide encoded by the nucleic acid of Claim 25.

FIGURE 1

SPLICE VARIANT SM1: 2.38 KB (Exon 9 splice variant)



SPLICE VARIANT SM2: 3' 116 NT INSERTION AFTER EXON 23



SEQUENCE LISTING

<110> Sabine Muller

Erik Foehr

Daniel J. Chin

Mirella Gonzalez-Zulueta

<120> USE OF BIOMOLECULAR TARGETS IN THE
TREATMENT AND VISUALIZATION OF BRAIN TUMORS

<130> AGYT-005WO

<150> 09/816,703

<151> 2001-03-23

<150> 09/983,000

<151> 2001-10-17

<160> 24

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 7941

<212> DNA

<213> Homo Sapiens

<220>

<221> CDS

<222> (148)...(7092)

<400> 1

cacacatacg cagcgacgat ctcacttcga tctatacact ggaggattaa aacaaacaaa 60
caaaaaaaaaac atttccttcg ctccccctcc ctctccactc tgagaagcag aggagccgca 120
cggcgaggggg ccgcagaccg tctggaa atg cga atc cta aag cgt ttc ctc gct 174

Met Arg Ile Leu Lys Arg Phe Leu Ala

1

5

tgc att cag ctc ctc tgt gtt tgc cgc ctg gat tgg gct aat gga tac 222

Cys Ile Gln Leu Leu Cys Val Cys Arg Leu Asp Trp Ala Asn Gly Tyr
 10 15 20 25

 tac aga caa cag aga aaa ctt gtt gaa gag att ggc tgg tcc tat aca 270
 Tyr Arg Gln Gln Arg Lys Leu Val Glu Glu Ile Gly Trp Ser Tyr Thr
 30 35 40

 gga gca ctg aat caa aaa aat tgg gga aag aaa tat cca aca tgt aat 318
 Gly Ala Leu Asn Gln Lys Asn Trp Gly Lys Lys Tyr Pro Thr Cys Asn
 45 50 55

 agc cca aaa caa tct cct atc aat att gat gaa gat ctt aca caa gta 366
 Ser Pro Lys Gln Ser Pro Ile Asn Ile Asp Glu Asp Leu Thr Gln Val
 60 65 70

 aat gtg aat ctt aag aaa ctt aaa ttt cag ggt tgg gat aaa aca tca 414
 Asn Val Asn Leu Lys Lys Leu Lys Phe Gln Gly Trp Asp Lys Thr Ser
 75 80 85

 ttg gaa aac aca ttc att cat aac act ggg aaa aca gtg gaa att aat 462
 Leu Glu Asn Thr Phe Ile His Asn Thr Gly Lys Thr Val Glu Ile Asn
 90 95 100 105

 ctc act aat gac tac cgt gtc agc gga gga gtt tca gaa atg gtg ttt 510
 Leu Thr Asn Asp Tyr Arg Val Ser Gly Gly Val Ser Glu Met Val Phe
 110 115 120

 aaa gca agc aag ata act ttt cac tgg gga aaa tgc aat atg tca tct 558
 Lys Ala Ser Lys Ile Thr Phe His Trp Gly Lys Cys Asn Met Ser Ser
 125 130 135

 gat gga tca gag cat agt tta gaa gga caa aaa ttt cca ctt gag atg 606
 Asp Gly Ser Glu His Ser Leu Glu Gly Gln Lys Phe Pro Leu Glu Met
 140 145 150

 caa atc tac tgc ttt gat gcg gac cga ttt tca agt ttt gag gaa gca 654
 Gln Ile Tyr Cys Phe Asp Ala Asp Arg Phe Ser Ser Phe Glu Glu Ala
 155 160 165

 gtc aaa gga aaa ggg aag tta aga gct tta tcc att ttg ttt gag gtt 702
 Val Lys Gly Lys Gly Lys Leu Arg Ala Leu Ser Ile Leu Phe Glu Val

170	175	180	185	
ggg aca gaa gaa aat ttg gat ttc aaa gcg att att gat gga gtc gaa				750
Gly Thr Glu Glu Asn Leu Asp Phe Lys Ala Ile Ile Asp Gly Val Glu				
	190	195	200	
agt gtt agt cgt ttt ggg aag cag gct gct tta gat cca ttc ata ctg				798
Ser Val Ser Arg Phe Gly Lys Gln Ala Ala Leu Asp Pro Phe Ile Leu				
	205	210	215	
ttg aac ctt ctg cca aac tca act gac aag tat tac att tac aat ggc				846
Leu Asn Leu Leu Pro Asn Ser Thr Asp Lys Tyr Tyr Ile Tyr Asn Gly				
	220	225	230	
tca ttg aca tct cct ccc tgc aca gac aca gtt gac tgg att gtt ttt				894
Ser Leu Thr Ser Pro Pro Cys Thr Asp Thr Val Asp Trp Ile Val Phe				
	235	240	245	
aaa gat aca gtt agc atc tct gaa agc cag ttg gct gtt ttt tgt gaa				942
Lys Asp Thr Val Ser Ile Ser Glu Ser Gln Leu Ala Val Phe Cys Glu				
250	255	260	265	
gtt ctt aca atg caa caa tct ggt tat gtc atg ctg atg gac tac tta				990
Val Leu Thr Met Gln Gln Ser Gly Tyr Val Met Leu Met Asp Tyr Leu				
	270	275	280	
caa aac aat ttt cga gag caa cag tac aag ttc tct aga cag gtg ttt				1038
Gln Asn Asn Phe Arg Glu Gln Gln Tyr Lys Phe Ser Arg Gln Val Phe				
	285	290	295	
tcc tca tac act gga aag gaa gag att cat gaa gca gtt tgt agt tca				1086
Ser Ser Tyr Thr Gly Lys Glu Glu Ile His Glu Ala Val Cys Ser Ser				
	300	305	310	
gaa cca gaa aat gtt cag gct gac cca gag aat tat acc agc ctt ctt				1134
Glu Pro Glu Asn Val Gln Ala Asp Pro Glu Asn Tyr Thr Ser Leu Leu				
	315	320	325	
gtt aca tgg gaa aga cct cga gtc gtt tat gat acc atg att gag aag				1182
Val Thr Trp Glu Arg Pro Arg Val Val Tyr Asp Thr Met Ile Glu Lys				
330	335	340	345	

ttt gca gtt ttg tac cag cag ttg gat gga gag gac caa acc aag cat	1230
Phe Ala Val Leu Tyr Gln Gln Leu Asp Gly Glu Asp Gln Thr Lys His	
350 355 360	
gaa ttt ttg aca gat ggc tat caa gac ttg ggt gct att ctc aat aat	1278
Glu Phe Leu Thr Asp Gly Tyr Gln Asp Leu Gly Ala Ile Leu Asn Asn	
365 370 375	
ttg cta ccc aat atg agt tat gtt ctt cag ata gta gcc ata tgc act	1326
Leu Leu Pro Asn Met Ser Tyr Val Leu Gln Ile Val Ala Ile Cys Thr	
380 385 390	
aat ggc tta tat gga aaa tac agc gac caa ctg att gtc gac atg cct	1374
Asn Gly Leu Tyr Gly Lys Tyr Ser Asp Gln Leu Ile Val Asp Met Pro	
395 400 405	
act gat aat cct gaa ctt gat ctt ttc cct gaa tta att gga act gaa	1422
Thr Asp Asn Pro Glu Leu Asp Leu Phe Pro Glu Leu Ile Gly Thr Glu	
410 415 420 425	
gaa ata atc aag gag gag gaa gag gga aaa gac att gaa gaa ggc gct	1470
Glu Ile Ile Lys Glu Glu Glu Glu Gly Lys Asp Ile Glu Glu Gly Ala	
430 435 440	
att gtg aat cct ggt aga gac agt gct aca aac caa atc agg aaa aag	1518
Ile Val Asn Pro Gly Arg Asp Ser Ala Thr Asn Gln Ile Arg Lys Lys	
445 450 455	
gaa ccc cag att tct acc aca aca cac tac aat cgc ata ggg acg aaa	1566
Glu Pro Gln Ile Ser Thr Thr Thr His Tyr Asn Arg Ile Gly Thr Lys	
460 465 470	
tac aat gaa gcc aag act aac cga tcc cca aca aga gga agt gaa ttc	1614
Tyr Asn Glu Ala Lys Thr Asn Arg Ser Pro Thr Arg Gly Ser Glu Phe	
475 480 485	
tct gga aag ggt gat gtt ccc aat aca tct tta aat tcc act tcc caa	1662
Ser Gly Lys Gly Asp Val Pro Asn Thr Ser Leu Asn Ser Thr Ser Gln	
490 495 500 505	

cca gtc act aaa tta gcc aca gaa aaa gat att tcc ttg act tct cag 1710
Pro Val Thr Lys Leu Ala Thr Glu Lys Asp Ile Ser Leu Thr Ser Gln
510 515 520

act gtg act gaa ctg cca cct cac act gtg gaa ggt act tca gcc tct 1758
Thr Val Thr Glu Leu Pro Pro His Thr Val Glu Gly Thr Ser Ala Ser
525 530 535

tta aat gat ggc tct aaa act gtt ctt aga tct cca cat atg aac ttg 1806
Leu Asn Asp Gly Ser Lys Thr Val Leu Arg Ser Pro His Met Asn Leu
540 545 550

tcg ggg act gca gaa tcc tta aat aca gtt tct ata aca gaa tat gag 1854
Ser Gly Thr Ala Glu Ser Leu Asn Thr Val Ser Ile Thr Glu Tyr Glu
555 560 565

gag gag agt tta ttg acc agt ttc aag ctt gat act gga gct gaa gat 1902
Glu Glu Ser Leu Leu Thr Ser Phe Lys Leu Asp Thr Gly Ala Glu Asp
570 575 580 585

tct tca ggc tcc agt ccc gca act tct gct atc cca ttc atc tct gag 1950
Ser Ser Gly Ser Ser Pro Ala Thr Ser Ala Ile Pro Phe Ile Ser Glu
590 595 600

aac ata tcc caa ggg tat ata ttt tcc tcc gaa aac cca gag aca ata 1998
Asn Ile Ser Gln Gly Tyr Ile Phe Ser Ser Glu Asn Pro Glu Thr Ile
605 610 615

aca tat gat gtc ctt ata cca gaa tct gct aga aat gct tcc gaa gat 2046
Thr Tyr Asp Val Leu Ile Pro Glu Ser Ala Arg Asn Ala Ser Glu Asp
620 625 630

tca act tca tca ggt tca gaa gaa tca cta aag gat cct tct atg gag 2094
Ser Thr Ser Ser Gly Ser Glu Glu Ser Leu Lys Asp Pro Ser Met Glu
635 640 645

gga aat gtg tgg ttt cct agc tct aca gac ata aca gca cag ccc gat 2142
Gly Asn Val Trp Phe Pro Ser Ser Thr Asp Ile Thr Ala Gln Pro Asp
650 655 660 665

gtt gga tca ggc aga gag agc ttt ctc cag act aat tac act gag ata 2190

Val Gly Ser Gly Arg Glu Ser Phe Leu Gln Thr Asn Tyr Thr Glu Ile	
670 675 680	
cg t gtt gat gaa tct gag aag aca acc aag tcc ttt tct gca ggc cca	2238
Arg Val Asp Glu Ser Glu Lys Thr Thr Lys Ser Phe Ser Ala Gly Pro	
685 690 695	
gtg atg tca cag ggt ccc tca gtt aca gat ctg gaa atg cca cat tat	2286
Val Met Ser Gln Gly Pro Ser Val Thr Asp Leu Glu Met Pro His Tyr	
700 705 710	
tct acc ttt gcc tac ttc cca act gag gta aca cct cat gct ttt acc	2334
Ser Thr Phe Ala Tyr Phe Pro Thr Glu Val Thr Pro His Ala Phe Thr	
715 720 725	
cca tcc tcc aga caa cag gat ttg gtc tcc acg gtc aac gtg gta tac	2382
Pro Ser Ser Arg Gln Gln Asp Leu Val Ser Thr Val Asn Val Val Tyr	
730 735 740 745	
tcg cag aca acc caa ccg gta tac aat ggt gag aca cct ctt caa cct	2430
Ser Gln Thr Thr Gln Pro Val Tyr Asn Gly Glu Thr Pro Leu Gln Pro	
750 755 760	
tcc tac agt agt gaa gtc ttt cct cta gtc acc cct ttg ttg ctt gac	2478
Ser Tyr Ser Ser Glu Val Phe Pro Leu Val Thr Pro Leu Leu Leu Asp	
765 770 775	
aat cag atc ctc aac act acc cct gct gct tca agt agt gat tcg gcc	2526
Asn Gln Ile Leu Asn Thr Thr Pro Ala Ala Ser Ser Ser Asp Ser Ala	
780 785 790	
ttg cat gct acg cct gta ttt ccc agt gtc gat gtg tca ttt gaa tcc	2574
Leu His Ala Thr Pro Val Phe Pro Ser Val Asp Val Ser Phe Glu Ser	
795 800 805	
atc ctg tct tcc tat gat ggt gca cct ttg ctt cca ttt tcc tct gct	2622
Ile Leu Ser Ser Tyr Asp Gly Ala Pro Leu Leu Pro Phe Ser Ser Ala	
810 815 820 825	
tcc ttc agt agt gaa ttg ttt cgc cat ctg cat aca gtt tct caa atc	2670
Ser Phe Ser Ser Glu Leu Phe Arg His Leu His Thr Val Ser Gln Ile	

830	835	840	
ctt cca caa gtt act tca gct acc gag agt gat aag gtg ccc ttg cat			2718
Leu Pro Gln Val Thr Ser Ala Thr Glu Ser Asp Lys Val Pro Leu His			
845	850	855	
gct tct ctg cca gtg gct ggg ggt gat ttg cta tta gag ccc agc ctt			2766
Ala Ser Leu Pro Val Ala Gly Gly Asp Leu Leu Leu Glu Pro Ser Leu			
860	865	870	
gct cag tat tct gat gtg ctg tcc act act cat gct gct tca gag acg			2814
Ala Gln Tyr Ser Asp Val Leu Ser Thr Thr His Ala Ala Ser Glu Thr			
875	880	885	
ctg gaa ttt ggt agt gaa tct ggt gtt ctt tat aaa acg ctt atg ttt			2862
Leu Glu Phe Gly Ser Glu Ser Gly Val Leu Tyr Lys Thr Leu Met Phe			
890	895	900	905
tct caa gtt gaa cca ccc agc agt gat gcc atg atg cat gca cgt tct			2910
Ser Gln Val Glu Pro Pro Ser Ser Asp Ala Met Met His Ala Arg Ser			
910	915	920	
tca ggg cct gaa cct tct tat gcc ttg tct gat aat gag ggc tcc caa			2958
Ser Gly Pro Glu Pro Ser Tyr Ala Leu Ser Asp Asn Glu Gly Ser Gln			
925	930	935	
cac atc ttc act gtt tct tac agt tct gca ata cct gtg cat gat tct			3006
His Ile Phe Thr Val Ser Tyr Ser Ser Ala Ile Pro Val His Asp Ser			
940	945	950	
gtg ggt gta act tat cag ggt tcc tta ttt agc ggc cct agc cat ata			3054
Val Gly Val Thr Tyr Gln Gly Ser Leu Phe Ser Gly Pro Ser His Ile			
955	960	965	
cca ata cct aag tct tcg tta ata acc cca act gca tca tta ctg cag			3102
Pro Ile Pro Lys Ser Ser Leu Ile Thr Pro Thr Ala Ser Leu Leu Gln			
970	975	980	985
cct act cat gcc ctc tct ggt gat ggg gaa tgg tct gga gcc tct tct			3150
Pro Thr His Ala Leu Ser Gly Asp Gly Glu Trp Ser Gly Ala Ser Ser			
990	995	1000	

gat agt gaa ttt ctt tta cct gac aca gat ggg ctg aca gcc ctt aac 3198
Asp Ser Glu Phe Leu Leu Pro Asp Thr Asp Gly Leu Thr Ala Leu Asn
1005 1010 1015

att tct tca cct gtt tct gta gct gaa ttt aca tat aca aca tct gtg 3246
Ile Ser Ser Pro Val Ser Val Ala Glu Phe Thr Tyr Thr Thr Ser Val
1020 1025 1030

ttt ggt gat gat aat aag gcg ctt tct aaa agt gaa ata ata tat gga 3294
Phe Gly Asp Asp Asn Lys Ala Leu Ser Lys Ser Glu Ile Ile Tyr Gly
1035 1040 1045

aat gag act gaa ctg caa att cct tct ttc aat gag atg gtt tac cct 3342
Asn Glu Thr Glu Leu Gln Ile Pro Ser Phe Asn Glu Met Val Tyr Pro
1050 1055 1060 1065

tct gaa agc aca gtc atg ccc aac atg tat gat aat gta aat aag ttg 3390
Ser Glu Ser Thr Val Met Pro Asn Met Tyr Asp Asn Val Asn Lys Leu
1070 1075 1080

aat gcg tct tta caa gaa acc tct gtt tcc att tct agc acc aag ggc 3438
Asn Ala Ser Leu Gln Glu Thr Ser Val Ser Ile Ser Ser Thr Lys Gly
1085 1090 1095

atg ttt cca ggg tcc ctt gct cat acc acc act aag gtt ttt gat cat 3486
Met Phe Pro Gly Ser Leu Ala His Thr Thr Thr Lys Val Phe Asp His
1100 1105 1110

gag att agt caa gtt cca gaa aat aac ttt tca gtt caa cct aca cat 3534
Glu Ile Ser Gln Val Pro Glu Asn Asn Phe Ser Val Gln Pro Thr His
1115 1120 1125

act gtc tct caa gca tct ggt gac act tcg ctt aaa cct gtg ctt agt 3582
Thr Val Ser Gln Ala Ser Gly Asp Thr Ser Leu Lys Pro Val Leu Ser
1130 1135 1140 1145

gca aac tca gag cca gca tcc tct gac cct gct tct agt gaa atg tta 3630
Ala Asn Ser Glu Pro Ala Ser Ser Asp Pro Ala Ser Ser Glu Met Leu
1150 1155 1160

tct cct tca act cag ctc tta ttt tat gag acc tca gct tct ttt agt	3678
Ser Pro Ser Thr Gln Leu Leu Phe Tyr Glu Thr Ser Ala Ser Phe Ser	
1165 1170 1175	
act gaa gta ttg cta caa cct tcc ttt cag gct tct gat gtt gac acc	3726
Thr Glu Val Leu Leu Gln Pro Ser Phe Gln Ala Ser Asp Val Asp Thr	
1180 1185 1190	
ttg ctt aaa act gtt ctt cca gct gtg ccc agt gat cca ata ttg gtt	3774
Leu Leu Lys Thr Val Leu Pro Ala Val Pro Ser Asp Pro Ile Leu Val	
1195 1200 1205	
gaa acc ccc aaa gtt gat aaa att agt tct aca atg ttg cat ctc att	3822
Glu Thr Pro Lys Val Asp Lys Ile Ser Ser Thr Met Leu His Leu Ile	
1210 1215 1220 1225	
gta tca aat tct gct tca agt gaa aac atg ctg cac tct aca tct gta	3870
Val Ser Asn Ser Ala Ser Ser Glu Asn Met Leu His Ser Thr Ser Val	
1230 1235 1240	
cca gtt ttt gat gtg tcg cct act tct cat atg cac tct gct tca ctt	3918
Pro Val Phe Asp Val Ser Pro Thr Ser His Met His Ser Ala Ser Leu	
1245 1250 1255	
caa ggt ttg acc att tcc tat gca agt gag aaa tat gaa cca gtt ttg	3966
Gln Gly Leu Thr Ile Ser Tyr Ala Ser Glu Lys Tyr Glu Pro Val Leu	
1260 1265 1270	
tta aaa agt gaa agt tcc cac caa gtg gta cct tct ttg tac agt aat	4014
Leu Lys Ser Glu Ser Ser His Gln Val Val Pro Ser Leu Tyr Ser Asn	
1275 1280 1285	
gat gag ttg ttc caa acg gcc aat ttg gag att aac cag gcc cat ccc	4062
Asp Glu Leu Phe Gln Thr Ala Asn Leu Glu Ile Asn Gln Ala His Pro	
1290 1295 1300 1305	
cca aaa gga agg cat gta ttt gct aca cct gtt tta tca att gat gaa	4110
Pro Lys Gly Arg His Val Phe Ala Thr Pro Val Leu Ser Ile Asp Glu	
1310 1315 1320	
cca tta aat aca cta ata aat aag ctt ata cat tcc gat gaa att tta	4158

Pro	Leu	Asn	Thr	Leu	Ile	Asn	Lys	Leu	Ile	His	Ser	Asp	Glu	Ile	Leu		
				1325				1330					1335				
acc	tcc	acc	aaa	agt	tct	gtt	act	ggg	aag	gta	ttt	gct	ggg	att	cca	4206	
Thr	Ser	Thr	Lys	Ser	Ser	Val	Thr	Gly	Lys	Val	Phe	Ala	Gly	Ile	Pro		
			1340					1345					1350				
aca	gtt	gct	tct	gat	aca	ttt	gta	tct	act	gat	cat	tct	gtt	cct	ata	4254	
Thr	Val	Ala	Ser	Asp	Thr	Phe	Val	Ser	Thr	Asp	His	Ser	Val	Pro	Ile		
			1355				1360					1365					
gga	aat	ggg	cat	gtt	gcc	att	aca	gct	gtt	tct	ccc	cac	aga	gat	ggg	4302	
Gly	Asn	Gly	His	Val	Ala	Ile	Thr	Ala	Val	Ser	Pro	His	Arg	Asp	Gly		
1370						1375					1380				1385		
tct	gta	acc	tca	aca	aag	ttg	ctg	ttt	cct	tct	aag	gca	act	tct	gag	4350	
Ser	Val	Thr	Ser	Thr	Lys	Leu	Leu	Phe	Pro	Ser	Lys	Ala	Thr	Ser	Glu		
					1390					1395				1400			
ctg	agt	cat	agt	gcc	aaa	tct	gat	gcc	ggg	tta	gtg	ggg	ggg	ggg	gaa	4398	
Leu	Ser	His	Ser	Ala	Lys	Ser	Asp	Ala	Gly	Leu	Val	Gly	Gly	Gly	Glu		
			1405					1410					1415				
gat	ggg	gac	act	gat	gat	gat	ggg	gat	gat	gat	gat	gac	aga	gat	agt	4446	
Asp	Gly	Asp	Thr	Asp	Asp	Asp	Gly	Asp	Asp	Asp	Asp	Asp	Arg	Asp	Ser		
			1420				1425					1430					
gat	ggc	tta	tcc	att	cat	aag	tgt	atg	tca	tgc	tca	tcc	tat	aga	gaa	4494	
Asp	Gly	Leu	Ser	Ile	His	Lys	Cys	Met	Ser	Cys	Ser	Ser	Tyr	Arg	Glu		
			1435				1440					1445					
tca	cag	gaa	aag	gta	atg	aat	gat	tca	gac	acc	cac	gaa	aac	agt	ctt	4542	
Ser	Gln	Glu	Lys	Val	Met	Asn	Asp	Ser	Asp	Thr	His	Glu	Asn	Ser	Leu		
1450						1455					1460				1465		
atg	gat	cag	aat	aat	cca	atc	tca	tac	tca	cta	tct	gag	aat	tct	gaa	4590	
Met	Asp	Gln	Asn	Asn	Pro	Ile	Ser	Tyr	Ser	Leu	Ser	Glu	Asn	Ser	Glu		
					1470					1475			1480				
gaa	gat	aat	aga	gtc	aca	agt	gta	tcc	tca	gac	agt	caa	act	ggg	atg	4638	
Glu	Asp	Asn	Arg	Val	Thr	Ser	Val	Ser	Ser	Asp	Ser	Gln	Thr	Gly	Met		

1485	1490	1495	
gac aga agt cct ggt aaa tca cca tca gca aat ggg cta tcc caa aag			4686
Asp Arg Ser Pro Gly Lys Ser Pro Ser Ala Asn Gly Leu Ser Gln Lys			
1500	1505	1510	
cac aat gat gga aaa gag gaa aat gac att cag act ggt agt gct ctg			4734
His Asn Asp Gly Lys Glu Glu Asn Asp Ile Gln Thr Gly Ser Ala Leu			
1515	1520	1525	
ctt cct ctc agc cct gaa tct aaa gca tgg gca gtt ctg aca agt gat			4782
Leu Pro Leu Ser Pro Glu Ser Lys Ala Trp Ala Val Leu Thr Ser Asp			
1530	1535	1540	1545
gaa gaa agt gga tca ggg caa ggt acc tca gat agc ctt aat gag aat			4830
Glu Glu Ser Gly Ser Gly Gln Gly Thr Ser Asp Ser Leu Asn Glu Asn			
1550	1555	1560	
gag act tcc aca gat ttc agt ttt gca gac act aat gaa aaa gat gct			4878
Glu Thr Ser Thr Asp Phe Ser Phe Ala Asp Thr Asn Glu Lys Asp Ala			
1565	1570	1575	
gat ggg atc ctg gca gca ggt gac tca gaa ata act cct gga ttc cca			4926
Asp Gly Ile Leu Ala Ala Gly Asp Ser Glu Ile Thr Pro Gly Phe Pro			
1580	1585	1590	
cag tcc cca aca tca tct gtt act agc gag aac tca gaa gtg ttc cac			4974
Gln Ser Pro Thr Ser Ser Val Thr Ser Glu Asn Ser Glu Val Phe His			
1595	1600	1605	
gtt tca gag gca gag gcc agt aat agt agc cat gag tct cgt att ggt			5022
Val Ser Glu Ala Glu Ala Ser Asn Ser Ser His Glu Ser Arg Ile Gly			
1610	1615	1620	1625
cta gct gag ggg ttg gaa tcc gag aag aag gca gtt ata ccc ctt gtg			5070
Leu Ala Glu Gly Leu Glu Ser Glu Lys Lys Ala Val Ile Pro Leu Val			
1630	1635	1640	
atc gtg tca gcc ctg act ttt atc tgt cta gtg gtt ctt gtg ggt att			5118
Ile Val Ser Ala Leu Thr Phe Ile Cys Leu Val Val Leu Val Gly Ile			
1645	1650	1655	

ctc atc tac tgg agg aaa tgc ttc cag act gca cac ttt tac tta gag	5166
Leu Ile Tyr Trp Arg Lys Cys Phe Gln Thr Ala His Phe Tyr Leu Glu	
1660 1665 1670	
gac agt aca tcc cct aga gtt ata tcc aca cct cca aca cct atc ttt	5214
Asp Ser Thr Ser Pro Arg Val Ile Ser Thr Pro Pro Thr Pro Ile Phe	
1675 1680 1685	
cca att tca gat gat gtc gga gca att cca ata aag cac ttt cca aag	5262
Pro Ile Ser Asp Asp Val Gly Ala Ile Pro Ile Lys His Phe Pro Lys	
1690 1695 1700 1705	
cat gtt gca gat tta cat gca agt agt ggg ttt act gaa gaa ttt gag	5310
His Val Ala Asp Leu His Ala Ser Ser Gly Phe Thr Glu Glu Phe Glu	
1710 1715 1720	
aca ctg aaa gag ttt tac cag gaa gtg cag agc tgt act gtt gac tta	5358
Thr Leu Lys Glu Phe Tyr Gln Glu Val Gln Ser Cys Thr Val Asp Leu	
1725 1730 1735	
ggt att aca gca gac agc tcc aac cac cca gac aac aag cac aag aat	5406
Gly Ile Thr Ala Asp Ser Ser Asn His Pro Asp Asn Lys His Lys Asn	
1740 1745 1750	
cga tac ata aat atc gtt gcc tat gat cat agc agg gtt aag cta gca	5454
Arg Tyr Ile Asn Ile Val Ala Tyr Asp His Ser Arg Val Lys Leu Ala	
1755 1760 1765	
cag ctt gct gaa aag gat ggc aaa ctg act gat tat atc aat gcc aat	5502
Gln Leu Ala Glu Lys Asp Gly Lys Leu Thr Asp Tyr Ile Asn Ala Asn	
1770 1775 1780 1785	
tat gtt gat ggc tac aac aga cca aaa gct tat att gct gcc caa ggc	5550
Tyr Val Asp Gly Tyr Asn Arg Pro Lys Ala Tyr Ile Ala Ala Gln Gly	
1790 1795 1800	
cca ctg aaa tcc aca gct gaa gat ttc tgg aga atg ata tgg gaa cat	5598
Pro Leu Lys Ser Thr Ala Glu Asp Phe Trp Arg Met Ile Trp Glu His	
1805 1810 1815	

aat gtg gaa gtt att gtc atg ata aca aac ctc gtg gag aaa gga agg	5646
Asn Val Glu Val Ile Val Met Ile Thr Asn Leu Val Glu Lys Gly Arg	
1820 1825 1830	
aga aaa tgt gat cag tac tgg cct gcc gat ggg agt gag gag tac ggg	5694
Arg Lys Cys Asp Gln Tyr Trp Pro Ala Asp Gly Ser Glu Glu Tyr Gly	
1835 1840 1845	
aac ttt ctg gtc act cag aag agt gtg caa gtg ctt gcc tat tat act	5742
Asn Phe Leu Val Thr Gln Lys Ser Val Gln Val Leu Ala Tyr Tyr Thr	
1850 1855 1860 1865	
gtg agg aat ttt act cta aga aac aca aaa ata aaa aag ggc tcc cag	5790
Val Arg Asn Phe Thr Leu Arg Asn Thr Lys Ile Lys Lys Gly Ser Gln	
1870 1875 1880	
aaa gga aga ccc agt gga cgt gtg gtc aca cag tat cac tac acg cag	5838
Lys Gly Arg Pro Ser Gly Arg Val Val Thr Gln Tyr His Tyr Thr Gln	
1885 1890 1895	
tgg cct gac atg gga gta cca gag tac tcc ctg cca gtg ctg acc ttt	5886
Trp Pro Asp Met Gly Val Pro Glu Tyr Ser Leu Pro Val Leu Thr Phe	
1900 1905 1910	
gtg aga aag gca gcc tat gcc aag cgc cat gca gtg ggg cct gtt gtc	5934
Val Arg Lys Ala Ala Tyr Ala Lys Arg His Ala Val Gly Pro Val Val	
1915 1920 1925	
gtc cac tgc agt gct gga gtt gga aga aca ggc aca tat att gtg cta	5982
Val His Cys Ser Ala Gly Val Gly Arg Thr Gly Thr Tyr Ile Val Leu	
1930 1935 1940 1945	
gac agt atg ttg cag cag att caa cac gaa gga act gtc aac ata ttt	6030
Asp Ser Met Leu Gln Gln Ile Gln His Glu Gly Thr Val Asn Ile Phe	
1950 1955 1960	
ggc ttc tta aaa cac atc cgt tca caa aga aat tat ttg gta caa act	6078
Gly Phe Leu Lys His Ile Arg Ser Gln Arg Asn Tyr Leu Val Gln Thr	
1965 1970 1975	
gag gag caa tat gtc ttc att cat gat aca ctg gtt gag gcc ata ctt	6126

Glu Glu Gln Tyr Val Phe Ile His Asp Thr Leu Val Glu Ala Ile Leu	
1980	1985
1990	
agt aaa gaa act gag gtg ctg gac agt cat att cat gcc tat gtt aat	6174
Ser Lys Glu Thr Glu Val Leu Asp Ser His Ile His Ala Tyr Val Asn	
1995	2000
2005	
gca ctc ctc att cct gga cca gca ggc aaa aca aag cta gag aaa caa	6222
Ala Leu Leu Ile Pro Gly Pro Ala Gly Lys Thr Lys Leu Glu Lys Gln	
2010	2015
2020	2025
ttc cag ctc ctg agc cag tca aat ata cag cag agt gac tat tct gca	6270
Phe Gln Leu Leu Ser Gln Ser Asn Ile Gln Gln Ser Asp Tyr Ser Ala	
2030	2035
2040	
gcc cta aag caa tgc aac agg gaa aag aat cga act tct tct atc atc	6318
Ala Leu Lys Gln Cys Asn Arg Glu Lys Asn Arg Thr Ser Ser Ile Ile	
2045	2050
2055	
cct gtg gaa aga tca agg gtt ggc att tca tcc ctg agt gga gaa ggc	6366
Pro Val Glu Arg Ser Arg Val Gly Ile Ser Ser Leu Ser Gly Glu Gly	
2060	2065
2070	
aca gac tac atc aat gcc tcc tat atc atg ggc tat tac cag agc aat	6414
Thr Asp Tyr Ile Asn Ala Ser Tyr Ile Met Gly Tyr Tyr Gln Ser Asn	
2075	2080
2085	
gaa ttc atc att acc cag cac cct ctc ctt cat acc atc aag gat ttc	6462
Glu Phe Ile Ile Thr Gln His Pro Leu Leu His Thr Ile Lys Asp Phe	
2090	2095
2100	2105
tgg agg atg ata tgg gac cat aat gcc caa ctg gtg gtt atg att cct	6510
Trp Arg Met Ile Trp Asp His Asn Ala Gln Leu Val Val Met Ile Pro	
2110	2115
2120	
gat ggc caa aac atg gca gaa gat gaa ttt gtt tac tgg cca aat aaa	6558
Asp Gly Gln Asn Met Ala Glu Asp Glu Phe Val Tyr Trp Pro Asn Lys	
2125	2130
2135	
gat gag cct ata aat tgt gag agc ttt aag gtc act ctt atg gct gaa	6606
Asp Glu Pro Ile Asn Cys Glu Ser Phe Lys Val Thr Leu Met Ala Glu	

2140	2145	2150	
gaa cac aaa tgt cta tct aat gag gaa aaa ctt ata att cag gac ttt			6654
Glu His Lys Cys Leu Ser Asn Glu Glu Lys Leu Ile Ile Gln Asp Phe			
2155	2160	2165	
atc tta gaa gct aca cag gat gat tat gta ctt gaa gtg agg cac ttt			6702
Ile Leu Glu Ala Thr Gln Asp Asp Tyr Val Leu Glu Val Arg His Phe			
2170	2175	2180	2185
cag tgt cct aaa tgg cca aat cca gat agc ccc att agt aaa act ttt			6750
Gln Cys Pro Lys Trp Pro Asn Pro Asp Ser Pro Ile Ser Lys Thr Phe			
2190	2195	2200	
gaa ctt ata agt gtt ata aaa gaa gaa gct gcc aat agg gat ggg cct			6798
Glu Leu Ile Ser Val Ile Lys Glu Glu Ala Ala Asn Arg Asp Gly Pro			
2205	2210	2215	
atg att gtt cat gat gag cat gga gga gtg acg gca gga act ttc tgt			6846
Met Ile Val His Asp Glu His Gly Gly Val Thr Ala Gly Thr Phe Cys			
2220	2225	2230	
gct ctg aca acc ctt atg cac caa cta gaa aaa gaa aat tcc gtg gat			6894
Ala Leu Thr Thr Leu Met His Gln Leu Glu Lys Glu Asn Ser Val Asp			
2235	2240	2245	
gtt tac cag gta gcc aag atg atc aat ctg atg agg cca gga gtc ttt			6942
Val Tyr Gln Val Ala Lys Met Ile Asn Leu Met Arg Pro Gly Val Phe			
2250	2255	2260	2265
gct gac att gag cag tat cag ttt ctc tac aaa gtg atc ctc agc ctt			6990
Ala Asp Ile Glu Gln Tyr Gln Phe Leu Tyr Lys Val Ile Leu Ser Leu			
2270	2275	2280	
gtg agc aca agg cag gaa gag aat cca tcc acc tct ctg gac agt aat			7038
Val Ser Thr Arg Gln Glu Glu Asn Pro Ser Thr Ser Leu Asp Ser Asn			
2285	2290	2295	
ggg gca gca ttg cct gat gga aat ata gct gag agc tta gag tct tta			7086
Gly Ala Ala Leu Pro Asp Gly Asn Ile Ala Glu Ser Leu Glu Ser Leu			
2300	2305	2310	

gtt taa cacagaaagg ggtgggggga ctcacatctg agcattgttt tcctcttcct 7142
 Val *

aaaattaggc aggaaaatca gtctagttct gttatctgtt gatttcccat cacctgacag 7202
 taactttcat gacataggat tctgccgcca aatttatatc attaacaatg tgtgcctttt 7262
 tgcaagactt gtaatttact tattatgttt gaactaaaat gattgaattt tacagtattt 7322
 ctaagaatgg aattgtggta tttttttctg tattgatttt aacagaaaat ttcaatttat 7382
 agaggttagg aattccaaac tacagaaaat gtttgttttt agtgtcaaat ttttagctgt 7442
 atttgtagca attatcaggt ttgctagaaa tataactttt aatacagtag cctgtaaata 7502
 aaacactctt ccatatgata ttcaacattt tacaactgca gtattcacct aaagtagaaa 7562
 taatctgtta cttattgtaa atactgccct agtgtctcca tggaccaaat ttatatattat 7622
 aattgtagat ttttatattt tactactgag tcaagttttc tagttctgtg taattgttta 7682
 gttaaatgac gtagttcatt agctggctct actctaccag ttttctgaca ttgtattgtg 7742
 ttacctaaagt cattaacttt gtttcagcat gtaattttta cttttgtgga aaatagaaat 7802
 accttcattt tgaaagaagt ttttatgaga ataacacctt accaaacatt gttcaaagg 7862
 tttttatcca aggaattgca aaaataaata taaatattgc cattaataaaa aaaaaaaaaa 7922
 aaaaaaaaaa aaaaaaaaaa 7941

<210> 2

<211> 2314

<212> PRT

<213> Homo Sapiens

<400> 2

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

115	120	125
His Trp Gly Lys Cys Asn Met Ser Ser Asp Gly Ser Glu His Ser Leu		
130	135	140
Glu Gly Gln Lys Phe Pro Leu Glu Met Gln Ile Tyr Cys Phe Asp Ala		
145	150	155
160		
Asp Arg Phe Ser Ser Phe Glu Glu Ala Val Lys Gly Lys Gly Lys Leu		
165	170	175
Arg Ala Leu Ser Ile Leu Phe Glu Val Gly Thr Glu Glu Asn Leu Asp		
180	185	190
Phe Lys Ala Ile Ile Asp Gly Val Glu Ser Val Ser Arg Phe Gly Lys		
195	200	205
Gln Ala Ala Leu Asp Pro Phe Ile Leu Leu Asn Leu Leu Pro Asn Ser		
210	215	220
Thr Asp Lys Tyr Tyr Ile Tyr Asn Gly Ser Leu Thr Ser Pro Pro Cys		
225	230	235
240		
Thr Asp Thr Val Asp Trp Ile Val Phe Lys Asp Thr Val Ser Ile Ser		
245	250	255
Glu Ser Gln Leu Ala Val Phe Cys Glu Val Leu Thr Met Gln Gln Ser		
260	265	270
Gly Tyr Val Met Leu Met Asp Tyr Leu Gln Asn Asn Phe Arg Glu Gln		
275	280	285
Gln Tyr Lys Phe Ser Arg Gln Val Phe Ser Ser Tyr Thr Gly Lys Glu		
290	295	300
Glu Ile His Glu Ala Val Cys Ser Ser Glu Pro Glu Asn Val Gln Ala		
305	310	315
320		
Asp Pro Glu Asn Tyr Thr Ser Leu Leu Val Thr Trp Glu Arg Pro Arg		
325	330	335
Val Val Tyr Asp Thr Met Ile Glu Lys Phe Ala Val Leu Tyr Gln Gln		
340	345	350
Leu Asp Gly Glu Asp Gln Thr Lys His Glu Phe Leu Thr Asp Gly Tyr		
355	360	365
Gln Asp Leu Gly Ala Ile Leu Asn Asn Leu Leu Pro Asn Met Ser Tyr		
370	375	380
Val Leu Gln Ile Val Ala Ile Cys Thr Asn Gly Leu Tyr Gly Lys Tyr		
385	390	395
400		
Ser Asp Gln Leu Ile Val Asp Met Pro Thr Asp Asn Pro Glu Leu Asp		
405	410	415
Leu Phe Pro Glu Leu Ile Gly Thr Glu Glu Ile Ile Lys Glu Glu Glu		
420	425	430
Glu Gly Lys Asp Ile Glu Glu Gly Ala Ile Val Asn Pro Gly Arg Asp		
435	440	445

Ser Ala Thr Asn Gln Ile Arg Lys Lys Glu Pro Gln Ile Ser Thr Thr
 450 455 460
 Thr His Tyr Asn Arg Ile Gly Thr Lys Tyr Asn Glu Ala Lys Thr Asn
 465 470 475 480
 Arg Ser Pro Thr Arg Gly Ser Glu Phe Ser Gly Lys Gly Asp Val Pro
 485 490 495
 Asn Thr Ser Leu Asn Ser Thr Ser Gln Pro Val Thr Lys Leu Ala Thr
 500 505 510
 Glu Lys Asp Ile Ser Leu Thr Ser Gln Thr Val Thr Glu Leu Pro Pro
 515 520 525
 His Thr Val Glu Gly Thr Ser Ala Ser Leu Asn Asp Gly Ser Lys Thr
 530 535 540
 Val Leu Arg Ser Pro His Met Asn Leu Ser Gly Thr Ala Glu Ser Leu
 545 550 555 560
 Asn Thr Val Ser Ile Thr Glu Tyr Glu Glu Glu Ser Leu Leu Thr Ser
 565 570 575
 Phe Lys Leu Asp Thr Gly Ala Glu Asp Ser Ser Gly Ser Ser Pro Ala
 580 585 590
 Thr Ser Ala Ile Pro Phe Ile Ser Glu Asn Ile Ser Gln Gly Tyr Ile
 595 600 605
 Phe Ser Ser Glu Asn Pro Glu Thr Ile Thr Tyr Asp Val Leu Ile Pro
 610 615 620
 Glu Ser Ala Arg Asn Ala Ser Glu Asp Ser Thr Ser Ser Gly Ser Glu
 625 630 635 640
 Glu Ser Leu Lys Asp Pro Ser Met Glu Gly Asn Val Trp Phe Pro Ser
 645 650 655
 Ser Thr Asp Ile Thr Ala Gln Pro Asp Val Gly Ser Gly Arg Glu Ser
 660 665 670
 Phe Leu Gln Thr Asn Tyr Thr Glu Ile Arg Val Asp Glu Ser Glu Lys
 675 680 685
 Thr Thr Lys Ser Phe Ser Ala Gly Pro Val Met Ser Gln Gly Pro Ser
 690 695 700
 Val Thr Asp Leu Glu Met Pro His Tyr Ser Thr Phe Ala Tyr Phe Pro
 705 710 715 720
 Thr Glu Val Thr Pro His Ala Phe Thr Pro Ser Ser Arg Gln Gln Asp
 725 730 735
 Leu Val Ser Thr Val Asn Val Val Tyr Ser Gln Thr Thr Gln Pro Val
 740 745 750
 Tyr Asn Gly Glu Thr Pro Leu Gln Pro Ser Tyr Ser Ser Glu Val Phe
 755 760 765
 Pro Leu Val Thr Pro Leu Leu Leu Asp Asn Gln Ile Leu Asn Thr Thr

770	775	780
Pro Ala Ala Ser Ser Ser Asp Ser Ala Leu His Ala Thr Pro Val Phe		
785	790	795
Pro Ser Val Asp Val Ser Phe Glu Ser Ile Leu Ser Ser Tyr Asp Gly		800
	805	810
Ala Pro Leu Leu Pro Phe Ser Ser Ala Ser Phe Ser Ser Glu Leu Phe		815
	820	825
Arg His Leu His Thr Val Ser Gln Ile Leu Pro Gln Val Thr Ser Ala		830
	835	840
Thr Glu Ser Asp Lys Val Pro Leu His Ala Ser Leu Pro Val Ala Gly		845
	850	855
Gly Asp Leu Leu Leu Glu Pro Ser Leu Ala Gln Tyr Ser Asp Val Leu		860
865	870	875
Ser Thr Thr His Ala Ala Ser Glu Thr Leu Glu Phe Gly Ser Glu Ser		880
	885	890
Gly Val Leu Tyr Lys Thr Leu Met Phe Ser Gln Val Glu Pro Pro Ser		895
	900	905
Ser Asp Ala Met Met His Ala Arg Ser Ser Gly Pro Glu Pro Ser Tyr		910
	915	920
Ala Leu Ser Asp Asn Glu Gly Ser Gln His Ile Phe Thr Val Ser Tyr		925
	930	935
Ser Ser Ala Ile Pro Val His Asp Ser Val Gly Val Thr Tyr Gln Gly		940
945	950	955
Ser Leu Phe Ser Gly Pro Ser His Ile Pro Ile Pro Lys Ser Ser Leu		960
	965	970
Ile Thr Pro Thr Ala Ser Leu Leu Gln Pro Thr His Ala Leu Ser Gly		975
	980	985
Asp Gly Glu Trp Ser Gly Ala Ser Ser Asp Ser Glu Phe Leu Leu Pro		990
	995	1000
Asp Thr Asp Gly Leu Thr Ala Leu Asn Ile Ser Ser Pro Val Ser Val		1005
	1010	1015
Ala Glu Phe Thr Tyr Thr Thr Ser Val Phe Gly Asp Asp Asn Lys Ala		1020
1025	1030	1035
Leu Ser Lys Ser Glu Ile Ile Tyr Gly Asn Glu Thr Glu Leu Gln Ile		1040
	1045	1050
Pro Ser Phe Asn Glu Met Val Tyr Pro Ser Glu Ser Thr Val Met Pro		1055
	1060	1065
Asn Met Tyr Asp Asn Val Asn Lys Leu Asn Ala Ser Leu Gln Glu Thr		1070
	1075	1080
Ser Val Ser Ile Ser Ser Thr Lys Gly Met Phe Pro Gly Ser Leu Ala		1085
	1090	1095
		1100

His Thr Thr Thr Lys Val Phe Asp His Glu Ile Ser Gln Val Pro Glu
 1105 1110 1115 1120
 Asn Asn Phe Ser Val Gln Pro Thr His Thr Val Ser Gln Ala Ser Gly
 1125 1130 1135
 Asp Thr Ser Leu Lys Pro Val Leu Ser Ala Asn Ser Glu Pro Ala Ser
 1140 1145 1150
 Ser Asp Pro Ala Ser Ser Glu Met Leu Ser Pro Ser Thr Gln Leu Leu
 1155 1160 1165
 Phe Tyr Glu Thr Ser Ala Ser Phe Ser Thr Glu Val Leu Leu Gln Pro
 1170 1175 1180
 Ser Phe Gln Ala Ser Asp Val Asp Thr Leu Leu Lys Thr Val Leu Pro
 1185 1190 1195 1200
 Ala Val Pro Ser Asp Pro Ile Leu Val Glu Thr Pro Lys Val Asp Lys
 1205 1210 1215
 Ile Ser Ser Thr Met Leu His Leu Ile Val Ser Asn Ser Ala Ser Ser
 1220 1225 1230
 Glu Asn Met Leu His Ser Thr Ser Val Pro Val Phe Asp Val Ser Pro
 1235 1240 1245
 Thr Ser His Met His Ser Ala Ser Leu Gln Gly Leu Thr Ile Ser Tyr
 1250 1255 1260
 Ala Ser Glu Lys Tyr Glu Pro Val Leu Leu Lys Ser Glu Ser Ser His
 1265 1270 1275 1280
 Gln Val Val Pro Ser Leu Tyr Ser Asn Asp Glu Leu Phe Gln Thr Ala
 1285 1290 1295
 Asn Leu Glu Ile Asn Gln Ala His Pro Pro Lys Gly Arg His Val Phe
 1300 1305 1310
 Ala Thr Pro Val Leu Ser Ile Asp Glu Pro Leu Asn Thr Leu Ile Asn
 1315 1320 1325
 Lys Leu Ile His Ser Asp Glu Ile Leu Thr Ser Thr Lys Ser Ser Val
 1330 1335 1340
 Thr Gly Lys Val Phe Ala Gly Ile Pro Thr Val Ala Ser Asp Thr Phe
 1345 1350 1355 1360
 Val Ser Thr Asp His Ser Val Pro Ile Gly Asn Gly His Val Ala Ile
 1365 1370 1375
 Thr Ala Val Ser Pro His Arg Asp Gly Ser Val Thr Ser Thr Lys Leu
 1380 1385 1390
 Leu Phe Pro Ser Lys Ala Thr Ser Glu Leu Ser His Ser Ala Lys Ser
 1395 1400 1405
 Asp Ala Gly Leu Val Gly Gly Gly Glu Asp Gly Asp Thr Asp Asp Asp
 1410 1415 1420
 Gly Asp Asp Asp Asp Asp Arg Asp Ser Asp Gly Leu Ser Ile His Lys

1425	1430	1435	1440
Cys Met Ser Cys Ser Ser Tyr Arg Glu Ser Gln Glu Lys Val Met Asn			
	1445	1450	1455
Asp Ser Asp Thr His Glu Asn Ser Leu Met Asp Gln Asn Asn Pro Ile			
	1460	1465	1470
Ser Tyr Ser Leu Ser Glu Asn Ser Glu Glu Asp Asn Arg Val Thr Ser			
	1475	1480	1485
Val Ser Ser Asp Ser Gln Thr Gly Met Asp Arg Ser Pro Gly Lys Ser			
	1490	1495	1500
Pro Ser Ala Asn Gly Leu Ser Gln Lys His Asn Asp Gly Lys Glu Glu			
1505	1510	1515	1520
Asn Asp Ile Gln Thr Gly Ser Ala Leu Leu Pro Leu Ser Pro Glu Ser			
	1525	1530	1535
Lys Ala Trp Ala Val Leu Thr Ser Asp Glu Glu Ser Gly Ser Gly Gln			
	1540	1545	1550
Gly Thr Ser Asp Ser Leu Asn Glu Asn Glu Thr Ser Thr Asp Phe Ser			
	1555	1560	1565
Phe Ala Asp Thr Asn Glu Lys Asp Ala Asp Gly Ile Leu Ala Ala Gly			
	1570	1575	1580
Asp Ser Glu Ile Thr Pro Gly Phe Pro Gln Ser Pro Thr Ser Ser Val			
1585	1590	1595	1600
Thr Ser Glu Asn Ser Glu Val Phe His Val Ser Glu Ala Glu Ala Ser			
	1605	1610	1615
Asn Ser Ser His Glu Ser Arg Ile Gly Leu Ala Glu Gly Leu Glu Ser			
	1620	1625	1630
Glu Lys Lys Ala Val Ile Pro Leu Val Ile Val Ser Ala Leu Thr Phe			
	1635	1640	1645
Ile Cys Leu Val Val Leu Val Gly Ile Leu Ile Tyr Trp Arg Lys Cys			
	1650	1655	1660
Phe Gln Thr Ala His Phe Tyr Leu Glu Asp Ser Thr Ser Pro Arg Val			
1665	1670	1675	1680
Ile Ser Thr Pro Pro Thr Pro Ile Phe Pro Ile Ser Asp Asp Val Gly			
	1685	1690	1695
Ala Ile Pro Ile Lys His Phe Pro Lys His Val Ala Asp Leu His Ala			
	1700	1705	1710
Ser Ser Gly Phe Thr Glu Glu Phe Glu Thr Leu Lys Glu Phe Tyr Gln			
	1715	1720	1725
Glu Val Gln Ser Cys Thr Val Asp Leu Gly Ile Thr Ala Asp Ser Ser			
	1730	1735	1740
Asn His Pro Asp Asn Lys His Lys Asn Arg Tyr Ile Asn Ile Val Ala			
1745	1750	1755	1760

Tyr Asp His Ser Arg Val Lys Leu Ala Gln Leu Ala Glu Lys Asp Gly
 1765 1770 1775
 Lys Leu Thr Asp Tyr Ile Asn Ala Asn Tyr Val Asp Gly Tyr Asn Arg
 1780 1785 1790
 Pro Lys Ala Tyr Ile Ala Ala Gln Gly Pro Leu Lys Ser Thr Ala Glu
 1795 1800 1805
 Asp Phe Trp Arg Met Ile Trp Glu His Asn Val Glu Val Ile Val Met
 1810 1815 1820
 Ile Thr Asn Leu Val Glu Lys Gly Arg Arg Lys Cys Asp Gln Tyr Trp
 1825 1830 1835 1840
 Pro Ala Asp Gly Ser Glu Glu Tyr Gly Asn Phe Leu Val Thr Gln Lys
 1845 1850 1855
 Ser Val Gln Val Leu Ala Tyr Tyr Thr Val Arg Asn Phe Thr Leu Arg
 1860 1865 1870
 Asn Thr Lys Ile Lys Lys Gly Ser Gln Lys Gly Arg Pro Ser Gly Arg
 1875 1880 1885
 Val Val Thr Gln Tyr His Tyr Thr Gln Trp Pro Asp Met Gly Val Pro
 1890 1895 1900
 Glu Tyr Ser Leu Pro Val Leu Thr Phe Val Arg Lys Ala Ala Tyr Ala
 1905 1910 1915 1920
 Lys Arg His Ala Val Gly Pro Val Val Val His Cys Ser Ala Gly Val
 1925 1930 1935
 Gly Arg Thr Gly Thr Tyr Ile Val Leu Asp Ser Met Leu Gln Gln Ile
 1940 1945 1950
 Gln His Glu Gly Thr Val Asn Ile Phe Gly Phe Leu Lys His Ile Arg
 1955 1960 1965
 Ser Gln Arg Asn Tyr Leu Val Gln Thr Glu Glu Gln Tyr Val Phe Ile
 1970 1975 1980
 His Asp Thr Leu Val Glu Ala Ile Leu Ser Lys Glu Thr Glu Val Leu
 1985 1990 1995 2000
 Asp Ser His Ile His Ala Tyr Val Asn Ala Leu Leu Ile Pro Gly Pro
 2005 2010 2015
 Ala Gly Lys Thr Lys Leu Glu Lys Gln Phe Gln Leu Leu Ser Gln Ser
 2020 2025 2030
 Asn Ile Gln Gln Ser Asp Tyr Ser Ala Ala Leu Lys Gln Cys Asn Arg
 2035 2040 2045
 Glu Lys Asn Arg Thr Ser Ser Ile Ile Pro Val Glu Arg Ser Arg Val
 2050 2055 2060
 Gly Ile Ser Ser Leu Ser Gly Glu Gly Thr Asp Tyr Ile Asn Ala Ser
 2065 2070 2075 2080
 Tyr Ile Met Gly Tyr Tyr Gln Ser Asn Glu Phe Ile Ile Thr Gln His

2085					2090					2095						
Pro	Leu	Leu	His	Thr	Ile	Lys	Asp	Phe	Trp	Arg	Met	Ile	Trp	Asp	His	
2100					2105					2110						
Asn	Ala	Gln	Leu	Val	Val	Met	Ile	Pro	Asp	Gly	Gln	Asn	Met	Ala	Glu	
2115					2120					2125						
Asp	Glu	Phe	Val	Tyr	Trp	Pro	Asn	Lys	Asp	Glu	Pro	Ile	Asn	Cys	Glu	
2130					2135					2140						
Ser	Phe	Lys	Val	Thr	Leu	Met	Ala	Glu	Glu	His	Lys	Cys	Leu	Ser	Asn	
2145					2150					2155					2160	
Glu	Glu	Lys	Leu	Ile	Ile	Gln	Asp	Phe	Ile	Leu	Glu	Ala	Thr	Gln	Asp	
2165					2170					2175						
Asp	Tyr	Val	Leu	Glu	Val	Arg	His	Phe	Gln	Cys	Pro	Lys	Trp	Pro	Asn	
2180					2185					2190						
Pro	Asp	Ser	Pro	Ile	Ser	Lys	Thr	Phe	Glu	Leu	Ile	Ser	Val	Ile	Lys	
2195					2200					2205						
Glu	Glu	Ala	Ala	Asn	Arg	Asp	Gly	Pro	Met	Ile	Val	His	Asp	Glu	His	
2210					2215					2220						
Gly	Gly	Val	Thr	Ala	Gly	Thr	Phe	Cys	Ala	Leu	Thr	Thr	Leu	Met	His	
2225					2230					2235					2240	
Gln	Leu	Glu	Lys	Glu	Asn	Ser	Val	Asp	Val	Tyr	Gln	Val	Ala	Lys	Met	
2245					2250					2255						
Ile	Asn	Leu	Met	Arg	Pro	Gly	Val	Phe	Ala	Asp	Ile	Glu	Gln	Tyr	Gln	
2260					2265					2270						
Phe	Leu	Tyr	Lys	Val	Ile	Leu	Ser	Leu	Val	Ser	Thr	Arg	Gln	Glu	Glu	
2275					2280					2285						
Asn	Pro	Ser	Thr	Ser	Leu	Asp	Ser	Asn	Gly	Ala	Ala	Leu	Pro	Asp	Gly	
2290					2295					2300						
Asn	Ile	Ala	Glu	Ser	Leu	Glu	Ser	Leu	Val							
2305					2310											

<210> 3

<211> 3091

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (148)...(1272)

<223> PTP-zeta SM1 exon 9 variant

<221> misc_feature

<222> (1262)...(1262)

<223> coding sequence alternative splice site

<221> misc_feature

<222> (1273)...(3091)

<223> 3' Untranslated Region

<400> 3

cacacatacg cacgcacgat ctcaacttga tctatacact ggaggattaa aacaaacaaa 60
caaaaaaaaaac atttccttcg ctccccctcc ctctccactc tgagaagcag aggagccgca 120
cggcgagggg cgcgagaccg tctggaa atg cga atc cta aag cgt ttc ctc gct 174

Met Arg Ile Leu Lys Arg Phe Leu Ala

1

5

tgc att cag ctc ctc tgt gtt tgc cgc ctg gat tgg gct aat gga tac 222
Cys Ile Gln Leu Leu Cys Val Cys Arg Leu Asp Trp Ala Asn Gly Tyr
10 15 20 25

tac aga caa cag aga aaa ctt gtt gaa gag att ggc tgg tcc tat aca 270
Tyr Arg Gln Gln Arg Lys Leu Val Glu Glu Ile Gly Trp Ser Tyr Thr
30 35 40

gga gca ctg aat caa aaa aat tgg gga aag aaa tat cca aca tgt aat 318
Gly Ala Leu Asn Gln Lys Asn Trp Gly Lys Lys Tyr Pro Thr Cys Asn
45 50 55

agc cca aaa caa tct cct atc aat att gat gaa gat ctt aca caa gta 366
Ser Pro Lys Gln Ser Pro Ile Asn Ile Asp Glu Asp Leu Thr Gln Val
60 65 70

aat gtg aat ctt aag aaa ctt aaa ttt cag ggt tgg gat aaa aca tca 414
Asn Val Asn Leu Lys Lys Leu Lys Phe Gln Gly Trp Asp Lys Thr Ser
75 80 85

ttg gaa aac aca ttc att cat aac act ggg aaa aca gtg gaa att aat 462
Leu Glu Asn Thr Phe Ile His Asn Thr Gly Lys Thr Val Glu Ile Asn
90 95 100 105

ctc act aat gac tac cgt gtc agc gga gga gtt tca gaa atg gtg ttt 510
Leu Thr Asn Asp Tyr Arg Val Ser Gly Gly Val Ser Glu Met Val Phe

110	115	120	
aaa gca agc aag ata act ttt cac tgg gga aaa tgc aat atg tca tct			558
Lys Ala Ser Lys Ile Thr Phe His Trp Gly Lys Cys Asn Met Ser Ser			
125	130	135	
gat gga tca gag cat agt tta gaa gga caa aaa ttt cca ctt gag atg			606
Asp Gly Ser Glu His Ser Leu Glu Gly Gln Lys Phe Pro Leu Glu Met			
140	145	150	
caa atc tac tgc ttt gat gcg gac cga ttt tca agt ttt gag gaa gca			654
Gln Ile Tyr Cys Phe Asp Ala Asp Arg Phe Ser Ser Phe Glu Glu Ala			
155	160	165	
gtc aaa gga aaa ggg aag tta aga gct tta tcc att ttg ttt gag gtt			702
Val Lys Gly Lys Gly Lys Leu Arg Ala Leu Ser Ile Leu Phe Glu Val			
170	175	180	185
ggg aca gaa gaa aat ttg gat ttc aaa gcg att att gat gga gtc gaa			750
Gly Thr Glu Glu Asn Leu Asp Phe Lys Ala Ile Ile Asp Gly Val Glu			
190	195	200	
agt gtt agt cgt ttt ggg aag cag gct gct tta gat cca ttc ata ctg			798
Ser Val Ser Arg Phe Gly Lys Gln Ala Ala Leu Asp Pro Phe Ile Leu			
205	210	215	
ttg aac ctt ctg cca aac tca act gac aag tat tac att tac aat ggc			846
Leu Asn Leu Leu Pro Asn Ser Thr Asp Lys Tyr Tyr Ile Tyr Asn Gly			
220	225	230	
tca ttg aca tct cct ccc tgc aca gac aca gtt gac tgg att gtt ttt			894
Ser Leu Thr Ser Pro Pro Cys Thr Asp Thr Val Asp Trp Ile Val Phe			
235	240	245	
aaa gat aca gtt agc atc tct gaa agc cag ttg gct gtt ttt tgt gaa			942
Lys Asp Thr Val Ser Ile Ser Glu Ser Gln Leu Ala Val Phe Cys Glu			
250	255	260	265
gtt ctt aca atg caa caa tct ggt tat gtc atg ctg atg gac tac tta			990
Val Leu Thr Met Gln Gln Ser Gly Tyr Val Met Leu Met Asp Tyr Leu			
270	275	280	

caa aac aat ttt cga gag caa cag tac aag ttc tct aga cag gtg ttt 1038
 Gln Asn Asn Phe Arg Glu Gln Gln Tyr Lys Phe Ser Arg Gln Val Phe
 285 290 295

tcc tca tac act gga aag gaa gag att cat gaa gca gtt tgt agt tca 1086
 Ser Ser Tyr Thr Gly Lys Glu Glu Ile His Glu Ala Val Cys Ser Ser
 300 305 310

gaa cca gaa aat gtt cag gct gac cca gag aat tat acc agc ctt ctt 1134
 Glu Pro Glu Asn Val Gln Ala Asp Pro Glu Asn Tyr Thr Ser Leu Leu
 315 320 325

gtt aca tgg gaa aga cct cga gtc gtt tat gat acc atg att gag aag 1182
 Val Thr Trp Glu Arg Pro Arg Val Val Tyr Asp Thr Met Ile Glu Lys
 330 335 340 345

ttt gca gtt ttg tac cag cag ttg gat gga gag gac caa acc aag cat 1230
 Phe Ala Val Leu Tyr Gln Gln Leu Asp Gly Glu Asp Gln Thr Lys His
 350 355 360

gaa ttt ttg aca gat ggc tat caa gac ttg gta act ata tga 1272
 Glu Phe Leu Thr Asp Gly Tyr Gln Asp Leu Val Thr Ile *
 365 370

tcagttgttt tacatagggg aacattataa tttaatttcc aaggtaagaa cttacaaatg 1332
 gttgtatatt attttcctcc attactttta gactttatgt gaagggtggg taggctgagt 1392
 atttttaaat ttaaaaaaaa atttttaaatt agaagctata ctaaattatg tttaaagtta 1452
 catttaatta aatggatata ataactttgc caacaataac actatagagt agatacatat 1512
 gacttatgaa ctggagatca tttagtgtgg cttttcttaa gatttcagtt gtagaatagt 1572
 gccagaatct cagtgccttg atacatttta tattgtgtct tccattacgc tatatcagca 1632
 caggaaaagt agagtagggg acatacaagt cctctttgtt gcacaaaaaa attttcagat 1692
 aacagctggg aagtcattgat tgggtcagaa ctttggggat gtaagaaaac attttcttaca 1752
 aaaagatcca cccctgcctc cctccaccag cgcattgcga taaagtacag attccctttg 1812
 tggcctgagc atgtcagtat taaactttgc tctggtaggg aagtgttggc catagattag 1872
 ggtgtagttg acaaaccttc atctggatgt aggtccagaa agtccccact gcagggttaa 1932
 ggacactgga ctctgcactc aggcacctag agtcctgcaa gtcctgggaa cctgcattta 1992
 aataaaaaatg cactattaat tatgtttcat atcatgtgga caaaatggat aaaatttttag 2052
 taacctttta attcagttgc ctggaatatg gagacacaat gacctgggaa aatcgtgaaa 2112
 taaatagtaa taaaaatggt tatttcataa ttacgtgaag aagataattc tattactgtt 2172
 cttgcatata tattgtcaag aaaaagagat aacttagttg ttcacttttt cacattgctc 2232

cttgttttgca aatgcccccc atttatttgt ctaaaatatt aatttttagt ttgtagtact 2292
 aatttatgaa tttgatgagt tctggctaaa aatgaaactt cctgaaacta aatctgattt 2352
 ttaaaaagca aaaaaaaaaa aaaagcctag ctttccagtt cttcataatt cacaaatacc 2412
 acaagtttaa ctaagcaaca ttgcataaac ttttccttag gttaataaaa tagaagtatt 2472
 ttccacggac cagggagaaa aagttttcta ggaaagatac ctagtgtgtt ggtagtccta 2532
 tgagaataac atttgtataa ttactaacat ctttctttta ggggtgctatt ctcaataatt 2592
 tgctacccaa tatgagttat gttcttcaga tagtagccat atgcactaat ggcttatatg 2652
 gaaaatacag cgaccaactg attgtcgaca tgcctactga taatcctggg aagtgccacc 2712
 agatacatct atatattaac tcaataaatg aggttagttt aattactgta tgcattgatg 2772
 ctttctctct atattctttt ggccaaaagg caaagtgatt ttctcttaag tctggattgc 2832
 cgggtaattt tttggggcat gggacccatt tctcattcag caggtctggg gccagacaat 2892
 aagtaaactt atccttaata ttggagttaa ccatttgtaa aataagagtg actaaacata 2952
 tttataacat tgtaataatc attaaatgaa aattgctatg taaatgttga gactgttatt 3012
 ttggataatt aagagttggg ttaatttgta tttatttcct cttttcagcc cccaaagcat 3072
 tatgtagtaa gtgtatata 3091

<210> 4

<211> 374

<212> PRT

<213> Homo sapiens

<400> 4

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

```

145          150          155          160
Asp Arg Phe Ser Ser Phe Glu Glu Ala Val Lys Gly Lys Gly Lys Leu
          165          170          175
Arg Ala Leu Ser Ile Leu Phe Glu Val Gly Thr Glu Glu Asn Leu Asp
          180          185          190
Phe Lys Ala Ile Ile Asp Gly Val Glu Ser Val Ser Arg Phe Gly Lys
          195          200          205
Gln Ala Ala Leu Asp Pro Phe Ile Leu Leu Asn Leu Leu Pro Asn Ser
          210          215          220
Thr Asp Lys Tyr Tyr Ile Tyr Asn Gly Ser Leu Thr Ser Pro Pro Cys
225          230          235          240
Thr Asp Thr Val Asp Trp Ile Val Phe Lys Asp Thr Val Ser Ile Ser
          245          250          255
Glu Ser Gln Leu Ala Val Phe Cys Glu Val Leu Thr Met Gln Gln Ser
          260          265          270
Gly Tyr Val Met Leu Met Asp Tyr Leu Gln Asn Asn Phe Arg Glu Gln
          275          280          285
Gln Tyr Lys Phe Ser Arg Gln Val Phe Ser Ser Tyr Thr Gly Lys Glu
          290          295          300
Glu Ile His Glu Ala Val Cys Ser Ser Glu Pro Glu Asn Val Gln Ala
305          310          315          320
Asp Pro Glu Asn Tyr Thr Ser Leu Leu Val Thr Trp Glu Arg Pro Arg
          325          330          335
Val Val Tyr Asp Thr Met Ile Glu Lys Phe Ala Val Leu Tyr Gln Gln
          340          345          350
Leu Asp Gly Glu Asp Gln Thr Lys His Glu Phe Leu Thr Asp Gly Tyr
          355          360          365
Gln Asp Leu Val Thr Ile
          370

```

<210> 5

<211> 8058

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (6229)...(6345)

<223> Exon 23a splice variant of PTP-zeta SM2

<221> misc_feature

<222> (7210)...(8058)

<223> 3' untranslated region

<400> 5

```

cacacatacg cacgcacgat ctccacttcca tctatacact ggaggattaa aacaaacaaa 60
caaaaaaaaaac atttccttcg ctccccctcc ctctccactc tgagaagcag aggagccgca 120
ggcgaggggg ccgcagaccg tctggaaatg cgaatcctaa agcgtttcct cgcttgcat 180
cagctcctct gtgtttgccc cctggattgg gctaattggat actacagaca acagagaaaa 240
cttgttgaag agattggctg gtcctataca ggagcactga atcaaaaaaa ttggggaaag 300
aaatatccaa catgtaatag cccaaaacaa tctcctatca atattgatga agatcttaca 360
caagtaaagt tgaatcttaa gaaacttaaa tttcagggtt gggataaaac atcattggaa 420
aacacattca ttcataacac tgggaaaaca gtggaaatta atctcactaa tgactaccgt 480
gtcagcggag gagtttcaga aatggtgttt aaagcaagca agataacttt tctctgggga 540
aaatgcaata tgtcatctga tggatcagag catagttagg aaggacaaaa atttccactt 600
gagatgcaaa tctactgctt tgatgctggc cgattttcaa gttttgagga agcagtcaaa 660
ggaaaaggga agttaagagc tttatccatt ttgtttgagg ttgggacaga agaaaatttg 720
gatttcaaag cgattattga tggagtcgaa agtggttagtc gttttgggaa gcaggctgct 780
ttagatccat tcatactgtt gaaccttctg ccaaaactcaa ctgacaagta ttacatttac 840
aatggctcat tgacatctcc tccctgcaca gacacagttg actggattgt ttttaaagat 900
acagtttagc tctctgaaaag ccagttggct gttttttgtg aagttcttac aatgcaacaa 960
tctggttatg tcattgctgat ggactactta caaaacaatt ttcgagagca acagtacaag 1020
ttctctagac aggtgttttc ctcatcactc ggaaaggaag agattcatga agcagtttgt 1080
agttcagaac cagaaaatgt tcaggctgac ccagagaatt ataccagcct tcttggtaca 1140
tgggaaagac ctcgagtcgt ttatgatacc atgattgaga agtttgagc tttgtaccag 1200
cagttggatg gagaggacca aaccaagcat gaatttttga cagatggcta tcaagacttg 1260
ggtgctattc tcaataattt gctacccaat atgagttatg ttcttcagat agtagccata 1320
tgcactaatg gcttatatgg aaaatacagc gaccaactga ttgtcgacat gcctactgat 1380
aatcctgaac ttgatctttt ccctgaatta attggaactg aagaaataat caaggaggag 1440
gaagagggaa aagacattga agaaggcgtt attgtgaatc ctggtagaga cagtgttaca 1500
aaccaaatca ggaaaaagga accccagatt tctaccacaa cacactacaa tcgcataggg 1560
acgaaataca atgaagccaa gactaacgca tccccacaa gaggaagtga attctctgga 1620
aagggtgatg ttcccaatac atctttaaat tccacttccc aaccagtcac taaattagcc 1680
acagaaaaag atatttcctt gacttctcag actgtgactg aactgccacc tcactctgtg 1740
gaaggtactt cagcctcttt aaatgatggc tctaaaactg ttcttagatc tccacatatg 1800
aacttgctcg ggactgcaga atccttaaata acagtttcta taacagaata tgaggaggag 1860
agtttattga ccagtttcaa gcttgatact ggagctgaag attcttcagg ctccagtcct 1920
gcaacttctg ctatcccatt catctctgag aacatatccc aagggtatat attttcctcc 1980
gaaaaccag agacaataac atatgatgtc cttataccag aatctgctag aaatgcttcc 2040
gaagattcaa cttcatcagg ttcagaagaa tactaaagg atccttctat ggagggaaat 2100
gtgtgggttc ctagctctac agacataaca gcacagcccg atgttgatc aggagagag 2160

```

agcttttctcc agactaatta cactgagata cgtgttgatg aatctgagaa gacaaccaag 2220
 tccttttctg caggcccagt gatgtcacag ggtccctcag ttacagatct ggaaatgcc 2280
 cattattcta cctttgccta cttcccaact gaggtaacac ctcatgcttt taccatcc 2340
 tccagacaac aggatttggc cttccacggtc aacgtggtat actcgcagac aaccaaccg 2400
 gtatacaatg gtgagacacc tcttcaacct tcctacagta gtgaagtctt tcctctagtc 2460
 acccctttgt tgcttgacaa tcagatcctc aacactaccc ctgctgcttc aagtagtgat 2520
 tcggccttgc atgctacgcc tgtatttccc agtgtcgatg tgtcatttga atccatcctg 2580
 tcttcctatg atgggtgcacc tttgcttcca ttttcctctg cttccttcag tagtgaattg 2640
 tttcgccatc tgcatacagt ttctcaaate cttccacaag ttacttcagc taccgagagt 2700
 gataaggtgc ccttgcatgc ttctctgcca gtggctgggg gtgatttgct attagagccc 2760
 agccttgctc agtattctga tgtgctgtcc actactcatg ctgcttcaga gacgctggaa 2820
 tttggtagtg aatctggtgt tctttataaa acgcttatgt tttctcaagt tgaaccaccc 2880
 agcagtgatg ccatgatgca tgcacgttct tcagggcctg aaccttctta tgccttgtct 2940
 gataatgagg gctcccaaca catcttccact gtttcttaca gttctgcaat acctgtgcat 3000
 gattctgtgg gtgtaactta tcaggggtcc ttatttagcg gccctagcca tataccaata 3060
 cctaagtctt cgttaataac cccaactgca tcattactgc agcctactca tgcctctctt 3120
 ggtgatgggg aatggtctgg agcctcttct gatagtgaat ttcttttacc tgacacagat 3180
 gggctgacag cccttaacat ttcttcacct gtttctgtag ctgaatttac atatacaaca 3240
 tctgtgtttg gtgatgataa taaggcgctt tctaaaagt aaataatata tggaaatgag 3300
 actgaactgc aaattccttc tttcaatgag atgggttacc cttctgaaag cacagtcatg 3360
 cccaacatgt atgataatgt aaataagttg aatgcgtctt tacaagaaac ctctgtttcc 3420
 atttctagca ccaagggcat gtttccaggg tcccttctc ataccaccac taagggtttt 3480
 gatcatgaga ttagtcaagt tccagaaaat aacttttcag ttcaacctac acatactgtc 3540
 tctcaagcat ctggtgacac ttcgcttaaa cctgtgctta gtgcaaaactc agagccagca 3600
 tcctctgacc ctgcttctag tgaaatgtta tctccttcaa ctgagctctt attttatgag 3660
 acctcagctt ctttttagtac tgaagtattg ctacaacctt cctttcaggc ttctgatgtt 3720
 gacaccttgc ttaaaactgt tcttccagct gtgccagtg atccaatatt gggtgaaacc 3780
 ccaaagttg ataaaattag ttctacaatg ttgcatctca ttgtatcaaa ttctgcttca 3840
 agtgaaaaca tgctgcactc tacatctgta ccagtttttg atgtgtcgcc tacttctcat 3900
 atgcactctg cttcacttca aggtttgacc atttctatg caagtgagaa atatgaacca 3960
 gttttgttaa aaagtgaag tttccaccaa gtggtacctt ctttgtacag taatgatgag 4020
 ttgttccaaa cggccaattt ggagattaac caggcccatc ccccaaaagg aaggcatgta 4080
 tttgctacac ctgttttatc aattgatgaa ccattaaata cactaataaa taagcttata 4140
 cattccgatg aaattttaac ctccaccaa agttctgtta ctggtaagggt atttgcgtgg 4200
 attccaacag ttgcttctga tacatttgta tctactgatc attctgttcc tataggaaat 4260
 gggcatgttg ccattacagc tgtttctccc cacagagatg gttctgtaac ctcaacaaag 4320
 ttgctgtttc cttctaaggc aacttctgag ctgagtcata gtgccaaatc tgatgccggg 4380
 ttagtgggtg gtggtgaaga tgggtgacact gatgatgatg gtgatgatga tgatgacaga 4440
 gatagtgatg gcttatccat tcataagtgt atgtcatgct catcctatag agaatacacag 4500
 gaaaaggtaa tgaatgattc agacaccac gaaaacagtc ttatggatca gaataatcca 4560
 atctcact cactatctga gaattctgaa gaagataata gaggcacaag tgtatcctca 4620

gacagtcaaa ctggtatgga cagaagtcct ggtaaatacac catcagcaaa tgggctatcc 4680
caaaagcaca atgatggaaa agaggaaaat gacattcaga ctggtagtgc tctgcttcct 4740
ctcagccctg aatctaaagc atgggcagtt ctgacaagtg atgaagaaag tggatcaggg 4800
caaggtacct cagatagcct taatgagaat gagacttcca cagatttcag ttttgagac 4860
actaatgaaa aagatgctga tgggatcctg gcagcaggtg actcagaaat aactcctgga 4920
ttcccacagt ccccaacatc atctgttact agcgagaact cagaagtgtt ccacgtttca 4980
gaggcagagg ccagtaatag tagccatgag tctcgtattg gtctagctga ggggttgaa 5040
tccgagaaga aggcagttat accccttggtg atcgtgtcag ccctgacttt tatctgtcta 5100
gtggttcttg tgggtattct catctactgg aggaaatgct tccagactgc acacttttac 5160
ttagaggaca gtacatccc tagagttata tccacacctc caacacctat ctttccaatt 5220
tcagatgatg tcggagcaat tccaataaag cactttccaa agcatgttgc agatttacat 5280
gcaagtagtg ggtttactga agaatttgag acactgaaag agttttacca ggaagtgcag 5340
agctgtactg ttgacttagg tattacagca gacagctcca accaccaga caacaagcac 5400
aagaatcgat acataaatat cgttgcctat gatcatagca gggttaagct agcacagctt 5460
gctgaaaagg atggcaaact gactgattat atcaatgcc aattatgttg tggctacaac 5520
agaccaaag cttatattgc tgcccaaggc ccactgaaat ccacagctga agatttctgg 5580
agaatgatat gggaaacataa tgtggaagtt attgtcatga taacaaacct cgtggagaaa 5640
ggaaggagaa aatgtgatca gtactggcct gccgatggga gtgaggagta cgggaacttt 5700
ctggctactc agaagagtgt gcaagtgtt gcctattata ctgtgaggaa ttttactcta 5760
agaaacacaa aaataaaaaa gggtccccag aaaggaagac ccagtggacg tgtggtcaca 5820
cagtatcact acacgcagtg gcctgacatg ggagtaccag agtactccct gccagtgtctg 5880
acctttgtga gaaaggcagc ctatgccaaag cgccatgcag tggggcctgt tgtcgtccac 5940
tgcagtgtctg gagttggaag aacaggcaca tatattgtgc tagacagtat gttgcagcag 6000
attcaacacg aaggaactgt caacatatct ggcttcttaa aacacatccg ttcacaaaga 6060
aattatttgg tacaaactga ggagcaatat gtcttcattc atgatacact ggttgaggcc 6120
atacttagta aagaaactga ggtgctggac agtcatattc atgcctatgt taatgcactc 6180
ctcattcctg gaccagcagg caaaacaaag ctagagaaac aattccaggg tctcactctg 6240
tcacccaggc tggagtgcag aggacaaatc tcggctcact gcaaccttcc tctccctggc 6300
ttaactgatc ctccctacctc agcctcccga gtggctggga ctatactcct gagccagtca 6360
aatatacagc agagtgacta ttctgcagcc ctaaagcaat gcaacagggg aaagaatcga 6420
acttcttcta tcatccctgt ggaaagatca agggttggca tttcatccct gagtggagaa 6480
ggcacagact acatcaatgc ctccctatct atgggctatt accagagcaa tgaattcatc 6540
attaccagc accctctcct tcataccatc aaggatttct ggaggatgat atgggacat 6600
aatgcccaac tgggtggttat gattcctgat ggccaaaaca tggcagaaga tgaatttgtt 6660
tactggccaa ataaagatga gcctataaat tgtgagagct ttaaggtcac tcttatggct 6720
gaagaacaca aatgtctatc taatgaggaa aaacttataa ttcaggactt tatcttagaa 6780
gctacacagg atgattatgt acttgaagtg aggcactttc agtgtcctaa atggccaaat 6840
ccagatagcc ccattagtaa aacttttgaa ctataaagtg ttataaaaga agaagctgcc 6900
aatagggatg ggcctatgat tgttcatgat gagcatggag gagtgcggc aggaactttc 6960
tgtgtcttga caacccttat gcaccaacta gaaaaagaaa attccgtgga tgtttaccag 7020
gtagccaaga tgatcaatct gatgaggcca ggagtctttg ctgacattga gcagtatcag 7080

```

tttctctaca aagtgatcct cagccttgtg agcacaaggc aggaagagaa tccatccacc 7140
tctctggaca gtaatgggtgc agcattgcct gatggaaata tagctgagag cttagagtct 7200
ttagtttaac acagaaaggg gtgggggggac tcacatctga gcattgtttt cctcttccta 7260
aaattaggca ggaaaatcag tctagtcttg ttatctgttg atttcccatc acctgacagt 7320
aactttcatg acataggatt ctgccgcaa atttatatca ttaacaatgt gtgccttttt 7380
gcaagacttg taatttactt attatgtttg aactaaaatg attgaatttt acagtatttc 7440
taagaatgga attgtggtat ttttttctgt attgatttta acagaaaatt tcaatttata 7500
gaggtagga attccaaact acagaaaatg tttgttttta gtgtcaaatt tttagctgta 7560
ttttagcaa ttatcagggt tgctagaaat ataactttta atacagtagc ctgtaaataa 7620
aacactcttc catatgatat tcaacatttt acaactgcag tattcaccta aagtagaaat 7680
aatctgttac ttattgtaaa tactgcccta gtgtctccat ggaccaaatt tatatttata 7740
attgtagatt tttatatattt actactgagt caagttttct agttctgtgt aattgttttag 7800
tttaatgacg tagttcatta gctggtctta ctctaccagt tttctgacat tgtattgtgt 7860
tacctaagtc attaaacttg tttcagcatg taattttaac ttttgtggaa aatagaaata 7920
ccttcatttt gaaagaagtt tttatgagaa taacacctta ccaaacattg ttcaaattgg 7980
ttttatccaa ggaattgcaa aaataaatat aaatattgcc attaaaaaaaa aaaaaaaaaa 8040
aaaaaaaaaa aaaaaaaaaa 8058

```

<210> 6

<211> 2353

<212> PRT

<213> Homo sapiens

<220>

<221> VARSPLIC

<222> (1)...(2353)

<223> PTP-zeta SM2 23a exon variant

<400> 6

```

Met Arg Ile Leu Lys Arg Phe Leu Ala Cys Ile Gln Leu Leu Cys Val
 1             5             10             15
Cys Arg Leu Asp Trp Ala Asn Gly Tyr Tyr Arg Gln Gln Arg Lys Leu
          20             25             30
Val Glu Glu Ile Gly Trp Ser Tyr Thr Gly Ala Leu Asn Gln Lys Asn
          35             40             45
Trp Gly Lys Lys Tyr Pro Thr Cys Asn Ser Pro Lys Gln Ser Pro Ile
          50             55             60
Asn Ile Asp Glu Asp Leu Thr Gln Val Asn Val Asn Leu Lys Lys Leu
65             70             75             80

```

Lys Phe Gln Gly Trp Asp Lys Thr Ser Leu Glu Asn Thr Phe Ile His
 85 90 95
 Asn Thr Gly Lys Thr Val Glu Ile Asn Leu Thr Asn Asp Tyr Arg Val
 100 105 110
 Ser Gly Gly Val Ser Glu Met Val Phe Lys Ala Ser Lys Ile Thr Phe
 115 120 125
 His Trp Gly Lys Cys Asn Met Ser Ser Asp Gly Ser Glu His Ser Leu
 130 135 140
 Glu Gly Gln Lys Phe Pro Leu Glu Met Gln Ile Tyr Cys Phe Asp Ala
 145 150 155 160
 Asp Arg Phe Ser Ser Phe Glu Glu Ala Val Lys Gly Lys Gly Lys Leu
 165 170 175
 Arg Ala Leu Ser Ile Leu Phe Glu Val Gly Thr Glu Glu Asn Leu Asp
 180 185 190
 Phe Lys Ala Ile Ile Asp Gly Val Glu Ser Val Ser Arg Phe Gly Lys
 195 200 205
 Gln Ala Ala Leu Asp Pro Phe Ile Leu Leu Asn Leu Leu Pro Asn Ser
 210 215 220
 Thr Asp Lys Tyr Tyr Ile Tyr Asn Gly Ser Leu Thr Ser Pro Pro Cys
 225 230 235 240
 Thr Asp Thr Val Asp Trp Ile Val Phe Lys Asp Thr Val Ser Ile Ser
 245 250 255
 Glu Ser Gln Leu Ala Val Phe Cys Glu Val Leu Thr Met Gln Gln Ser
 260 265 270
 Gly Tyr Val Met Leu Met Asp Tyr Leu Gln Asn Asn Phe Arg Glu Gln
 275 280 285
 Gln Tyr Lys Phe Ser Arg Gln Val Phe Ser Ser Tyr Thr Gly Lys Glu
 290 295 300
 Glu Ile His Glu Ala Val Cys Ser Ser Glu Pro Glu Asn Val Gln Ala
 305 310 315 320
 Asp Pro Glu Asn Tyr Thr Ser Leu Leu Val Thr Trp Glu Arg Pro Arg
 325 330 335
 Val Val Tyr Asp Thr Met Ile Glu Lys Phe Ala Val Leu Tyr Gln Gln
 340 345 350
 Leu Asp Gly Glu Asp Gln Thr Lys His Glu Phe Leu Thr Asp Gly Tyr
 355 360 365
 Gln Asp Leu Gly Ala Ile Leu Asn Asn Leu Leu Pro Asn Met Ser Tyr
 370 375 380
 Val Leu Gln Ile Val Ala Ile Cys Thr Asn Gly Leu Tyr Gly Lys Tyr
 385 390 395 400
 Ser Asp Gln Leu Ile Val Asp Met Pro Thr Asp Asn Pro Glu Leu Asp

34

Leu Val Ser Thr Val Asn Val Val Tyr Ser Gln Thr Thr Gln Pro Val
 740 745 750
 Tyr Asn Gly Glu Thr Pro Leu Gln Pro Ser Tyr Ser Ser Glu Val Phe
 755 760 765
 Pro Leu Val Thr Pro Leu Leu Leu Asp Asn Gln Ile Leu Asn Thr Thr
 770 775 780
 Pro Ala Ala Ser Ser Ser Asp Ser Ala Leu His Ala Thr Pro Val Phe
 785 790 795 800
 Pro Ser Val Asp Val Ser Phe Glu Ser Ile Leu Ser Ser Tyr Asp Gly
 805 810 815
 Ala Pro Leu Leu Pro Phe Ser Ser Ala Ser Phe Ser Ser Glu Leu Phe
 820 825 830
 Arg His Leu His Thr Val Ser Gln Ile Leu Pro Gln Val Thr Ser Ala
 835 840 845
 Thr Glu Ser Asp Lys Val Pro Leu His Ala Ser Leu Pro Val Ala Gly
 850 855 860
 Gly Asp Leu Leu Leu Glu Pro Ser Leu Ala Gln Tyr Ser Asp Val Leu
 865 870 875 880
 Ser Thr Thr His Ala Ala Ser Glu Thr Leu Glu Phe Gly Ser Glu Ser
 885 890 895
 Gly Val Leu Tyr Lys Thr Leu Met Phe Ser Gln Val Glu Pro Pro Ser
 900 905 910
 Ser Asp Ala Met Met His Ala Arg Ser Ser Gly Pro Glu Pro Ser Tyr
 915 920 925
 Ala Leu Ser Asp Asn Glu Gly Ser Gln His Ile Phe Thr Val Ser Tyr
 930 935 940
 Ser Ser Ala Ile Pro Val His Asp Ser Val Gly Val Thr Tyr Gln Gly
 945 950 955 960
 Ser Leu Phe Ser Gly Pro Ser His Ile Pro Ile Pro Lys Ser Ser Leu
 965 970 975
 Ile Thr Pro Thr Ala Ser Leu Leu Gln Pro Thr His Ala Leu Ser Gly
 980 985 990
 Asp Gly Glu Trp Ser Gly Ala Ser Ser Asp Ser Glu Phe Leu Leu Pro
 995 1000 1005
 Asp Thr Asp Gly Leu Thr Ala Leu Asn Ile Ser Ser Pro Val Ser Val
 1010 1015 1020
 Ala Glu Phe Thr Tyr Thr Thr Ser Val Phe Gly Asp Asp Asn Lys Ala
 1025 1030 1035 1040
 Leu Ser Lys Ser Glu Ile Ile Tyr Gly Asn Glu Thr Glu Leu Gln Ile
 1045 1050 1055
 Pro Ser Phe Asn Glu Met Val Tyr Pro Ser Glu Ser Thr Val Met Pro

1060	1065	1070
Asn Met Tyr Asp Asn Val Asn Lys Leu Asn Ala Ser Leu Gln Glu Thr		
1075	1080	1085
Ser Val Ser Ile Ser Ser Thr Lys Gly Met Phe Pro Gly Ser Leu Ala		
1090	1095	1100
His Thr Thr Thr Lys Val Phe Asp His Glu Ile Ser Gln Val Pro Glu		
1105	1110	1115
Asn Asn Phe Ser Val Gln Pro Thr His Thr Val Ser Gln Ala Ser Gly		
1125	1130	1135
Asp Thr Ser Leu Lys Pro Val Leu Ser Ala Asn Ser Glu Pro Ala Ser		
1140	1145	1150
Ser Asp Pro Ala Ser Ser Glu Met Leu Ser Pro Ser Thr Gln Leu Leu		
1155	1160	1165
Phe Tyr Glu Thr Ser Ala Ser Phe Ser Thr Glu Val Leu Leu Gln Pro		
1170	1175	1180
Ser Phe Gln Ala Ser Asp Val Asp Thr Leu Leu Lys Thr Val Leu Pro		
1185	1190	1195
Ala Val Pro Ser Asp Pro Ile Leu Val Glu Thr Pro Lys Val Asp Lys		
1205	1210	1215
Ile Ser Ser Thr Met Leu His Leu Ile Val Ser Asn Ser Ala Ser Ser		
1220	1225	1230
Glu Asn Met Leu His Ser Thr Ser Val Pro Val Phe Asp Val Ser Pro		
1235	1240	1245
Thr Ser His Met His Ser Ala Ser Leu Gln Gly Leu Thr Ile Ser Tyr		
1250	1255	1260
Ala Ser Glu Lys Tyr Glu Pro Val Leu Leu Lys Ser Glu Ser Ser His		
1265	1270	1275
Gln Val Val Pro Ser Leu Tyr Ser Asn Asp Glu Leu Phe Gln Thr Ala		
1285	1290	1295
Asn Leu Glu Ile Asn Gln Ala His Pro Pro Lys Gly Arg His Val Phe		
1300	1305	1310
Ala Thr Pro Val Leu Ser Ile Asp Glu Pro Leu Asn Thr Leu Ile Asn		
1315	1320	1325
Lys Leu Ile His Ser Asp Glu Ile Leu Thr Ser Thr Lys Ser Ser Val		
1330	1335	1340
Thr Gly Lys Val Phe Ala Gly Ile Pro Thr Val Ala Ser Asp Thr Phe		
1345	1350	1355
Val Ser Thr Asp His Ser Val Pro Ile Gly Asn Gly His Val Ala Ile		
1365	1370	1375
Thr Ala Val Ser Pro His Arg Asp Gly Ser Val Thr Ser Thr Lys Leu		
1380	1385	1390

Leu Phe Pro Ser Lys Ala Thr Ser Glu Leu Ser His Ser Ala Lys Ser
 1395 1400 1405
 Asp Ala Gly Leu Val Gly Gly Gly Glu Asp Gly Asp Thr Asp Asp Asp
 1410 1415 1420
 Gly Asp Asp Asp Asp Asp Arg Asp Ser Asp Gly Leu Ser Ile His Lys
 1425 1430 1435 1440
 Cys Met Ser Cys Ser Ser Tyr Arg Glu Ser Gln Glu Lys Val Met Asn
 1445 1450 1455
 Asp Ser Asp Thr His Glu Asn Ser Leu Met Asp Gln Asn Asn Pro Ile
 1460 1465 1470
 Ser Tyr Ser Leu Ser Glu Asn Ser Glu Glu Asp Asn Arg Val Thr Ser
 1475 1480 1485
 Val Ser Ser Asp Ser Gln Thr Gly Met Asp Arg Ser Pro Gly Lys Ser
 1490 1495 1500
 Pro Ser Ala Asn Gly Leu Ser Gln Lys His Asn Asp Gly Lys Glu Glu
 1505 1510 1515 1520
 Asn Asp Ile Gln Thr Gly Ser Ala Leu Leu Pro Leu Ser Pro Glu Ser
 1525 1530 1535
 Lys Ala Trp Ala Val Leu Thr Ser Asp Glu Glu Ser Gly Ser Gly Gln
 1540 1545 1550
 Gly Thr Ser Asp Ser Leu Asn Glu Asn Glu Thr Ser Thr Asp Phe Ser
 1555 1560 1565
 Phe Ala Asp Thr Asn Glu Lys Asp Ala Asp Gly Ile Leu Ala Ala Gly
 1570 1575 1580
 Asp Ser Glu Ile Thr Pro Gly Phe Pro Gln Ser Pro Thr Ser Ser Val
 1585 1590 1595 1600
 Thr Ser Glu Asn Ser Glu Val Phe His Val Ser Glu Ala Glu Ala Ser
 1605 1610 1615
 Asn Ser Ser His Glu Ser Arg Ile Gly Leu Ala Glu Gly Leu Glu Ser
 1620 1625 1630
 Glu Lys Lys Ala Val Ile Pro Leu Val Ile Val Ser Ala Leu Thr Phe
 1635 1640 1645
 Ile Cys Leu Val Val Leu Val Gly Ile Leu Ile Tyr Trp Arg Lys Cys
 1650 1655 1660
 Phe Gln Thr Ala His Phe Tyr Leu Glu Asp Ser Thr Ser Pro Arg Val
 1665 1670 1675 1680
 Ile Ser Thr Pro Pro Thr Pro Ile Phe Pro Ile Ser Asp Asp Val Gly
 1685 1690 1695
 Ala Ile Pro Ile Lys His Phe Pro Lys His Val Ala Asp Leu His Ala
 1700 1705 1710
 Ser Ser Gly Phe Thr Glu Glu Phe Glu Thr Leu Lys Glu Phe Tyr Gln

1715	1720	1725
Glu Val Gln Ser Cys Thr	Val Asp Leu Gly Ile Thr	Ala Asp Ser Ser
1730	1735	1740
Asn His Pro Asp Asn Lys	His Lys Asn Arg Tyr Ile	Asn Ile Val Ala
1745	1750	1755
Tyr Asp His Ser Arg Val	Lys Leu Ala Gln Leu Ala	Glu Lys Asp Gly
1765	1770	1775
Lys Leu Thr Asp Tyr Ile	Asn Ala Asn Tyr Val Asp	Gly Tyr Asn Arg
1780	1785	1790
Pro Lys Ala Tyr Ile Ala	Ala Gln Gly Pro Leu Lys	Ser Thr Ala Glu
1795	1800	1805
Asp Phe Trp Arg Met Ile	Trp Glu His Asn Val Glu	Val Ile Val Met
1810	1815	1820
Ile Thr Asn Leu Val Glu	Lys Gly Arg Arg Lys Cys	Asp Gln Tyr Trp
1825	1830	1835
Pro Ala Asp Gly Ser Glu	Glu Tyr Gly Asn Phe Leu	Val Thr Gln Lys
1845	1850	1855
Ser Val Gln Val Leu Ala	Tyr Tyr Thr Val Arg Asn	Phe Thr Leu Arg
1860	1865	1870
Asn Thr Lys Ile Lys Lys	Gly Ser Gln Lys Gly Arg	Pro Ser Gly Arg
1875	1880	1885
Val Val Thr Gln Tyr His	Tyr Thr Gln Trp Pro Asp	Met Gly Val Pro
1890	1895	1900
Glu Tyr Ser Leu Pro Val	Leu Thr Phe Val Arg Lys	Ala Ala Tyr Ala
1905	1910	1915
Lys Arg His Ala Val Gly	Pro Val Val Val His Cys	Ser Ala Gly Val
1925	1930	1935
Gly Arg Thr Gly Thr Tyr	Ile Val Leu Asp Ser Met	Leu Gln Gln Ile
1940	1945	1950
Gln His Glu Gly Thr Val	Asn Ile Phe Gly Phe Leu	Lys His Ile Arg
1955	1960	1965
Ser Gln Arg Asn Tyr Leu	Val Gln Thr Glu Glu Gln	Tyr Val Phe Ile
1970	1975	1980
His Asp Thr Leu Val Glu	Ala Ile Leu Ser Lys Glu	Thr Glu Val Leu
1985	1990	1995
Asp Ser His Ile His Ala	Tyr Val Asn Ala Leu Leu	Ile Pro Gly Pro
2005	2010	2015
Ala Gly Lys Thr Lys Leu	Glu Lys Gln Phe Gln Gly	Leu Thr Leu Ser
2020	2025	2030
Pro Arg Leu Glu Cys Arg	Gly Thr Ile Ser Ala His	Cys Asn Leu Pro
2035	2040	2045

Leu Pro Gly Leu Thr Asp Pro Pro Thr Ser Ala Ser Arg Val Ala Gly
 2050 2055 2060
 Thr Ile Leu Leu Ser Gln Ser Asn Ile Gln Gln Ser Asp Tyr Ser Ala
 2065 2070 2075 2080
 Ala Leu Lys Gln Cys Asn Arg Glu Lys Asn Arg Thr Ser Ser Ile Ile
 2085 2090 2095
 Pro Val Glu Arg Ser Arg Val Gly Ile Ser Ser Leu Ser Gly Glu Gly
 2100 2105 2110
 Thr Asp Tyr Ile Asn Ala Ser Tyr Ile Met Gly Tyr Tyr Gln Ser Asn
 2115 2120 2125
 Glu Phe Ile Ile Thr Gln His Pro Leu Leu His Thr Ile Lys Asp Phe
 2130 2135 2140
 Trp Arg Met Ile Trp Asp His Asn Ala Gln Leu Val Val Met Ile Pro
 2145 2150 2155 2160
 Asp Gly Gln Asn Met Ala Glu Asp Glu Phe Val Tyr Trp Pro Asn Lys
 2165 2170 2175
 Asp Glu Pro Ile Asn Cys Glu Ser Phe Lys Val Thr Leu Met Ala Glu
 2180 2185 2190
 Glu His Lys Cys Leu Ser Asn Glu Glu Lys Leu Ile Ile Gln Asp Phe
 2195 2200 2205
 Ile Leu Glu Ala Thr Gln Asp Asp Tyr Val Leu Glu Val Arg His Phe
 2210 2215 2220
 Gln Cys Pro Lys Trp Pro Asn Pro Asp Ser Pro Ile Ser Lys Thr Phe
 2225 2230 2235 2240
 Glu Leu Ile Ser Val Ile Lys Glu Glu Ala Ala Asn Arg Asp Gly Pro
 2245 2250 2255
 Met Ile Val His Asp Glu His Gly Gly Val Thr Ala Gly Thr Phe Cys
 2260 2265 2270
 Ala Leu Thr Thr Leu Met His Gln Leu Glu Lys Glu Asn Ser Val Asp
 2275 2280 2285
 Val Tyr Gln Val Ala Lys Met Ile Asn Leu Met Arg Pro Gly Val Phe
 2290 2295 2300
 Ala Asp Ile Glu Gln Tyr Gln Phe Leu Tyr Lys Val Ile Leu Ser Leu
 2305 2310 2315 2320
 Val Ser Thr Arg Gln Glu Glu Asn Pro Ser Thr Ser Leu Asp Ser Asn
 2325 2330 2335
 Gly Ala Ala Leu Pro Asp Gly Asn Ile Ala Glu Ser Leu Glu Ser Leu
 2340 2345 2350
 Val

<210> 7

<211> 1518

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (22)...(1503)

<223> Coding sequence ARP2

<400> 7

```

aaccaccatt ttgcaaggac c atg agg cca ctg tgc gtg aca tgc tgg tgg      51
                        Met Arg Pro Leu Cys Val Thr Cys Trp Trp
                          1             5             10

ctc gga ctg ctg gct gcc atg gga gct gtt gca ggc cag gag gac ggt      99
Leu Gly Leu Leu Ala Ala Met Gly Ala Val Ala Gly Gln Glu Asp Gly
                      15             20             25

ttt gag ggc act gag gag ggc tcg cca aga gag ttc att tac cta aac     147
Phe Glu Gly Thr Glu Glu Gly Ser Pro Arg Glu Phe Ile Tyr Leu Asn
                      30             35             40

agg tac aag cgg gcg ggc gag tcc cag gac aag tgc acc tac acc ttc     195
Arg Tyr Lys Arg Ala Gly Glu Ser Gln Asp Lys Cys Thr Tyr Thr Phe
                      45             50             55

att gtg ccc cag cag cgg gtc acg ggt gcc atc tgc gtc aac tcc aag     243
Ile Val Pro Gln Gln Arg Val Thr Gly Ala Ile Cys Val Asn Ser Lys
                      60             65             70

gag cct gag gtg ctt ctg gag aac cga gtg cat aag cag gag cta gag     291
Glu Pro Glu Val Leu Leu Glu Asn Arg Val His Lys Gln Glu Leu Glu
                      75             80             85             90

ctg ctc aac aat gag ctg ctc aag cag aag cgg cag atc gag aca ctg     339
Leu Leu Asn Asn Glu Leu Leu Lys Gln Lys Arg Gln Ile Glu Thr Leu
                      95             100            105

cag cag ctg gtg gag gtg gac ggc ggc att gtg agc gag gtg aag ctg     387

```

Gln Gln Leu Val Glu Val Asp Gly Gly Ile Val Ser Glu Val Lys Leu	
110 115 120	
ctg cgc aag gag agc cgc aac atg aac tcg cgg gtc acg cag ctc tac	435
Leu Arg Lys Glu Ser Arg Asn Met Asn Ser Arg Val Thr Gln Leu Tyr	
125 130 135	
atg cag ctc ctg cac gag atc atc cgc aag cgg gac aac gcg ttg gag	483
Met Gln Leu Leu His Glu Ile Ile Arg Lys Arg Asp Asn Ala Leu Glu	
140 145 150	
ctc tcc cag ctg gag aac agg atc ctg aac cag aca gcc gac atg ctg	531
Leu Ser Gln Leu Glu Asn Arg Ile Leu Asn Gln Thr Ala Asp Met Leu	
155 160 165 170	
cag ctg gcc agc aag tac aag gac ctg gag cac aag tac cag cac ctg	579
Gln Leu Ala Ser Lys Tyr Lys Asp Leu Glu His Lys Tyr Gln His Leu	
175 180 185	
gcc aca ctg gcc cac aac caa tca gag atc atc gcg cag ctt gag gag	627
Ala Thr Leu Ala His Asn Gln Ser Glu Ile Ile Ala Gln Leu Glu Glu	
190 195 200	
cac tgc cag agg gtg ccc tcg gcc agg ccc gtc ccc cag cca ccc ccc	675
His Cys Gln Arg Val Pro Ser Ala Arg Pro Val Pro Gln Pro Pro Pro	
205 210 215	
gct gcc ccg ccc cgg gtc tac caa cca ccc acc tac aac cgc atc atc	723
Ala Ala Pro Pro Arg Val Tyr Gln Pro Pro Thr Tyr Asn Arg Ile Ile	
220 225 230	
aac cag atc tct acc aac gag atc cag agt gac cag aac ctg aag gtg	771
Asn Gln Ile Ser Thr Asn Glu Ile Gln Ser Asp Gln Asn Leu Lys Val	
235 240 245 250	
ctg cca ccc cct ctg ccc act atg ccc act ctc acc agc ctc cca tct	819
Leu Pro Pro Pro Leu Pro Thr Met Pro Thr Leu Thr Ser Leu Pro Ser	
255 260 265	
tcc acc gac aag ccg tcg ggc cca tgg aga gac tgc ctg cag gcc ctg	867
Ser Thr Asp Lys Pro Ser Gly Pro Trp Arg Asp Cys Leu Gln Ala Leu	

270	275	280	
gag gat ggc cac gac acc agc tcc atc tac ctg gtg aag ccg gag aac			915
Glu Asp Gly His Asp Thr Ser Ser Ile Tyr Leu Val Lys Pro Glu Asn			
285	290	295	
acc aac cgc ctc atg cag gtg tgg tgc gac cag aga cac gac ccc ggg			963
Thr Asn Arg Leu Met Gln Val Trp Cys Asp Gln Arg His Asp Pro Gly			
300	305	310	
ggc tgg acc gtc atc cag aga cgc ctg gat ggc tct gtt aac ttc ttc			1011
Gly Trp Thr Val Ile Gln Arg Arg Leu Asp Gly Ser Val Asn Phe Phe			
315	320	325	330
agg aac tgg gag acg tac aag caa ggg ttt ggg aac att gat ggc gaa			1059
Arg Asn Trp Glu Thr Tyr Lys Gln Gly Phe Gly Asn Ile Asp Gly Glu			
335	340	345	
tac tgg ctg ggc ctg gag aac att tac tgg ctg acg aac caa ggc aac			1107
Tyr Trp Leu Gly Leu Glu Asn Ile Tyr Trp Leu Thr Asn Gln Gly Asn			
350	355	360	
tac aaa ctc ctg gtg acc atg gag gac tgg tcc ggc cgc aaa gtc ttt			1155
Tyr Lys Leu Leu Val Thr Met Glu Asp Trp Ser Gly Arg Lys Val Phe			
365	370	375	
gca gaa tac gcc agt ttc cgc ctg gaa cct gag agc gag tat tat aag			1203
Ala Glu Tyr Ala Ser Phe Arg Leu Glu Pro Glu Ser Glu Tyr Tyr Lys			
380	385	390	
ctg cgg ctg ggg cgc tac cat ggc aat gcg ggt gac tcc ttt aca tgg			1251
Leu Arg Leu Gly Arg Tyr His Gly Asn Ala Gly Asp Ser Phe Thr Trp			
395	400	405	410
cac aac ggc aag cag ttc acc acc ctg gac aga gat cat gat gtc tac			1299
His Asn Gly Lys Gln Phe Thr Thr Leu Asp Arg Asp His Asp Val Tyr			
415	420	425	
aca gga aac tgt gcc cac tac cag aag gga ggc tgg tgg tat aac gcc			1347
Thr Gly Asn Cys Ala His Tyr Gln Lys Gly Gly Trp Trp Tyr Asn Ala			
430	435	440	

tgt gcc cac tcc aac ctc aac ggg gtc tgg tac cgc ggg ggc cat tac 1395
Cys Ala His Ser Asn Leu Asn Gly Val Trp Tyr Arg Gly Gly His Tyr
445 450 455

cgg agc cgc tac cag gac gga gtc tac tgg gct gag ttc cga gga ggc 1443
Arg Ser Arg Tyr Gln Asp Gly Val Tyr Trp Ala Glu Phe Arg Gly Gly
460 465 470

tct tac tca ctc aag aaa gtg gtg atg atg atc cga ccg aac ccc aac 1491
Ser Tyr Ser Leu Lys Lys Val Val Met Met Ile Arg Pro Asn Pro Asn
475 480 485 490

acc ttc cac taa gccagctccc cctcc 1518
Thr Phe His *

<210> 8

<211> 2133

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (58) ... (969)

<223> SPARC/osteonectin

<400> 8

cgggagagcg cgctctgcct gccgcctgcc tgctgccac tgagggttcc cagcacc atg 60
Met
1

agg gcc tgg atc ttc ttt ctc ctt tgc ctg gcc ggg agg gcc ttg gca 108
Arg Ala Trp Ile Phe Phe Leu Leu Cys Leu Ala Gly Arg Ala Leu Ala
5 10 15

gcc cct cag caa gaa gcc ctg cct gat gag aca gag gtg gtg gaa gaa 156
Ala Pro Gln Gln Glu Ala Leu Pro Asp Glu Thr Glu Val Val Glu Glu
20 25 30

act gtg gca gag gtg act gag gta tct gtg gga gct aat cct gtc cag	204
Thr Val Ala Glu Val Thr Glu Val Ser Val Gly Ala Asn Pro Val Gln	
35 40 45	
gtg gaa gta gga gaa ttt gat gat ggt gca gag gaa acc gaa gag gag	252
Val Glu Val Gly Glu Phe Asp Asp Gly Ala Glu Glu Thr Glu Glu Glu	
50 55 60 65	
gtg gtg gcg gaa aat ccc tgc cag aac cac cac tgc aaa cac ggc aag	300
Val Val Ala Glu Asn Pro Cys Gln Asn His His Cys Lys His Gly Lys	
70 75 80	
gtg tgc gag ctg gat gag aac aac acc ccc atg tgc gtg tgc cag gac	348
Val Cys Glu Leu Asp Glu Asn Asn Thr Pro Met Cys Val Cys Gln Asp	
85 90 95	
ccc acc agc tgc cca gcc ccc att ggc gag ttt gag aag gtg tgc agc	396
Pro Thr Ser Cys Pro Ala Pro Ile Gly Glu Phe Glu Lys Val Cys Ser	
100 105 110	
aat gac aac aag acc ttc gac tct tcc tgc cac ttc ttt gcc aca aag	444
Asn Asp Asn Lys Thr Phe Asp Ser Ser Cys His Phe Phe Ala Thr Lys	
115 120 125	
tgc acc ctg gag ggc acc aag aag ggc cac aag ctc cac ctg gac tac	492
Cys Thr Leu Glu Gly Thr Lys Lys Gly His Lys Leu His Leu Asp Tyr	
130 135 140 145	
atc ggg cct tgc aaa tac atc ccc cct tgc ctg gac tct gag ctg acc	540
Ile Gly Pro Cys Lys Tyr Ile Pro Pro Cys Leu Asp Ser Glu Leu Thr	
150 155 160	
gaa ttc ccc ctg cgc atg cgg gac tgg ctc aag aac gtc ctg gtc acc	588
Glu Phe Pro Leu Arg Met Arg Asp Trp Leu Lys Asn Val Leu Val Thr	
165 170 175	
ctg tat gag agg gat gag gac aac aac ctt ctg act gag aag cag aag	636
Leu Tyr Glu Arg Asp Glu Asp Asn Asn Leu Leu Thr Glu Lys Gln Lys	
180 185 190	
ctg cgg gtg aag aag atc cat gag aat gag aag cgc ctg gag gca gga	684

Leu Arg Val Lys Lys Ile His Glu Asn Glu Lys Arg Leu Glu Ala Gly
 195 200 205

 gac cac ccc gtg gag ctg ctg gcc cgg gac ttc gag aag aac tat aac 732
 Asp His Pro Val Glu Leu Leu Ala Arg Asp Phe Glu Lys Asn Tyr Asn
 210 215 220 225

 atg tac atc ttc cct gta cac tgg cag ttc ggc cag ctg gac cag cac 780
 Met Tyr Ile Phe Pro Val His Trp Gln Phe Gly Gln Leu Asp Gln His
 230 235 240

 ccc att gac ggg tac ctc tcc cac acc gag ctg gct cca ctg cgt gct 828
 Pro Ile Asp Gly Tyr Leu Ser His Thr Glu Leu Ala Pro Leu Arg Ala
 245 250 255

 ccc ctc atc ccc atg gag cat tgc acc acc cgc ttt ttc gag acc tgt 876
 Pro Leu Ile Pro Met Glu His Cys Thr Thr Arg Phe Phe Glu Thr Cys
 260 265 270

 gac ctg gac aat gac aag tac atc gcc ctg gat gag tgg gcc ggc tgc 924
 Asp Leu Asp Asn Asp Lys Tyr Ile Ala Leu Asp Glu Trp Ala Gly Cys
 275 280 285

 ttc ggc atc aag cag aag gat atc gac aag gat ctt gtg atc taa 969
 Phe Gly Ile Lys Gln Lys Asp Ile Asp Lys Asp Leu Val Ile *
 290 295 300

 atccactcct tccacagtac cggattctct ctttaaccct ccccttcgtg tttcccccaa 1029
 tgtttaaaat gtttggtggtg tttgttggtc tgcctggaga caaggtgcta acatagattt 1089
 aagtgaatac attaacggtg ctaaaaatga aaattctaac ccaagacatg acattccttag 1149
 ctgtaactta actattaagg ccttttccac acgcattaat agtcccattt ttctcttgcc 1209
 atttgtagct ttgcccattg tcttattggc acatgggtgg acacggatct gctgggctct 1269
 gccttaaaca cacattgcag cttcaacttt tctctttagt gttctgtttg aaactaatac 1329
 ttaccgagtc agactttgtg ttcatttcat ttcagggtct tggctgcctg tgggcttccc 1389
 caggtggcct ggaggtgggc aaaggaagt aacagacaca cgatgttgtc aaggatgggt 1449
 ttgggactag aggtcagtg gtgggagaga tcctgcaga atccaccaac cagaacgtgg 1509
 tttgcctgag gctgtaactg agagaaagat tctggggctg tcttatgaaa atatagacat 1569
 tctcacataa gccagttca tcaccatttc ctctttacc tttcagtga gtttcttttc 1629
 acattaggct gttggttcaa acttttggga gcacggactg tcagttctct gggaagtgg 1689
 cagcgcatcc tgcagggtt ctctctctct gtcttttggga gaaccagggc tcttctcagg 1749
 ggctctaggg actgccaggc tgtttcagcc aggaaggcca aaatcaagag tgagatgtag 1809

```

aaagttgtaa aatagaaaaa gtggagttgg tgaatcggtt gttctttcct cacatttgga 1869
tgattgtcat aagggttttta gcatgttcct ccttttcttc accctcccct ttgttcttct 1929
attaatcaag agaaacttca aagttaatgg gatggtcgga tctcacaggc tgagaactcg 1989
ttcacctcca agcatttcat gaaaaagctg cttcttatta atcatacaaa ctctcaccat 2049
gatgtgaaga gtttcacaaa tctttcaaaa taaaaagtaa tgacttagaa actgaaaaaa 2109
aaaaaaaaaa aaaaaaaaaa aaaa                                     2133

```

<210> 9

<211> 4586

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (195)...(4367)

<223> C-Met Proto Oncogene

<400> 9

```

gaattccgcc ctcgccgcc gcggcgcccc gagcgctttg tgagcagatg cggagccgag 60
tggagggcgc gagccagatg cggggcgaca gctgacttgc tgagaggagg cggggaggcg 120
cggagcgcgc gtgtggtcct tgcgccgctg acttctccac tggttcctgg gcaccgaaag 180
ataaacctct cata atg aag gcc ccc gct gtg ctt gca cct ggc atc ctc 230
      Met Lys Ala Pro Ala Val Leu Ala Pro Gly Ile Leu
              1              5              10

```

```

gtg ctc ctg ttt acc ttg gtg cag agg agc aat ggg gag tgt aaa gag 278
Val Leu Leu Phe Thr Leu Val Gln Arg Ser Asn Gly Glu Cys Lys Glu
      15              20              25

```

```

gca cta gca aag tcc gag atg aat gtg aat atg aag tat cag ctt ccc 326
Ala Leu Ala Lys Ser Glu Met Asn Val Asn Met Lys Tyr Gln Leu Pro
      30              35              40

```

```

aac ttc acc gcg gaa aca ccc atc cag aat gtc att cta cat gag cat 374
Asn Phe Thr Ala Glu Thr Pro Ile Gln Asn Val Ile Leu His Glu His
      45              50              55              60

```

```

cac att ttc ctt ggt gcc act aac tac att tat gtt tta aat gag gaa 422
His Ile Phe Leu Gly Ala Thr Asn Tyr Ile Tyr Val Leu Asn Glu Glu
      65              70              75

```

gac ctt cag aag gtt gct gag tac aag act ggg cct gtg ctg gaa cac	470
Asp Leu Gln Lys Val Ala Glu Tyr Lys Thr Gly Pro Val Leu Glu His	
80 85 90	
cca gat tgt ttc cca tgt cag gac tgc agc agc aaa gcc aat tta tca	518
Pro Asp Cys Phe Pro Cys Gln Asp Cys Ser Ser Lys Ala Asn Leu Ser	
95 100 105	
gga ggt gtt tgg aaa gat aac atc aac atg gct cta gtt gtc gac acc	566
Gly Gly Val Trp Lys Asp Asn Ile Asn Met Ala Leu Val Val Asp Thr	
110 115 120	
tac tat gat gat caa ctc att agc tgt ggc agc gtc aac aga ggg acc	614
Tyr Tyr Asp Asp Gln Leu Ile Ser Cys Gly Ser Val Asn Arg Gly Thr	
125 130 135 140	
tgc cag cga cat gtc ttt ccc cac aat cat act gct gac ata cag tcg	662
Cys Gln Arg His Val Phe Pro His Asn His Thr Ala Asp Ile Gln Ser	
145 150 155	
gag gtt cac tgc ata ttc tcc cca cag ata gaa gag ccc agc cag tgt	710
Glu Val His Cys Ile Phe Ser Pro Gln Ile Glu Glu Pro Ser Gln Cys	
160 165 170	
cct gac tgt gtg gtg agc gcc ctg gga gcc aaa gtc ctt tca tct gta	758
Pro Asp Cys Val Val Ser Ala Leu Gly Ala Lys Val Leu Ser Ser Val	
175 180 185	
aag gac cgg ttc atc aac ttc ttt gta ggc aat acc ata aat tct tct	806
Lys Asp Arg Phe Ile Asn Phe Phe Val Gly Asn Thr Ile Asn Ser Ser	
190 195 200	
tat ttc cca gat cat cca ttg cat tcg ata tca gtg aga agg cta aag	854
Tyr Phe Pro Asp His Pro Leu His Ser Ile Ser Val Arg Arg Leu Lys	
205 210 215 220	
gaa acg aaa gat ggt ttt atg ttt ttg acg gac cag tcc tac att gat	902
Glu Thr Lys Asp Gly Phe Met Phe Leu Thr Asp Gln Ser Tyr Ile Asp	
225 230 235	
gtt tta cct gag ttc aga gat tct tac ccc att aag tat gtc cat gcc	950

Val	Leu	Pro	Glu	Phe	Arg	Asp	Ser	Tyr	Pro	Ile	Lys	Tyr	Val	His	Ala		
			240					245					250				
ttt	gaa	agc	aac	aat	ttt	att	tac	ttc	ttg	acg	gtc	caa	agg	gaa	act	998	
Phe	Glu	Ser	Asn	Asn	Phe	Ile	Tyr	Phe	Leu	Thr	Val	Gln	Arg	Glu	Thr		
			255					260					265				
cta	gat	gct	cag	act	ttt	cac	aca	aga	ata	atc	agg	ttc	tgt	tcc	ata	1046	
Leu	Asp	Ala	Gln	Thr	Phe	His	Thr	Arg	Ile	Ile	Arg	Phe	Cys	Ser	Ile		
			270					275					280				
aac	tct	gga	ttg	cat	tcc	tac	atg	gaa	atg	cct	ctg	gag	tgt	att	ctc	1094	
Asn	Ser	Gly	Leu	His	Ser	Tyr	Met	Glu	Met	Pro	Leu	Glu	Cys	Ile	Leu		
285						290				295					300		
aca	gaa	aag	aga	aaa	aag	aga	tcc	aca	aag	aag	gaa	gtg	ttt	aat	ata	1142	
Thr	Glu	Lys	Arg	Lys	Lys	Arg	Ser	Thr	Lys	Lys	Glu	Val	Phe	Asn	Ile		
				305						310					315		
ctt	cag	gct	gcg	tat	gtc	agc	aag	cct	ggg	gcc	cag	ctt	gct	aga	caa	1190	
Leu	Gln	Ala	Ala	Tyr	Val	Ser	Lys	Pro	Gly	Ala	Gln	Leu	Ala	Arg	Gln		
				320						325					330		
ata	gga	gcc	agc	ctg	aat	gat	gac	att	ctt	ttc	ggg	gtg	ttc	gca	caa	1238	
Ile	Gly	Ala	Ser	Leu	Asn	Asp	Asp	Ile	Leu	Phe	Gly	Val	Phe	Ala	Gln		
				335						340					345		
agc	aag	cca	gat	tct	gcc	gaa	cca	atg	gat	cga	tct	gcc	atg	tgt	gca	1286	
Ser	Lys	Pro	Asp	Ser	Ala	Glu	Pro	Met	Asp	Arg	Ser	Ala	Met	Cys	Ala		
				350						355					360		
ttc	cct	atc	aaa	tat	gtc	aac	gac	ttc	ttc	aac	aag	atc	gtc	aac	aaa	1334	
Phe	Pro	Ile	Lys	Tyr	Val	Asn	Asp	Phe	Phe	Asn	Lys	Ile	Val	Asn	Lys		
365						370					375				380		
aac	aat	gtg	aga	tgt	ctc	cag	cat	ttt	tac	gga	ccc	aat	cat	gag	cac	1382	
Asn	Asn	Val	Arg	Cys	Leu	Gln	His	Phe	Tyr	Gly	Pro	Asn	His	Glu	His		
						385					390				395		
tgc	ttt	aat	agg	aca	ctt	ctg	aga	aat	tca	tca	ggc	tgt	gaa	gcg	cgc	1430	
Cys	Phe	Asn	Arg	Thr	Leu	Leu	Arg	Asn	Ser	Ser	Gly	Cys	Glu	Ala	Arg		

400	405	410	
cgt gat gaa tat cga aca gag ttt acc aca gct ttg cag cgc gtt gac			1478
Arg Asp Glu Tyr Arg Thr Glu Phe Thr Thr Ala Leu Gln Arg Val Asp			
415	420	425	
tta ttc atg ggt caa ttc agc gaa gtc ctc tta aca tct ata tcc acc			1526
Leu Phe Met Gly Gln Phe Ser Glu Val Leu Leu Thr Ser Ile Ser Thr			
430	435	440	
ttc att aaa gga gac ctc acc ata gct aat ctt ggg aca tca gag ggt			1574
Phe Ile Lys Gly Asp Leu Thr Ile Ala Asn Leu Gly Thr Ser Glu Gly			
445	450	455	460
cgc ttc atg cag gtt gtg gtt tct cga tca gga cca tca acc cct cat			1622
Arg Phe Met Gln Val Val Val Ser Arg Ser Gly Pro Ser Thr Pro His			
465	470	475	
gtg aat ttt ctc ctg gac tcc cat cca gtg tct cca gaa gtg att gtg			1670
Val Asn Phe Leu Leu Asp Ser His Pro Val Ser Pro Glu Val Ile Val			
480	485	490	
gag cat aca tta aac caa aat ggc tac aca ctg gtt atc act ggg aag			1718
Glu His Thr Leu Asn Gln Asn Gly Tyr Thr Leu Val Ile Thr Gly Lys			
495	500	505	
aag atc acg aag atc cca ttg aat ggc ttg ggc tgc aga cat ttc cag			1766
Lys Ile Thr Lys Ile Pro Leu Asn Gly Leu Gly Cys Arg His Phe Gln			
510	515	520	
tcc tgc agt caa tgc ctc tct gcc cca ccc ttt gtt cag tgt ggc tgg			1814
Ser Cys Ser Gln Cys Leu Ser Ala Pro Pro Phe Val Gln Cys Gly Trp			
525	530	535	540
tgc cac gac aaa tgt gtg cga tcg gag gaa tgc ctg agc ggg aca tgg			1862
Cys His Asp Lys Cys Val Arg Ser Glu Glu Cys Leu Ser Gly Thr Trp			
545	550	555	
act caa cag atc tgt ctg cct gca atc tac aag gtt ttc cca aat agt			1910
Thr Gln Gln Ile Cys Leu Pro Ala Ile Tyr Lys Val Phe Pro Asn Ser			
560	565	570	

gca ccc ctt gaa gga ggg aca agg ctg acc ata tgt ggc tgg gac ttt	1958
Ala Pro Leu Glu Gly Gly Thr Arg Leu Thr Ile Cys Gly Trp Asp Phe	
575 580 585	
gga ttt cgg agg aat aat aaa ttt gat tta aag aaa act aga gtt ctc	2006
Gly Phe Arg Arg Asn Asn Lys Phe Asp Leu Lys Lys Thr Arg Val Leu	
590 595 600	
ctt gga aat gag agc tgc acc ttg act tta agt gag agc acg atg aat	2054
Leu Gly Asn Glu Ser Cys Thr Leu Thr Leu Ser Glu Ser Thr Met Asn	
605 610 615 620	
aca ttg aaa tgc aca gtt ggt cct gcc atg aat aag cat ttc aat atg	2102
Thr Leu Lys Cys Thr Val Gly Pro Ala Met Asn Lys His Phe Asn Met	
625 630 635	
tcc ata att att tca aat ggc cac ggg aca aca caa tac agt aca ttc	2150
Ser Ile Ile Ile Ser Asn Gly His Gly Thr Thr Gln Tyr Ser Thr Phe	
640 645 650	
tcc tat gtg gat cct gta ata aca agt att tcg ccg aaa tac ggt cct	2198
Ser Tyr Val Asp Pro Val Ile Thr Ser Ile Ser Pro Lys Tyr Gly Pro	
655 660 665	
atg gct ggt ggc act tta ctt act tta act gga aat tac cta aac agt	2246
Met Ala Gly Gly Thr Leu Leu Thr Leu Thr Gly Asn Tyr Leu Asn Ser	
670 675 680	
ggg aat tct aga cac att tca att ggt gga aaa aca tgt act tta aaa	2294
Gly Asn Ser Arg His Ile Ser Ile Gly Gly Lys Thr Cys Thr Leu Lys	
685 690 695 700	
agt gtg tca aac agt att ctt gaa tgt tat acc cca gcc caa acc att	2342
Ser Val Ser Asn Ser Ile Leu Glu Cys Tyr Thr Pro Ala Gln Thr Ile	
705 710 715	
tca act gag ttt gct gtt aaa ttg aaa att gac tta gcc aac cga gag	2390
Ser Thr Glu Phe Ala Val Lys Leu Lys Ile Asp Leu Ala Asn Arg Glu	
720 725 730	

aca agc atc ttc agt tac cgt gaa gat ccc att gtc tat gaa att cat	2438
Thr Ser Ile Phe Ser Tyr Arg Glu Asp Pro Ile Val Tyr Glu Ile His	
735 740 745	
cca acc aaa tct ttt att agt ggt ggg agc aca ata aca ggt gtt ggg	2486
Pro Thr Lys Ser Phe Ile Ser Gly Gly Ser Thr Ile Thr Gly Val Gly	
750 755 760	
aaa aac ctg aat tca gtt agt gtc ccg aga atg gtc ata aat gtg cat	2534
Lys Asn Leu Asn Ser Val Ser Val Pro Arg Met Val Ile Asn Val His	
765 770 775 780	
gaa gca gga agg aac ttt aca gtg gca tgt caa cat cgc tct aat tca	2582
Glu Ala Gly Arg Asn Phe Thr Val Ala Cys Gln His Arg Ser Asn Ser	
785 790 795	
gag ata atc tgt tgt acc act cct tcc ctg caa cag ctg aat ctg caa	2630
Glu Ile Ile Cys Cys Thr Thr Pro Ser Leu Gln Gln Leu Asn Leu Gln	
800 805 810	
ctc ccc ctg aaa acc aaa gcc ttt ttc atg tta gat ggg atc ctt tcc	2678
Leu Pro Leu Lys Thr Lys Ala Phe Phe Met Leu Asp Gly Ile Leu Ser	
815 820 825	
aaa tac ttt gat ctc att tat gta cat aat cct gtg ttt aag cct ttt	2726
Lys Tyr Phe Asp Leu Ile Tyr Val His Asn Pro Val Phe Lys Pro Phe	
830 835 840	
gaa aag cca gtg atg atc tca atg ggc aat gaa aat gta ctg gaa att	2774
Glu Lys Pro Val Met Ile Ser Met Gly Asn Glu Asn Val Leu Glu Ile	
845 850 855 860	
aag gga aat gat att gac cct gaa gca gtt aaa ggt gaa gtg tta aaa	2822
Lys Gly Asn Asp Ile Asp Pro Glu Ala Val Lys Gly Glu Val Leu Lys	
865 870 875	
gtt gga aat aag agc tgt gag aat ata cac tta cat tct gaa gcc gtt	2870
Val Gly Asn Lys Ser Cys Glu Asn Ile His Leu His Ser Glu Ala Val	
880 885 890	
tta tgc acg gtc ccc aat gac ctg ctg aaa ttg aac agc gag cta aat	2918

Leu Cys Thr Val Pro Asn Asp	Leu Leu Lys Leu Asn Ser Glu Leu Asn	
895	900	905
ata gag tgg aag caa gca att tct tca acc gtc ctt gga aaa gta ata		2966
Ile Glu Trp Lys Gln Ala Ile Ser Ser Thr Val Leu Gly Lys Val Ile		
910	915	920
gtt caa cca gat cag aat ttc aca gga ttg att gct ggt gtt gtc tca		3014
Val Gln Pro Asp Gln Asn Phe Thr Gly Leu Ile Ala Gly Val Val Ser		
925	930	935 940
ata tca aca gca ctg tta tta cta ctt ggg ttt ttc ctg tgg ctg aaa		3062
Ile Ser Thr Ala Leu Leu Leu Leu Leu Gly Phe Phe Leu Trp Leu Lys		
945	950	955
aag aga aag caa att aaa gat ctg ggc agt gaa tta gtt cgc tac gat		3110
Lys Arg Lys Gln Ile Lys Asp Leu Gly Ser Glu Leu Val Arg Tyr Asp		
960	965	970
gca aga gta cac act cct cat ttg gat agg ctt gta agt gcc cga agt		3158
Ala Arg Val His Thr Pro His Leu Asp Arg Leu Val Ser Ala Arg Ser		
975	980	985
gta agc cca act aca gaa atg gtt tca aat gaa tct gta gac tac cga		3206
Val Ser Pro Thr Thr Glu Met Val Ser Asn Glu Ser Val Asp Tyr Arg		
990	995	1000
gct act ttt cca gaa gat cag ttt cct aat tca tct cag aac ggt tca		3254
Ala Thr Phe Pro Glu Asp Gln Phe Pro Asn Ser Ser Gln Asn Gly Ser		
1005	1010	1015 1020
tgc cga caa gtg cag tat cct ctg aca gac atg tcc ccc atc cta act		3302
Cys Arg Gln Val Gln Tyr Pro Leu Thr Asp Met Ser Pro Ile Leu Thr		
1025	1030	1035
agt ggg gac tct gat ata tcc agt cca tta ctg caa aat act gtc cac		3350
Ser Gly Asp Ser Asp Ile Ser Ser Pro Leu Leu Gln Asn Thr Val His		
1040	1045	1050
att gac ctc agt gct cta aat cca gag ctg gtc cag gca gtg cag cat		3398
Ile Asp Leu Ser Ala Leu Asn Pro Glu Leu Val Gln Ala Val Gln His		

1055	1060	1065	
gta gtg att ggg ccc agt agc ctg att gtg cat ttc aat gaa gtc ata			3446
Val Val Ile Gly Pro Ser Ser Leu Ile Val His Phe Asn Glu Val Ile			
1070	1075	1080	
gga aga ggg cat ttt ggt tgt gta tat cat ggg act ttg ttg gac aat			3494
Gly Arg Gly His Phe Gly Cys Val Tyr His Gly Thr Leu Leu Asp Asn			
1085	1090	1095	1100
gat ggc aag aaa att cac tgt gct gtg aaa tcc ttg aac aga atc act			3542
Asp Gly Lys Lys Ile His Cys Ala Val Lys Ser Leu Asn Arg Ile Thr			
	1105	1110	1115
gac ata gga gaa gtt tcc caa ttt ctg acc gag gga atc atc atg aaa			3590
Asp Ile Gly Glu Val Ser Gln Phe Leu Thr Glu Gly Ile Ile Met Lys			
	1120	1125	1130
gat ttt agt cat ccc aat gtc ctc tcg ctc ctg gga atc tgc ctg cga			3638
Asp Phe Ser His Pro Asn Val Leu Ser Leu Leu Gly Ile Cys Leu Arg			
	1135	1140	1145
agt gaa ggg tct ccg ctg gtg gtc cta cca tac atg aaa cat gga gat			3686
Ser Glu Gly Ser Pro Leu Val Val Leu Pro Tyr Met Lys His Gly Asp			
1150	1155	1160	
ctt cga aat ttc att cga aat gag act cat aat cca act gta aaa gat			3734
Leu Arg Asn Phe Ile Arg Asn Glu Thr His Asn Pro Thr Val Lys Asp			
1165	1170	1175	1180
ctt att ggc ttt ggt ctt caa gta gcc aaa ggc atg aaa tat ctt gca			3782
Leu Ile Gly Phe Gly Leu Gln Val Ala Lys Gly Met Lys Tyr Leu Ala			
	1185	1190	1195
agc aaa aag ttt gtc cac aga gac ttg gct gca aga aac tgt atg ctg			3830
Ser Lys Lys Phe Val His Arg Asp Leu Ala Ala Arg Asn Cys Met Leu			
1200	1205	1210	
gat gaa aaa ttc aca gtc aag gtt gct gat ttt ggt ctt gcc aga gac			3878
Asp Glu Lys Phe Thr Val Lys Val Ala Asp Phe Gly Leu Ala Arg Asp			
1215	1220	1225	

atg tat gat aaa gaa tac tat agt gta cac aac aaa aca ggt gca aag	3926
Met Tyr Asp Lys Glu Tyr Tyr Ser Val His Asn Lys Thr Gly Ala Lys	
1230 1235 1240	
ctg cca gtg aag tgg atg gct ttg gaa agt ctg caa act caa aag ttt	3974
Leu Pro Val Lys Trp Met Ala Leu Glu Ser Leu Gln Thr Gln Lys Phe	
1245 1250 1255 1260	
acc acc aag tca gat gtg tgg tcc ttt ggc gtc gtc ctc tgg gag ctg	4022
Thr Thr Lys Ser Asp Val Trp Ser Phe Gly Val Val Leu Trp Glu Leu	
1265 1270 1275	
atg aca aga gga gcc cca cct tat cct gac gta aac acc ttt gat ata	4070
Met Thr Arg Gly Ala Pro Pro Tyr Pro Asp Val Asn Thr Phe Asp Ile	
1280 1285 1290	
act gtt tac ttg ttg caa ggg aga aga ctc cta caa ccc gaa tac tgc	4118
Thr Val Tyr Leu Leu Gln Gly Arg Arg Leu Leu Gln Pro Glu Tyr Cys	
1295 1300 1305	
cca gac ccc tta tat gaa gta atg cta aaa tgc tgg cac cct aaa gcc	4166
Pro Asp Pro Leu Tyr Glu Val Met Leu Lys Cys Trp His Pro Lys Ala	
1310 1315 1320	
gaa atg cgc cca tcc ttt tct gaa ctg gtg tcc cgg ata tca gcg atc	4214
Glu Met Arg Pro Ser Phe Ser Glu Leu Val Ser Arg Ile Ser Ala Ile	
1325 1330 1335 1340	
ttc tct act ttc att ggg gag cac tat gtc cat gtg aac gct act tat	4262
Phe Ser Thr Phe Ile Gly Glu His Tyr Val His Val Asn Ala Thr Tyr	
1345 1350 1355	
gtg aac gta aaa tgt gtc gct ccg tat cct tct ctg ttg tca tca gaa	4310
Val Asn Val Lys Cys Val Ala Pro Tyr Pro Ser Leu Leu Ser Ser Glu	
1360 1365 1370	
gat aac gct gat gat gag gtg gac aca cga cca gcc tcc ttc tgg gag	4358
Asp Asn Ala Asp Asp Glu Val Asp Thr Arg Pro Ala Ser Phe Trp Glu	
1375 1380 1385	

aca tca tag tgctagtact atgtcaaagc aacagtccac actttgtcca 4407

Thr Ser *

1390

atgggtttttt cactgcctga cctttaaaag gccatcgata ttctttgctc cttgccaaat 4467

tgcactatta ataggacttg tattgttatt taaattactg gattctaagg aattttcttat 4527

ctgacagagc atcagaacca gaggtttggt cccacaggcc agggaccaat gcgctgcag 4586

<210> 10

<211> 1807

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (29)...(1114)

<223> human CD44

<400> 10

cccgcgccct ccgttcgctc cggacacc atg gac aag ttt tgg tgg cac gca 52

Met Asp Lys Phe Trp Trp His Ala

1

5

gcc tgg gga ctc tgc ctc gtg ccg ctg agc ctg gcg cag atc gat ttg 100

Ala Trp Gly Leu Cys Leu Val Pro Leu Ser Leu Ala Gln Ile Asp Leu

10

15

20

aat ata acc tgc cgc ttt gca ggt gta ttc cac gtg gag aaa aat ggt 148

Asn Ile Thr Cys Arg Phe Ala Gly Val Phe His Val Glu Lys Asn Gly

25

30

35

40

cgc tac agc atc tct cgg acg gag gcc gct gac ctc tgc aag gct ttc 196

Arg Tyr Ser Ile Ser Arg Thr Glu Ala Ala Asp Leu Cys Lys Ala Phe

45

50

55

aat agc acc ttg ccc aca atg gcc cag atg gag aaa gct ctg agc atc 244

Asn Ser Thr Leu Pro Thr Met Ala Gln Met Glu Lys Ala Leu Ser Ile

60

65

70

gga ttt gag acc tgc agg tat ggg ttc ata gaa ggg cac gtg gtg att 292

Gly Phe Glu Thr Cys Arg Tyr Gly Phe Ile Glu Gly His Val Val Ile

75	80	85	
ccc cgg atc cac ccc aac tcc atc tgt gca gca aac aac aca ggg gtg			340
Pro Arg Ile His Pro Asn Ser Ile Cys Ala Ala Asn Asn Thr Gly Val			
90	95	100	
tac atc ctc aca tcc aac acc tcc cag tat gac aca tat tgc ttc aat			388
Tyr Ile Leu Thr Ser Asn Thr Ser Gln Tyr Asp Thr Tyr Cys Phe Asn			
105	110	115	120
gct tca gct cca cct gaa gaa gat tgt aca tca gtc aca gac ctg ccc			436
Ala Ser Ala Pro Pro Glu Glu Asp Cys Thr Ser Val Thr Asp Leu Pro			
125	130	135	
aat gcc ttt gat gga cca att acc ata act att gtt aac cgt gat ggc			484
Asn Ala Phe Asp Gly Pro Ile Thr Ile Thr Ile Val Asn Arg Asp Gly			
140	145	150	
acc cgc tat gtc cag aaa gga gaa tac aga acg aat cct gaa gac atc			532
Thr Arg Tyr Val Gln Lys Gly Glu Tyr Arg Thr Asn Pro Glu Asp Ile			
155	160	165	
tac ccc agc aac cct act gat gat gac gtg agc agc ggc tcc tcc agt			580
Tyr Pro Ser Asn Pro Thr Asp Asp Asp Val Ser Ser Gly Ser Ser Ser			
170	175	180	
gaa agg agc agc act tca gga ggt tac atc ttt tac acc ttt tct act			628
Glu Arg Ser Ser Thr Ser Gly Gly Tyr Ile Phe Tyr Thr Phe Ser Thr			
185	190	195	200
gta cac ccc atc cca gac gaa gac agt ccc tgg atc acc gac agc aca			676
Val His Pro Ile Pro Asp Glu Asp Ser Pro Trp Ile Thr Asp Ser Thr			
205	210	215	
gac aga atc cct gct acc aga gac caa gac aca ttc cac ccc agt ggg			724
Asp Arg Ile Pro Ala Thr Arg Asp Gln Asp Thr Phe His Pro Ser Gly			
220	225	230	
ggg tcc cat acc act cat gga tct gaa tca gat gga cac tca cat ggg			772
Gly Ser His Thr Thr His Gly Ser Glu Ser Asp Gly His Ser His Gly			
235	240	245	

```

agt caa gaa ggt gga gca aac aca acc tct ggt cct ata agg aca ccc      820
Ser Gln Glu Gly Gly Ala Asn Thr Thr Ser Gly Pro Ile Arg Thr Pro
      250              255              260

caa att cca gaa tgg ctg atc atc ttg gca tcc ctc ttg gcc ttg gct      868
Gln Ile Pro Glu Trp Leu Ile Ile Leu Ala Ser Leu Leu Ala Leu Ala
      265              270              275              280

ttg att ctt gca gtt tgc att gca gtc aac agt cga aga agg tgt ggg      916
Leu Ile Leu Ala Val Cys Ile Ala Val Asn Ser Arg Arg Arg Cys Gly
              285              290              295

cag aag aaa aag cta gtg atc aac agt ggc aat gga gct gtg gag gac      964
Gln Lys Lys Lys Leu Val Ile Asn Ser Gly Asn Gly Ala Val Glu Asp
              300              305              310

aga aag cca agt gga ctc aac gga gag gcc agc aag tct cag gaa atg      1012
Arg Lys Pro Ser Gly Leu Asn Gly Glu Ala Ser Lys Ser Gln Glu Met
              315              320              325

gtg cat ttg gtg aac aag gag tcg tca gaa act cca gac cag ttt atg      1060
Val His Leu Val Asn Lys Glu Ser Ser Glu Thr Pro Asp Gln Phe Met
              330              335              340

aca gct gat gag aca agg aac ctg cag aat gtg gac atg aag att ggg      1108
Thr Ala Asp Glu Thr Arg Asn Leu Gln Asn Val Asp Met Lys Ile Gly
      345              350              355              360

gtg taa cacctacacc attatcttgg aaagaaacaa ccgttggaata cataaccatt      1164
Val  *

```

```

acagggagct gggacactta acagatgcaa tgtgctactg attgtttcat tgcgaatctt 1224
tttttagcata aaatcttcta ctctttttgt tttttgtgtt ttgttcttta aagtcaggtc 1284
caatttgtaa aaacagcatt gctttctgaa attagggccc aattaataat cagcaagaat 1344
ttgatcggtc cagttccac ttggaggcct ttcattccctc ggggtgtgcta tggatggctt 1404
ctaacaaaaa ctacacatat gtattcctga tcgccaacct ttccccacc agctaaggac 1464
atttcccagg gttaataggg cctgggtccct gggaggaaat ttgaatgggt ccattttgcc 1524
cttccatagc ctaatccctg ggcattgctt tccactgagg ttgggggttg ggggtgtacta 1584
gttacacatc ttcaacagac cccctctaga aatttttcag atgcttcttg gagacacca 1644

```

aaggggtgaag ctatttatct gtagtaaact atttatctgt gtttttgaaa tattaacccc 1704
 tggatcagtc ctttgatcag tataatTTTT taaagttact ttgtcagagg cacaaaaggg 1764
 tttaaactga ttcataataa atatctgtac ttcttcgatc ttc 1807

<210> 11

<211> 3366

<212> DNA

<213> Rattus norvegicus

<220>

<221> CDS

<222> (1)...(1547)

<400> 11

tgg aga aag ctc tga gca tcg gat ttg aga cct gca ggt atg ggt tca 48
 Trp Arg Lys Leu * Ala Ser Asp Leu Arg Pro Ala Gly Met Gly Ser
 1 5 10 15

tag aag ggc acg tgg tga ttc ccc gga tcc acc cca act cca tct gtg 96
 * Lys Gly Thr Trp * Phe Pro Gly Ser Thr Pro Thr Pro Ser Val
 20 25

cag caa aca aca cag ggg tgt aca tcc tca cat cca aca cct ccc agt 144
 Gln Gln Thr Thr Gln Gly Cys Thr Ser Ser His Pro Thr Pro Pro Ser
 30 35 40 45

atg aca cat att gct tca atg ctt cag ctc cac ctg aag aag att gta 192
 Met Thr His Ile Ala Ser Met Leu Gln Leu His Leu Lys Lys Ile Val
 50 55 60

cat cag tca cag acc tgc cca atg cct ttg atg gac caa tta cca taa 240
 His Gln Ser Gln Thr Cys Pro Met Pro Leu Met Asp Gln Leu Pro *
 65 70 75

cta ttg tta acc gtg atg gca ccc gct atg tcc aga aag gag aat aca 288
 Leu Leu Leu Thr Val Met Ala Pro Ala Met Ser Arg Lys Glu Asn Thr
 80 85 90

gaa cga atc ctg aag aca tct acc cca gca acc cta ctg atg atg acg 336
 Glu Arg Ile Leu Lys Thr Ser Thr Pro Ala Thr Leu Leu Met Met Thr
 95 100 105

tga gca gcg gct cct cca gtg aaa gga gca gca ctt cag gag gtt aca	384
* Ala Ala Ala Pro Pro Val Lys Gly Ala Ala Leu Gln Glu Val Thr	
110 115 120	
tct ttt aca cct ttt cta ctg tac acc cca tcc cag acg aag aca gtc	432
Ser Phe Thr Pro Phe Leu Leu Tyr Thr Pro Ser Gln Thr Lys Thr Val	
125 130 135	
cct gga tca ccg aca gca cag aca gaa tcc ctg cta cca tcc agg caa	480
Pro Gly Ser Pro Thr Ala Gln Thr Glu Ser Leu Leu Pro Ser Arg Gln	
140 145 150 155	
ctc cta gta gta caa cgg aag aaa cag cta ccc aga agg aac agt ggt	528
Leu Leu Val Val Gln Arg Lys Lys Gln Leu Pro Arg Arg Asn Ser Gly	
160 165 170	
ttg gca aca gat ggc atg agg gat atc gcc aaa cac cca gag aag act	576
Leu Ala Thr Asp Gly Met Arg Asp Ile Ala Lys His Pro Glu Lys Thr	
175 180 185	
ccc att cga caa cag gga cag ctg cag cct cag ctc ata cca gcc atc	624
Pro Ile Arg Gln Gln Gly Gln Leu Gln Pro Gln Leu Ile Pro Ala Ile	
190 195 200	
caa tgc aag gaa gga caa cac caa gcc cag agg aca gtt cct gga ctg	672
Gln Cys Lys Glu Gly Gln His Gln Ala Gln Arg Thr Val Pro Gly Leu	
205 210 215	
att tct tca acc caa tct cac acc cca tgg gac gag gtc atc aag cag	720
Ile Ser Ser Thr Gln Ser His Thr Pro Trp Asp Glu Val Ile Lys Gln	
220 225 230 235	
gaa gaa gga tgg ata tgg act cca gtc ata gta caa cgc ttc agc cta	768
Glu Glu Gly Trp Ile Trp Thr Pro Val Ile Val Gln Arg Phe Ser Leu	
240 245 250	
ctg caa atc caa aca cag gtt tgg tgg aag att tgg aca gga cag gac	816
Leu Gln Ile Gln Thr Gln Val Trp Trp Lys Ile Trp Thr Gly Gln Asp	
255 260 265	

```

ctc ttt caa tga caa cgc agc aga gta att ctc aga gct tct cta cat      864
Leu Phe Gln  *  Gln Arg Ser Arg Val Ile Leu Arg Ala Ser Leu His
      270                      275                      280

cac atg aag gct tgg aag aag ata aag acc atc caa caa ctt cta ctc      912
His Met Lys Ala Trp Lys Lys Ile Lys Thr Ile Gln Gln Leu Leu Leu
      285                      290                      295

tga cat caa gca ata gga atg atg tca cag gtg gaa gaa gag acc caa      960
*  His Gln Ala Ile Gly Met Met Ser Gln Val Glu Glu Glu Thr Gln
      300                      305                      310

atc att ctg aag gct caa cta ctt tac tgg aag gtt ata cct ctc att      1008
Ile Ile Leu Lys Ala Gln Leu Leu Tyr Trp Lys Val Ile Pro Leu Ile
      315                      320                      325

acc cac aca cga agg aaa gca gga cct tca tcc cag tga cct cag cta      1056
Thr His Thr Arg Arg Lys Ala Gly Pro Ser Ser Gln  *  Pro Gln Leu
      330                      335                      340

aga ctg ggt cct ttg gag tta ctg cag tta ctg ttg gag att cca act      1104
Arg Leu Gly Pro Leu Glu Leu Leu Gln Leu Leu Leu Glu Ile Pro Thr
      345                      350                      355                      360

cta atg tca atc gtt cct tat cag gag acc aag aca cat tcc acc cca      1152
Leu Met Ser Ile Val Pro Tyr Gln Glu Thr Lys Thr His Ser Thr Pro
      365                      370                      375

gtg ggg ggt ccc ata cca ctc atg gat ctg aat cag atg gac act cac      1200
Val Gly Gly Pro Ile Pro Leu Met Asp Leu Asn Gln Met Asp Thr His
      380                      385                      390

atg gga gtc aag aag gtg gag caa aca caa cct ctg gtc cta taa gga      1248
Met Gly Val Lys Lys Val Glu Gln Thr Gln Pro Leu Val Leu  *  Gly
      395                      400                      405

cac ccc aaa ttc cag aat ggc tga tca tct tgg cat ccc tct tgg cct      1296
His Pro Lys Phe Gln Asn Gly  *  Ser Ser Trp His Pro Ser Trp Pro
      410                      415                      420

tgg ctt tga ttc ttg cag ttt gca ttg cag tca aca gtc gaa gaa ggt      1344

```

Trp Leu * Phe Leu Gln Phe Ala Leu Gln Ser Thr Val Glu Glu Gly
 425 430 435

 gtg ggc aga aga aaa agc tag tga tca aca gtg gca atg gag ctg tgg 1392
 Val Gly Arg Arg Lys Ser * * Ser Thr Val Ala Met Glu Leu Trp
 440 445 450

 agg aca gaa agc caa gtg gac tca acg gag agg cca gca agt ctc agg 1440
 Arg Thr Glu Ser Gln Val Asp Ser Thr Glu Arg Pro Ala Ser Leu Arg
 455 460 465

 aaa tgg tgc att tgg tga aca agg agt cgt cag aaa ctc cag acc agt 1488
 Lys Trp Cys Ile Trp * Thr Arg Ser Arg Gln Lys Leu Gln Thr Ser
 470 475 480

 tta tga cag ctg atg aga caa gga acc tgc aga atg tgg aca tga aga 1536
 Leu * Gln Leu Met Arg Gln Gly Thr Cys Arg Met Trp Thr * Arg
 485 490 495

 ttg ggg tgt aa cacctacacc attatcttgg aaagaaacaa ccggttgaaa 1587
 Leu Gly Cys

 cataaccatt acagggagct gggacactta acagatgcaa tgtgctactg attgtttcat 1647
 tgcgaatctt ttttagcata aaattttcta ctcttatgga gaaagctctg agcatcggat 1707
 ttgagacctg caggtatggg ttcatagaag ggcacgtggt gattccccgg atccacccca 1767
 actccatctg tgcagcaaac aacacagggg tgtacatcct cacatccaac acctcccagt 1827
 atgacacata ttgcttcaat gcttcagctc cacctgaaga agattgtaca tcagtcacag 1887
 acctgcccac tgcctttgat ggaccaatta ccataactat tgttaaccgt gatggcacc 1947
 gctatgtcca gaaaggagaa tacagaacga atcctgaaga catctacccc agcaacccta 2007
 ctgatgatga cgtgagcagc ggctcctcca gtgaaaggag cagcacttca ggagggttaca 2067
 tctttttacac cttttctact gtacacccca tccagacga agacagtccc tggatcaccg 2127
 acagcacaga cagaatccct gctaccatcc aggcaactcc tagtagtaca acggaagaaa 2187
 cagctaccca gaaggaacag tggtttggca acagatggca tgagggatat cgccaaacac 2247
 ccagagaaga ctcccattcg acaacagggg cagctgcagc ctcagctcat accagccatc 2307
 caatgcaagg aaggacaaca ccaagcccag aggacagttc ctggactgat ttcttcaacc 2367
 caatctcaca ccccatggga cgaggtcatc aagcaggaag aaggatggat atggactcca 2427
 gtcatagtac aacgcttcag cctactgcaa atccaaacac aggtttggtg gaagatttgg 2487
 acaggacagg acctctttca atgacaacgc agcagagtaa ttctcagagc ttctctacat 2547
 cacatgaagg cttggaagaa gataaagacc atccaacaac ttctactctg acatcaagca 2607
 ataggaatga tgtcacaggt ggaagaagag acccaaatca ttctgaaggc tcaactactt 2667

tactggaagg ttatacctct cattacccac acacgaagga aagcaggacc ttcatcccag 2727
 tgacctcagc taagactggg tcctttggag ttactgcagt tactgttgga gattccaact 2787
 ctaatgtcaa tcgttcctta tcaggagacc aagacacatt ccaccccagt ggggggtccc 2847
 ataccactca tggatctgaa tcagatggac actcacatgg gagtcaagaa ggtggagcaa 2907
 acacaacctc tggtcctata aggacacccc aaattccaga atggctgatc atcttggcat 2967
 ccctctttggc cttggcctttg attcttgcag tttgcattgc agtcaacagt cgaagaaggt 3027
 gtgggcagaa gaaaaagcta gtgatcaaca gtggcaatgg agctgtggag gacagaaagc 3087
 caagtggact caacggagag gccagcaagt ctcaggaaat ggtgcatttg gtgaacaagg 3147
 agtcgtcaga aactccagac cagtttatga cagctgatga gacaaggaac ctgcagaatg 3207
 tggacatgaa gattggggtg taacacctac accattatct tggaaagaaa caaccgttgg 3267
 aaacataacc attacagga gctgggacac ttaacagatg caatgtgcta ctgattgttt 3327
 cattgcgaat ctttttttagc ataaaatttt ctactctta 3366

<210> 12

<211> 2878

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (48)...(2783)

<223> BEHAB

<400> 12

tgtggcactg cctgcgtacc caaccccagc cctgggtagc ctgcagc atg gcc cag 56

Met Ala Gln

1

ctg ttc ctg ccc ctg ctg gca gcc ctg gtc ctg gcc cag gct cct gca 104

Leu Phe Leu Pro Leu Leu Ala Ala Leu Val Leu Ala Gln Ala Pro Ala

5

10

15

gct tta gca gat gtt ctg gaa gga gac agc tca gag gac cgc gct ttt 152

Ala Leu Ala Asp Val Leu Glu Gly Asp Ser Ser Glu Asp Arg Ala Phe

20

25

30

35

cgc gtg cgc atc gcg ggc gac gcg cca ctg cag ggc gtg ctc ggc ggc 200

Arg Val Arg Ile Ala Gly Asp Ala Pro Leu Gln Gly Val Leu Gly Gly

40

45

50

gcc ctc acc atc cct tgc cac gtc cac tac ctg cgg cca ccg ccg agc 248

Ala	Leu	Thr	Ile	Pro	Cys	His	Val	His	Tyr	Leu	Arg	Pro	Pro	Pro	Ser	
			55					60						65		
cgc	cgg	gct	gtg	ctg	ggc	tct	ccg	cgg	gtc	aag	tgg	act	ttc	ctg	tcc	296
Arg	Arg	Ala	Val	Leu	Gly	Ser	Pro	Arg	Val	Lys	Trp	Thr	Phe	Leu	Ser	
			70					75						80		
cgg	ggc	cgg	gag	gca	gag	gtg	ctg	gtg	gcg	cgg	gga	gtg	cgc	gtc	aag	344
Arg	Gly	Arg	Glu	Ala	Glu	Val	Leu	Val	Ala	Arg	Gly	Val	Arg	Val	Lys	
			85					90						95		
gtg	aac	gag	gcc	tac	cgg	ttc	cgc	gtg	gca	ctg	cct	gcg	tac	cca	gcg	392
Val	Asn	Glu	Ala	Tyr	Arg	Phe	Arg	Val	Ala	Leu	Pro	Ala	Tyr	Pro	Ala	
100						105				110					115	
tcg	ctc	acc	gac	gtc	tcc	ctg	gcg	ctg	agc	gag	ctg	cgc	ccc	aac	gac	440
Ser	Leu	Thr	Asp	Val	Ser	Leu	Ala	Leu	Ser	Glu	Leu	Arg	Pro	Asn	Asp	
					120				125					130		
tca	ggt	atc	tat	cgc	tgt	gag	gtc	cag	cac	ggc	atc	gat	gac	agc	agc	488
Ser	Gly	Ile	Tyr	Arg	Cys	Glu	Val	Gln	His	Gly	Ile	Asp	Asp	Ser	Ser	
			135					140						145		
gac	gct	gtg	gag	gtc	aag	gtc	aaa	ggg	gtc	gtc	ttt	ctc	tac	cga	gag	536
Asp	Ala	Val	Glu	Val	Lys	Val	Lys	Gly	Val	Val	Phe	Leu	Tyr	Arg	Glu	
			150					155						160		
ggc	tct	gcc	cgc	tat	gct	ttc	tcc	ttt	tct	ggg	gcc	cag	gag	gcc	tgt	584
Gly	Ser	Ala	Arg	Tyr	Ala	Phe	Ser	Phe	Ser	Gly	Ala	Gln	Glu	Ala	Cys	
			165				170					175				
gcc	cgc	att	gga	gcc	cac	atc	gcc	acc	ccg	gag	cag	ctc	tat	gcc	gcc	632
Ala	Arg	Ile	Gly	Ala	His	Ile	Ala	Thr	Pro	Glu	Gln	Leu	Tyr	Ala	Ala	
180						185					190				195	
tac	ctt	ggg	ggc	tat	gag	caa	tgt	gat	gct	ggc	tgg	ctg	tcg	gat	cag	680
Tyr	Leu	Gly	Gly	Tyr	Glu	Gln	Cys	Asp	Ala	Gly	Trp	Leu	Ser	Asp	Gln	
					200				205					210		
acc	gtg	agg	tat	ccc	atc	cag	acc	cca	cga	gag	gcc	tgt	tac	gga	gac	728
Thr	Val	Arg	Tyr	Pro	Ile	Gln	Thr	Pro	Arg	Glu	Ala	Cys	Tyr	Gly	Asp	

215	220	225	
atg gat ggc ttc ccc ggg gtc cgg aac tat ggt gtg gtg gac ccg gat			776
Met Asp Gly Phe Pro Gly Val Arg Asn Tyr Gly Val Val Asp Pro Asp			
230	235	240	
gac ctc tat gat gtg tac tgt tat gct gaa gac cta aat gga gaa ttg			824
Asp Leu Tyr Asp Val Tyr Cys Tyr Ala Glu Asp Leu Asn Gly Glu Leu			
245	250	255	
ttc ctg ggt gac cct cca gag aag ctg aca ttg gag gaa gca cgg gcg			872
Phe Leu Gly Asp Pro Pro Glu Lys Leu Thr Leu Glu Glu Ala Arg Ala			
260	265	270	275
tac tgc cag gag cgg ggt gca gag att gcc acc acg ggc caa ctg tat			920
Tyr Cys Gln Glu Arg Gly Ala Glu Ile Ala Thr Thr Gly Gln Leu Tyr			
280	285	290	
gca gcc tgg gat ggt ggc ctg gac cac tgc agc cca ggg tgg cta gct			968
Ala Ala Trp Asp Gly Gly Leu Asp His Cys Ser Pro Gly Trp Leu Ala			
295	300	305	
gat ggc agt gtg cgc tac ccc atc gtc aca ccc agc cag cgc tgt ggt			1016
Asp Gly Ser Val Arg Tyr Pro Ile Val Thr Pro Ser Gln Arg Cys Gly			
310	315	320	
ggg ggc ttg cct ggt gtc aag act ctc ttc ctc ttc ccc aac cag act			1064
Gly Gly Leu Pro Gly Val Lys Thr Leu Phe Leu Phe Pro Asn Gln Thr			
325	330	335	
ggc ttc ccc aat aag cac agc cgc ttc aac gtc tac tgc ttc cga gac			1112
Gly Phe Pro Asn Lys His Ser Arg Phe Asn Val Tyr Cys Phe Arg Asp			
340	345	350	355
tcg gcc cag cct tct gcc atc cct gag gcc tcc aac cca gcc tcc aac			1160
Ser Ala Gln Pro Ser Ala Ile Pro Glu Ala Ser Asn Pro Ala Ser Asn			
360	365	370	
cca gcc tct gat gga cta gag gct atc gtc aca gtg aca gag acc ctg			1208
Pro Ala Ser Asp Gly Leu Glu Ala Ile Val Thr Val Thr Glu Thr Leu			
375	380	385	

gag gaa ctg cag ctg cct cag gaa gcc aca gag agt gaa tcc cgt ggg	1256
Glu Glu Leu Gln Leu Pro Gln Glu Ala Thr Glu Ser Glu Ser Arg Gly	
390 395 400	
gcc atc tac tcc atc ccc atc atg gag gac gga gga ggt gga agc tcc	1304
Ala Ile Tyr Ser Ile Pro Ile Met Glu Asp Gly Gly Gly Gly Ser Ser	
405 410 415	
act cca gaa gac cca gca gag gcc cct agg acg ctc cta gaa ttt gaa	1352
Thr Pro Glu Asp Pro Ala Glu Ala Pro Arg Thr Leu Leu Glu Phe Glu	
420 425 430 435	
aca caa tcc atg gta ccg ccc acg ggg ttc tca gaa gag gaa ggt aag	1400
Thr Gln Ser Met Val Pro Pro Thr Gly Phe Ser Glu Glu Glu Gly Lys	
440 445 450	
gca ttg gag gaa gaa gag aaa tat gaa gat gaa gaa gag aaa gag gag	1448
Ala Leu Glu Glu Glu Glu Lys Tyr Glu Asp Glu Glu Glu Lys Glu Glu	
455 460 465	
gaa gaa gaa gag gag gag gtg gag gat gag gct ctg tgg gca tgg ccc	1496
Glu Glu Glu Glu Glu Glu Val Glu Asp Glu Ala Leu Trp Ala Trp Pro	
470 475 480	
agc gag ctc agc agc ccg ggc cct gag gcc tct ctc ccc act gag cca	1544
Ser Glu Leu Ser Ser Pro Gly Pro Glu Ala Ser Leu Pro Thr Glu Pro	
485 490 495	
gca gcc cag gag gag tca ctc tcc cag gcg cca gca agg gca gtc ctg	1592
Ala Ala Gln Glu Glu Ser Leu Ser Gln Ala Pro Ala Arg Ala Val Leu	
500 505 510 515	
cag cct ggt gca tca cca ctt cct gat gga gag tca gaa gct tcc agg	1640
Gln Pro Gly Ala Ser Pro Leu Pro Asp Gly Glu Ser Glu Ala Ser Arg	
520 525 530	
cct cca agg gtc cat gga cca cct act gag act ctg ccc act ccc agg	1688
Pro Pro Arg Val His Gly Pro Pro Thr Glu Thr Leu Pro Thr Pro Arg	
535 540 545	

gag agg aac cta gca tcc cca tca cct tcc act ctg gtt gag gca aga	1736
Glu Arg Asn Leu Ala Ser Pro Ser Pro Ser Thr Leu Val Glu Ala Arg	
550 555 560	
gag gtg ggg gag gca act ggt ggt cct gag cta tct ggg gtc cct cga	1784
Glu Val Gly Glu Ala Thr Gly Gly Pro Glu Leu Ser Gly Val Pro Arg	
565 570 575	
gga gag agc gag gag aca gga agc tcc gag ggt gcc cct tcc ctg ctt	1832
Gly Glu Ser Glu Glu Thr Gly Ser Ser Glu Gly Ala Pro Ser Leu Leu	
580 585 590 595	
cca gcc aca cgg gcc cct gag ggt acc agg gag ctg gag gcc ccc tct	1880
Pro Ala Thr Arg Ala Pro Glu Gly Thr Arg Glu Leu Glu Ala Pro Ser	
600 605 610	
gaa gat aat tct gga aga act gcc cca gca ggg acc tca gtg cag gcc	1928
Glu Asp Asn Ser Gly Arg Thr Ala Pro Ala Gly Thr Ser Val Gln Ala	
615 620 625	
cag cca gtg ctg ccc act gac agc gcc agc cga ggt gga gtg gcc gtg	1976
Gln Pro Val Leu Pro Thr Asp Ser Ala Ser Arg Gly Gly Val Ala Val	
630 635 640	
gtc ccc gca tca ggt gac tgt gtc ccc agc ccc tgc cac aat ggt ggg	2024
Val Pro Ala Ser Gly Asp Cys Val Pro Ser Pro Cys His Asn Gly Gly	
645 650 655	
aca tgc ttg gag gag gag gaa ggg gtc cgc tgc cta tgt ctg cct ggc	2072
Thr Cys Leu Glu Glu Glu Glu Gly Val Arg Cys Leu Cys Leu Pro Gly	
660 665 670 675	
tat ggg ggg gac ctg tgc gat gtt ggc ctc cgc ttc tgc aac ccc ggc	2120
Tyr Gly Gly Asp Leu Cys Asp Val Gly Leu Arg Phe Cys Asn Pro Gly	
680 685 690	
tgg gac gcc ttc cag ggc gcc tgc tac aag cac ttt tcc aca cga agg	2168
Trp Asp Ala Phe Gln Gly Ala Cys Tyr Lys His Phe Ser Thr Arg Arg	
695 700 705	
agc tgg gag gag gca gag acc cag tgc cgg atg tac ggc gcg cat ctg	2216

Ser Trp Glu Glu Ala Glu Thr Gln Cys Arg Met Tyr Gly Ala His Leu
 710 715 720
 gcc agc atc agc aca ccc gag gaa cag gac ttc atc aac aac cgg tac 2264
 Ala Ser Ile Ser Thr Pro Glu Glu Gln Asp Phe Ile Asn Asn Arg Tyr
 725 730 735
 cgg gag tac cag tgg atc gga ctc aac gac agg acc atc gaa ggc gac 2312
 Arg Glu Tyr Gln Trp Ile Gly Leu Asn Asp Arg Thr Ile Glu Gly Asp
 740 745 750 755
 ttc ttg tgg tcg gat ggc gtc ccc ctg ctc tat gag aac tgg aac cct 2360
 Phe Leu Trp Ser Asp Gly Val Pro Leu Leu Tyr Glu Asn Trp Asn Pro
 760 765 770
 ggg cag cct gac agc tac ttc ctg tct gga gag aac tgc gtg gtc atg 2408
 Gly Gln Pro Asp Ser Tyr Phe Leu Ser Gly Glu Asn Cys Val Val Met
 775 780 785
 gtg tgg cat gat cag gga caa tgg agt gac gtg ccc tgc aac tac cac 2456
 Val Trp His Asp Gln Gly Gln Trp Ser Asp Val Pro Cys Asn Tyr His
 790 795 800
 ctg tcc tac acc tgc aag atg ggg ctg gtg tcc tgt ggg ccg cca ccg 2504
 Leu Ser Tyr Thr Cys Lys Met Gly Leu Val Ser Cys Gly Pro Pro Pro
 805 810 815
 gag ctg ccc ctg gct caa gtg ttc ggc cgc cca cgg ctg cgc tat gag 2552
 Glu Leu Pro Leu Ala Gln Val Phe Gly Arg Pro Arg Leu Arg Tyr Glu
 820 825 830 835
 gtg gac act gtg ctt cgc tac cgg tgc cgg gaa gga ctg gcc cag cgc 2600
 Val Asp Thr Val Leu Arg Tyr Arg Cys Arg Glu Gly Leu Ala Gln Arg
 840 845 850
 aat ctg ccg ctg atc cga tgc caa gag aac ggt cgt tgg gag gcc ccc 2648
 Asn Leu Pro Leu Ile Arg Cys Gln Glu Asn Gly Arg Trp Glu Ala Pro
 855 860 865
 cag atc tcc tgt gtg ccc aga aga cct gcc cga gct ctg cac cca gag 2696
 Gln Ile Ser Cys Val Pro Arg Arg Pro Ala Arg Ala Leu His Pro Glu

870 875 880

gag gac cca gaa gga cgt cag ggg agg cta ctg gga cgc tgg aag gcg 2744
 Glu Asp Pro Glu Gly Arg Gln Gly Arg Leu Leu Gly Arg Trp Lys Ala
 885 890 895

ctg ttg atc ccc cct tcc agc ccc atg cca ggt ccc tag ggggcaaggc 2793
 Leu Leu Ile Pro Pro Ser Ser Pro Met Pro Gly Pro *
 900 905 910

cttgaacact gccggccaca gcactgcct gtcacccaaa ttttccctca caccctgcgc 2853
 tcaccacagg aagtgacaac atgac 2878

<210> 13
 <211> 1726
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (137)...(898)

<400> 13

gcgcgacggc gcaggggctg cggggagcgc cgcgcaggcc gtgcagttcc tagcgaggag 60
 gcgccgccgc cattgccgct ctctcgggtga gcgcagcccc gctctccggg ccgggccttc 120
 gcggggccacc ggcgcc atg ggc cag tgc ggc atc acc tcc tcc aag acc gtg 172
 Met Gly Gln Cys Gly Ile Thr Ser Ser Lys Thr Val
 1 5 10

ctg gtc ttt ctc aac ctc atc ttc tgg ggg gca gct ggc att tta tgc 220
 Leu Val Phe Leu Asn Leu Ile Phe Trp Gly Ala Ala Gly Ile Leu Cys
 15 20 25

tat gtg gga gcc tat gtc ttc atc act tat gat gac tat gac cac ttc 268
 Tyr Val Gly Ala Tyr Val Phe Ile Thr Tyr Asp Asp Tyr Asp His Phe
 30 35 40

ttt gaa gat gtg tac acg ctc atc cct gct gta gtg atc ata gct gta 316
 Phe Glu Asp Val Tyr Thr Leu Ile Pro Ala Val Val Ile Ile Ala Val
 45 50 55 60

gga gcc ctg ctt ttc atc att ggg cta att ggc tgc tgt gcc aca atc 364
 Gly Ala Leu Leu Phe Ile Ile Gly Leu Ile Gly Cys Cys Ala Thr Ile
 65 70 75

cgg gaa agt cgc tgt gga ctt gcc acg ttt gtc atc atc ctg ctc ttg 412
 Arg Glu Ser Arg Cys Gly Leu Ala Thr Phe Val Ile Ile Leu Leu Leu
 80 85 90

gtt ttt gtc aca gaa gtt gtt gta gtg gtt ttg gga tat gtt tac aga 460
 Val Phe Val Thr Glu Val Val Val Val Val Leu Gly Tyr Val Tyr Arg
 95 100 105

gca aag gtg gaa aat gag gtt gat cgc agc att cag aaa gtg tat aag 508
 Ala Lys Val Glu Asn Glu Val Asp Arg Ser Ile Gln Lys Val Tyr Lys
 110 115 120

acc tac aat gga acc aac cct gat gct gct agc cgg gct att gat tat 556
 Thr Tyr Asn Gly Thr Asn Pro Asp Ala Ala Ser Arg Ala Ile Asp Tyr
 125 130 135 140

gta cag aga cag ctg cat tgt tgt gga att cac aac tac tca gac tgg 604
 Val Gln Arg Gln Leu His Cys Cys Gly Ile His Asn Tyr Ser Asp Trp
 145 150 155

gaa aat aca gat tgg ttc aaa gaa acc aaa aac cag agt gtc cct ctt 652
 Glu Asn Thr Asp Trp Phe Lys Glu Thr Lys Asn Gln Ser Val Pro Leu
 160 165 170

agc tgc tgc aga gag act gcc agc aat tgt aat ggc agc ctg gcc cac 700
 Ser Cys Cys Arg Glu Thr Ala Ser Asn Cys Asn Gly Ser Leu Ala His
 175 180 185

cct tcc gac ctc tat gct gag ggg tgt gag gct cta gta gtg aag aag 748
 Pro Ser Asp Leu Tyr Ala Glu Gly Cys Glu Ala Leu Val Val Lys Lys
 190 195 200

cta caa gaa atc atg atg cat gtg atc tgg gcc gca ctg gca ttt gca 796
 Leu Gln Glu Ile Met Met His Val Ile Trp Ala Ala Leu Ala Phe Ala
 205 210 215 220

gct att cag ctg ctg ggc atg ctg tgt gct tgc atc gtg ttg tgc aga 844

Ala Ile Gln Leu Leu Gly Met Leu Cys Ala Cys Ile Val Leu Cys Arg
 225 230 235

agg agt aga gat cct gct tac gag ctc ctc atc act ggc gga acc tat 892
 Arg Ser Arg Asp Pro Ala Tyr Glu Leu Leu Ile Thr Gly Gly Thr Tyr
 240 245 250

gca tag ttgacaactc aagcctgagc tttttggtct tgttctgatt tggaaggtga 948
 Ala *

attgagcagg tctgctgctg ttggcctctg gagttcattt agttaaagca catgtacact 1008
 ggtgttgagc agagcagctt ggcttttcat gtgccacact acttacctac tacctgcgac 1068
 tttctttttc cttgttctag ctgactcttc atgcccctaa gattttaagt acgatgggtga 1128
 acgttctaata ttcagaacca attgagagtc atgtagtgtg gtagaattaa aggaggacac 1188
 gagcctgctt ctgttacctc caagtggtaa caggactgat gccgaaatgt caccaggtcc 1248
 tttcagtctt cacagtggag aactcttggc caaagggttt tgcggggagg aggaggaaac 1308
 cagctttctg gttaagggtta acaccagatg gtgccctca ttggtgtcct tttaaaaaat 1368
 atttactgta gtccaataag atagcagctg taaaaaatga ctaaaataga ttgtaggatc 1428
 atatggcgta tatcttgggt catcttcaaa atcagagact gagctttgaa actagtgggt 1488
 tttaatcaaa gttggcttta taggaggagt ataatgtatg cactactgtt ttaaaagaat 1548
 tagtgtgagt gtgtttttgt atgaatgagc ccattcatgg taagtcttaa gcttgttgga 1608
 aataatgtac ccatgtagac tagcaaaata gtatgtagat gtgatctcag ttgtaaatag 1668
 aaaaatctaa ttcaataaac tctgtatcag cccccaaaaa aaaaaaaaaa aaaaaaaaaa 1726

<210> 14

<211> 3395

<212> DNA

<213> Homo sapiens

<400> 14

gccttcatgc ccttgtactg gcttctgggg ccgcacttcc attcttcttc cccacaagcc 60
 gattattgtt ttgaaagcat gaaacaaaac acaatcactc ccctgggaca gattccttta 120
 cgatctgtca gtgcagagaa attgtgcttc cacggtgagg ggagtggaca agttgggagg 180
 ggcaggcgct catcctcctg tacctgtctt gcgaaagtgg ttggtggtaa cctgatttcc 240
 catctctctc tgacagttgc tgggatacaa acgaccacag tgtgccctgg tgggtcatac 300
 gaataccgat tttaatattcc atcatcgtaa agtccctgtt gacatgtgga ttatctgtgg 360
 aatcacacac ttgggcagac agcaagcctg ccaccttcta ttaaataagaa ctgaccccat 420
 tccacgcac actcatcctt cccggaaaag gacctcttca ggcatgtgga cccccagagg 480
 cagttcccag ggaaggaggg agggactgag ggagcggagg ggagggtcga aatcaggtgg 540
 acttcagggc aaattagggg tccacgtgtc tgtttcacia acttattatc tcttgtgcag 600

```

gtcaattttg tccttttcat tagtattata cgaattttgc tgcagaagtt aacatcccca 660
gatgtcggcg gcaacgacca gtctcagtac aagtgagtgt gtgtcccggc ctggccgtcc 720
ctgatgggga gtgccgtgtg tgagacagga gacccccacg ctgccaggaa ggccctggag 780
gtgcggcgac tccgccgcct ccgccactgt cacacctcga ggctgggccc cagggtcttg 840
ctcctgcccc aggtcagag agggccctgga cagcatttcc ctggcgaagt aacaaatcac 900
tccaaaattht ggtggtttta aacacccgcc atctagatag ggtgtgattc tgcacctgat 960
gacctgggag gactcagcag gggcacttct ggccttggca actgtggaag tcaggcacca 1020
tttcttttat ttcaaggcca gaagaaagtc ctctccagt g aacttgcttg tttaaagcatc 1080
cgccactta catgttatac agcctgtctt tttctccttg accacacagt attttttatt 1140
ttacatttaa ttctgtctcc ccctcagaca tggccatgtg tccatggccg tgtgtctact 1200
gccagcccta ataccaggcc ttgttactta ataaaaattc agaaagcgtt tcctgcccac 1260
gtgattacat gaaaatggca ttgtgatctc ccttcagtag atgaagacag ggcagacgtt 1320
cagccagtca cccccagtca tccagcta at tcatgatgat ccggggaatc tataaagtaa 1380
tgcaaagact ccataaattht atatcaagaa tttatthttt tttaaaaagg taaaaaccaa 1440
aattcaagta taaagtaa ac actggcttgg tgtggaagt gttggccag cccagaaggt 1500
gcttcgggtt cagactgacc ggggggtggaa gggcggtgt gcacagagct gtaccatggc 1560
aggaaccctg ccgccagcc taggaggggt cctggcctgc atttcctttt tttccacct 1620
tagaaaacag ggatggattt ggggaggggc ggttagacag gagcatggcc aacagcccca 1680
aagctgggtg cagataaccg ccagtgccct gagggaagtc actttgccac tccatgcctc 1740
agcttcccca tctgtaacat ggaagtaata atgtcactaa ggtcaggagg atcaatgcag 1800
tctatattht ccggatattht gctgcctggc accagagtat tgcaaaagag tctggtaaac 1860
cagagaatct aattctgtgtg tgtagtgtct aaatccagg agtatccccg tctctgcttg 1920
cccacacgag ggatgggatt ccatgtcaga ccccgcatgg cagcatgacc ttgtccctgc 1980
tcaggggagt ccgggtggtg ctgtgtggca gagacccccg aataagatcg ttcttggttg 2040
cctttcaggt tgaacttgga tttcctgtga atcttgtctt agtttctatg ctgacctttg 2100
cattggttct ttgggagtggt ttgttaaatt ggacttggag atggatgctt ttatttgga 2160
acaaaatctg ttccctttcc aggacaacag aaaaggccga gacacccttg aggccacacg 2220
aggtccttht tgtgtaactc tggtcacaga agtagcaaaa cctccctgga taggggcaa 2280
gagctcagcc gcagataggc ctcttttcca gcaaaactca gatgcgttaa tagagggaat 2340
caggcaggtc ctctgccgct gactgaggga cctggctccc acggagccca gctgttacga 2400
gctgcaggtc cgggagcagg atcggttctc atcatthtaa cctgcgtgtg gcagcggtg 2460
aggccagcct tcttgtccca tgctgcctc aggatgggc gggctcccag gcggggcctg 2520
tgtaggttht gggtcactca cagggaatct tgcccctgcc agccatgaac ccaggacac 2580
agcgtgtgtc cttgtgaggg aatttgggag aaacggacac ttccggaatg cacgttccca 2640
ggccctgagt gctagatatg cacaggggcc tcagacataa acgtaaagaa acagtggtht 2700
ttcttgttg gcggggactc cagcagtggg gaggaggtgt ggagccctc ccacaccag 2760
ctgccccctc ctgagaagga cccagccagg cttgcagagg gtccgagggt ccagcagctg 2820
agctggctca tctccttctg tgcaggaggc tggccaagtc cacgctcctg cttatccgc 2880
tgttcggcgt cactacatg gtgtttgccc gtthtcccat cagcatctcc tccaaatacc 2940
agatactgtt tgagctgtgc ctcggtcgt tccaggtagg tgtgcggagg cggccgcct 3000
ccccgagcg cgggtcctgc ctgacctct cttctcattg cagggcctgg tgggtggcgt 3060

```

```

cctctactgt ttcctgaaca gtgaggtaag ttcattggccg ccttggaacc aagcacaggt 3120
actcacttgc ttcctgagat gctgcccagc ctggtgccgc agccccaca catgcctgtc 3180
ttccgcgggc tcgagttact gcccgggacc ccacagttct gtgtcccat cagaaaaccc 3240
acaacgacac aggcaaacc actcttcttg gccagcttc caaacccgc cgtccttccg 3300
tccaccctaa gtaacatcga ctgccgtccg tctcggaacc ctctcgggaa actcccggcc 3360
gggtctaaaa agaattctgt tttcccatg gtcggg                                     3395

```

<210> 15

<211> 3143

<212> DNA

<213> Homo sapiens

<400> 15

```

ggggaagtgt gggagcaggt gggctgggca gtggcagaaa cctgatgaca caatctcgcc 60
gcctccctgt gttggtggag gatgtctgca gcagcattta aattctggga gggcttggtt 120
gtcagcagca gcaggaggag gcagagacag catcgctggg accagactcg tctcaggcca 180
gttgcagcct tctcagccaa acgccgacca aggtacagct tcagtttgct actgggttgt 240
gcattcagct gaatttcatg gggaagtcca aattctaagg aaaaaatgt ggtagtataa 300
aaaggtatca ctgttgtaac ctatgaagat gtcagctatt cttttgaaat attttgcagg 360
aaaactcact accatgagaa ttgcagtgat ttgcttttgc ctctaggca tcacctgtgc 420
cataccagtg agtacagttg catcttaaag aaaattctctg aaaataactg aattgtgtgc 480
ttccatgtgc taggaggaca ttcttgtaat ctttcttcat cttttctgtt tctaaggtta 540
aacaggctga ttctggaagt tctgaggaaa agcaggtaag catcttttat gtttttata 600
agttaaataca ttactcaat tatggcgaga ggtgcaagaa acgtatttgc tgcgatcaaa 660
tgagttcata tttgtaaagc aatttgaaag agtgcctagc ccacagtaag tgctacataa 720
gagtttggtt aatgaatctg caaaaaaaaa aaaaattaca aaaaggtagc taagggtccg 780
gggtgactata tgcttccatc aagactagtg aagaatgggt gttttttcca ttcattcccta 840
catttctttt tttaataatg ataaacatgc aacttttttg tagctttaca acaaataccc 900
agatgctgtg gccacatggc taaaccctga cccatctcag aagcagaatc tcctagcccc 960
acaggatattt ttaacttctc cataattaaa ctacagtgat gaaagatagc cacactcagg 1020
ccatttgggc tgctcagatg aatcctgccc tgctgctgg caaacatgtg cttaggacat 1080
tgactgatct gccatgttgg cttctctctg tggttaagcca tccacagatg aggctgaaaa 1140
ataaaaaactg ctttggtatta aaaagggttaa cttttgaata aaaaagctag gcatgtgtga 1200
tgcgactata cacgtgccat tccttcttca gaatgctgtg tcctctgaag aaaccaatga 1260
ctttaaacaa gaggtagtt ctcatcttca atcagaggcc catcatgcct tgaagagatg 1320
aaagaaggca ttgcctggat tctcttctga tgaaatttca ttagcaagtt ttccagctaa 1380
ttggcagctc aaaacttgct cataaataaa acatgtattt actaaatatc agaaatacta 1440
ggtttcctcg gataacctaa aagccatggg atgtactgtg aatgcaaaga ttctgaaact 1500
aaataaaaaag aaagatagta aaagactaat gtgctataaa ggctaaggga aaataaaaaac 1560
ccatatatta attttcccg ccattctaat tttcagaacc ttccaagtaa gtccaacgaa 1620
agccatgacc acatggatga tatggatgat gaagatgatg atgaccatgt ggacagccag 1680

```

```

gactccattg actcgaacga ctctgatgat gtagatgaca ctgatgattc tcaccagtct 1740
gatgagtctc accattctga tgaatctgat gaactgggtca ctgattttcc cacggacctg 1800
ccagcaaccg aagttttcac tdcagttgtc cccacagtag acacatatga tggccgaggt 1860
gatagtgtgg tttatggact gaggtcaaaa tctaagaagt ttgcgagacc tgacatccag 1920
gtaaatcctt taacagacac acctgatggg tctgactagc gctcaagtct aggaaaccac 1980
agttttgcata ttcatctatt cattcatcca ttcatctatc cattcagcaa gaattcattc 2040
atattctact ttatgaccat tgaatacaaaa tctttttctg cttggcggtt tttgtaagtc 2100
tacataatct ctctctagat ttgattctca aacacaattc tactttttga aatcctggat 2160
caaagtaaca tgctagtatt atttcagcca gatttagaca attttttagta taagatgacc 2220
taaaagctag agagtggaaa aggattacca tattcccatc cctagccgtt catataatta 2280
ttcttcattt gtgccgtgat tcagtaccct gatgctacag acgaggacat cacctcacac 2340
atggaaagcg aggagttgaa tgggtgcatac aaggccatcc ccgttgccca ggacctgaac 2400
gcgccttctg attgggacag ccgtgggaag gacagttatg aaacgagtca gctggatgac 2460
cagagtgtctg aaaccacag ccacaagcag tccagattat ataagcggaa agccaatgat 2520
gagagcaatg agcattccga tgtgattgat agtcaggaac tttccaaagt cagccgtgaa 2580
ttccacagcc atgaatttca cagccatgaa gatatgctgg ttgtagacct caaaagtaag 2640
gaagaagata aacacctgaa atttcgtatt tctcatgaat tagatagtgc atcttctgag 2700
gtcaattaaa aggagaaaaa atacaatttc tcactttgca tttagtcaaa agaaaaaatg 2760
ctttatagca aaatgaaaga gaacatgaaa tgcttctttc tcagtttatt ggttgaatgt 2820
gtatctatct gagtctggaa ataactaatg tgtttgataa ttagtttagt ttgtggcttc 2880
atggaaactc cctgtaaaca aaagcttcag ggttatgtct atgttcattc tatagaagaa 2940
atgcaaacta tcactgtatt ttaatatctg ttattctctc atgaatagaa atttatgtag 3000
aagcaaaca aatactttta cccacttaaa aagagaatat aacattttat gtcactataa 3060
tcttttgttt tttaagttag tgtatatctt gttgtgatta tcttttgtgg tgtgaataaa 3120
tcttttatct tgaatgtaat aag 3143

```

<210> 16

<211> 2650

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1542)...(2048)

<223> Pleiotrophin/OSF-1

<400> 16

```

gaattcggca cgagctgcag ggtcaggagg agaatcgtgg ggccaggagg gcagaggcac 60
actccatctt cgtgctcctc acaggccctg ctcctctgcc tgctaaggac acaggggaagg 120
gggtccccac ctcaagtgcct gcctcccttc cctgtgcctg tgtacctggc agtcacagcc 180
acctggcgtg tcccagaaac caaccggctg acctcatctc ctgcccggcc ccacctccat 240

```

tggcctttggc ttttggcggt tgtgctgccc gaccctttct cctgtccgga tgcgcagggc 300
 agggcctgag ccgtcgagct gcacccacag caggctgcct ttggtgactc accgggtgaa 360
 cggggggcatt gcgaggcatc cctccctgg gtttggctcc tgcccacggg gctgacagta 420
 gaaatcacag gctgtgagac agctggagcc cagctctgct tgaacctatt ttaggtctct 480
 gatccccgct tcctctttag actcccctag agctcagcca gtgctcaacc tgaggctggg 540
 ggtctctgag gaagagttag ttggagctga ggggtctggg gctgtcccct gagagagggg 600
 ccagaggcag tgtcaagagc cgggcagtct gattgtggct caccctccat cactcccagg 660
 gcccctggcc cagcagccgc agtcccaac cacaatatcc tttgggggtt ggcctacgga 720
 gctggggcgg atgacccca aatagccctg gcagattccc cctagaccg cccgcaccat 780
 ggtcaggcat gcccctcctc atcgctggca cagcccagag ggtataaaca gtgctggagg 840
 ctggcggggc aggccagctg agtcctgagc agcagcccag cggatcctga gaacttcagg 900
 gtgagtttgg ggacccttga ttgttctttc tttttcgcta ttgtaaaatt catgttatat 960
 ggagggggca aagttttcag ggtgttgttt agaattggaa gatgtccctt gtatcaccat 1020
 ggaccctcat gataattttg tttctttcac tttctactct gttgacaacc attgtctcct 1080
 cttattttct tttcattttc tgtaactttt tcgttaaact ttagcttgca tttgtaacga 1140
 atttttaaat tcacttttgt ttatttgtca gattgtaagt actttctcta atcacttttt 1200
 tttcaaggca atcagggtat attatattgt acttcagcac agtttttagag aacaattgtt 1260
 ataattaaat gataaggtag aatatttctg catataaatt ctggctggcg tggaaatatt 1320
 cttattggta gaaacaacta catcctggtc atcatcctgc ctttctcttt atggttacaa 1380
 tgatatacac tgtttgagat gaggataaaa tactctgagt ccaaaccggg cccctctgct 1440
 aaccatgttc atgccttctt ctttttcta cagctcctgg gcaacgtgct ggttgttgtg 1500
 ctgtctcatc attttggcaa agaattaatt ccaactcaaa a atg cag gct caa cag 1556

Met Gln Ala Gln Gln

1

5

tac cag cag cag cgt cga aaa ttt gca gct gcc ttc ttg gca ttc att 1604
 Tyr Gln Gln Gln Arg Arg Lys Phe Ala Ala Ala Phe Leu Ala Phe Ile

10

15

20

ttc ata ctg gca gct gtg gat act gct gaa gca ggg aag aaa gag aaa 1652
 Phe Ile Leu Ala Ala Val Asp Thr Ala Glu Ala Gly Lys Lys Glu Lys

25

30

35

cca gaa aaa aaa gtg aag aag tct gac tgt gga gaa tgg cag tgg agt 1700
 Pro Glu Lys Lys Val Lys Lys Ser Asp Cys Gly Glu Trp Gln Trp Ser

40

45

50

gtg tgt gtg ccc acc agt gga gac tgt ggg ctg ggc aca cgg gag ggc 1748
 Val Cys Val Pro Thr Ser Gly Asp Cys Gly Leu Gly Thr Arg Glu Gly

55

60

65

act cgg act gga gct gag tgc aag caa acc atg aag acc cag aga tgt 1796
 Thr Arg Thr Gly Ala Glu Cys Lys Gln Thr Met Lys Thr Gln Arg Cys
 70 75 80 85

 aag atc ccc tgc aac tgg aag aag caa ttt ggc gcg gag tgc aaa tac 1844
 Lys Ile Pro Cys Asn Trp Lys Lys Gln Phe Gly Ala Glu Cys Lys Tyr
 90 95 100

 cag ttc cag gcc tgg gga gaa tgt gac ctg aac aca gcc ctg aag acc 1892
 Gln Phe Gln Ala Trp Gly Glu Cys Asp Leu Asn Thr Ala Leu Lys Thr
 105 110 115

 aga act gga agt ctg aag cga gcc ctg cac aat gcc gaa tgc cag aag 1940
 Arg Thr Gly Ser Leu Lys Arg Ala Leu His Asn Ala Glu Cys Gln Lys
 120 125 130

 act gtc acc atc tcc aag ccc tgt ggc aaa ctg acc aag ccc aaa cct 1988
 Thr Val Thr Ile Ser Lys Pro Cys Gly Lys Leu Thr Lys Pro Lys Pro
 135 140 145

 caa gca gaa tct aag aag aag aaa aag gaa ggc aag aaa cag gag aag 2036
 Gln Ala Glu Ser Lys Lys Lys Lys Lys Glu Gly Lys Lys Gln Glu Lys
 150 155 160 165

 atg ctg gat taa aagatgtcac ctgtggaaca taaaaaggac atcagcaaac 2088
 Met Leu Asp *

aggatcaatt cactcctcag gtgcaggctg cctatcagaa ggtgggtggct ggtgtggcca 2148
 atgccctggc tcacaaatac cactgagatc tttttccctc tgccaaaaat tatggggaca 2208
 tcatgaagcc ccttgagcat ctgacttctg gctaataaag gaaatttatt ttcattgcaa 2268
 tagtgtgttg gaattttttg tgtctctcac tcggaaggac atatgggagg gcaaatcatt 2328
 taaaacatca gaatgagtat ttgggtttaga gtttggcaac atatgccata tgctggctgc 2388
 catgaacaaa ggtgggtata aagaggtcat cagtatatga aacagccccc tgctgtccat 2448
 tccttattcc atagaaaagc cttgacttga ggtagattt tttttatatt ttgttttgtg 2508
 ttattttttt ctttaacatc cctaaaattt tccttacatg ttttactagc cagatttttc 2568
 ctctctctct gactactccc agtcatagct gtccctcttc tcttatggag atccctcgac 2628
 ctgcaggcag gcatgcaagc tt 2650

<210> 17

<211> 22

<212> DNA

<213> Homo sapiens

<400> 17

cagcagttgg atggaagagg ac

22

<210> 18

<211> 22

<212> DNA

<213> Homo sapiens

<400> 18

cactgagatt ctggcactat tc

22

<210> 19

<211> 21

<212> DNA

<213> Homo sapiens

<400> 19

aacaattcca gggctctcact c

21

<210> 20

<211> 21

<212> DNA

<213> Homo sapiens

<400> 20

ttgactggct caggagtata g

21

<210> 21

<211> 21

<212> DNA

<213> Homo sapiens

<400> 21

ctgataatga gggctcccaa c

21

<210> 22

<211> 24

<212> DNA

<213> Homo sapiens

<400> 22

ctctgcactt cctggtaaaa ctct 24

<210> 23

<211> 22

<212> DNA

<213> Homo sapiens

<400> 23

cagcagttgg atggaagagg ac 22

<210> 24

<211> 24

<212> DNA

<213> Homo sapiens

<400> 24

ctctgcactt cctggtaaaa ctct 24

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/08992

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : A61K 51/00; A61M 36/14

US CL : 424/9.1

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/9.1

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Registry, Caplus

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MAEDA, N. et al. A Receptor-like Protein-tyrosine Phosphatase PTP/RPTP beta Binds a Heparin-binding Growth Factor Midkine. The Journal of Biological Chemistry. 1999, Vol. 274, No. 18, pages 12474-12479, see entire document.	1-6, 8-28
A,P	US 6,312,686 B1 (STADDON et al) 06 November 2001 (06.11.2001), see entire document.	1-6, 8-28

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

☐ See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

11 June 2002 (11.06.2002)

Date of mailing of the international search report

28 JUN 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

D. Jones

Telephone No. (703) 308-1235