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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2001 (12.04.2001)

PCT

(10) International Publication Number
WO 01/25268 A1

(51) International Patent Classification[7]: C07K 14/00 // C12N 15/12, C07K 14/705

(21) International Application Number: PCT/DK00/00556

(22) International Filing Date: 4 October 2000 (04.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 1999 01420 4 October 1999 (04.10.1999) DK

(71) Applicants and
(72) Inventors: **SCHROTZ-KING, Petra** [DE/DK]; Læderhatten 8, DK-5220 Odense SØ (DK). **KING, Angus** [GB/DK]; Læderhatten 8, DK-5220 Odense SØ (DK). **MANN, Matthias** [DE/DK]; Grønningen 10, DK-5230 Odense M (DK). **ANDERSEN, Jens** [DK/DK]; Over Holluf Vej 66, DK-5220 Odense SØ (DK). **KUESTER, Bernhard** [DE/DK]; Langegade 43, DK-5000 Odense C (DK).

(74) Agent: **PLOUGMANN, VINGTOFT & PARTNERS A/S**; Sankt Annæ Plads 11, P.O. Box 3007, DK-1021 Copenhagen K (DK).

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

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

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

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

WO 01/25268 A1

(54) Title: HUMAN SEIZURE RELATED PROTEINS

(57) Abstract: The present invention relates to three new isolated and identified genes which code for novel proteins belonging to membrane receptor molecules and a truncated secreted version of said receptors. They show strong homology to a family of proteins that are termed seizure related proteins and they are potentially involved in the control or generation of seizures such as epileptic seizures or other neurological disorders. The invention discloses nucleotide sequences encoding three new polypeptides PSK-1, PSK-2 and PSK-3. The invention further relates to the manufacture of the disclosed nucleotide and polypeptide sequences and their use for the identification of potential drug targets, as well as to antibodies and nucleotide sequences for use in diagnosis and/or prognosis of neurological disorders.

Human seizure related proteins

Field of invention

The present invention relates to three new isolated and identified genes of which two code for novel membrane receptor molecules and one codes for a secreted molecule. The receptor molecules and the secreted molecule show strong homology to a familly of proteins that are termed seizure related proteins and they are potentially involved in the control or generation of seizures such as epileptic seizures or other neurological disorders. The three disclosed nucleotide sequences encode three polypeptides PSK-1, PSK-2 and PSK-3, which are disclosed in the present application. The invention further relates to the manufacture of the disclosed nucleotide and polypeptide sequences and their use for the identification of potential drug targets, as well as to antibodies and nucleotide sequences for use in diagnosis and/or prognosis of neurological disorders.

General background

Epilepsy affects at least 1 to 2 million people in the United States and 20 to 40 million people worldwide.

Seizures, for example epileptic seizures, are caused by an anomaly in the electric activity of nerve cells. Seizures can be caused by a variety of different reasons, such as:

- More than 50 brain disorders.
- Injury of a portion of the cortex (gray matter) of the brain or injury of the whole brain in the past.
- Acute injury or acute stroke.
- Some brain infections such as abscess, meningitis, encephalitis, parasitic infection (cysticercosis, toxoplasmosis, etc) .
- Degenerative conditions such as Alzheimer's disorder, end-stage multiple sclerosis (MS).
- Tumors often cause seizures.
- Genetic or congenital disorders.
- Vascular lesions including stroke, AVM, venous thrombosis, hypertensive encephalopathy and disorders of blood or the blood vessels (arteritis)

- Chemical disturbances of the brain, (drugs, poisons or organs that do not function properly).
- An inherited tendency to seizures.
- Sleep loss.
- Sudden stopping of drugs (sleep-provoking or tranquilizing drugs) medications or alcohol
- Idiopathic - no cause found.

Epilepsy can be caused by:

- Infections such as abscess, meningitis, encephalitis, parasitic infection (cysticercosis, toxoplasmosis, etc.).
- Brain injury.
- Tumors.
- Vascular lesions including stroke, AVM, venous thrombosis, hypertensive encephalopathy and disorders of blood or the blood vessels (arteritis).
- Degenerative conditions such as Alzheimer's disorder, end-stage MS.
- Genetic or congenital disorders.
- Idiopathic - no cause found.

In the mature brain, "status epilepticus" causes neuronal loss in hippocampal fields CA1 and CA3 and in the dentate hilus. This results in aberrant sprouting of granule cell axons in these areas and in long-term deficits in learning memory and behavior (for review see: Holmes G. L. and Yehezkel, B-A., Seizures in the developing brain: Perhaps not so benign after all, Neuron, Vol. 21, 1231-1234, 1998). The immature brain is more susceptible to seizures but the consequences are less severe than in the adult brain.

Today, most of the drugs used in treating epilepsy were discovered to have anti-epileptic properties by chance. However, a more systematic search for new anti-epileptic drugs is thus necessary to find potentially more effective ways to treat seizures and epilepsy and mechanisms that underlie seizure processes constitute valuable potential target areas for the development of new anti-epileptic drugs (AEDs) as well as drugs for a variety of related neurological disorders.

It is essential to understand the cellular, molecular, and genetic mechanisms underlying seizure processes to develop therapeutic agents designed to suppress seizure-provoking processes and to enhance the natural protective mechanisms of the brain.

Due to technical advances in molecular biology, involving gene mapping and gene identification, it is possible to examine the heritability of various epilepsy syndromes using "reverse genetics" or positional cloning. Abnormal proteins can be identified and characterized through isolation and cloning of the corresponding genes. Methods used to obtain DNA of interest involve digestion of genomic DNA with specific restriction endonucleases or amplification of DNA by polymerase chain reaction technology. Gel electrophoresis is the basis for the separation of differently sized DNA. Genetic epilepsies, however, are extremely variable and since the causes and basic mechanisms of epilepsies have only started to be revealed, there is still no cure for the disease.

The fusion between molecular genetics, neurosciences and clinical epilepsies will provide the base for the research of epilepsies and the primary goal, namely, understanding the mechanisms of human epilepsies.

The knowledge gained from such connections will certainly bear on the diagnosis of the sub-varieties of epilepsies and is already promoting new methods of treatment of the disease. In the most common of human epilepsies, ie, temporal lobe epilepsy, a priority challenge is to analyze paroxysmal depolarization shifts in hippocampal slices in vitro, ie slices excised from known sites of epileptogenicity. Parallel experiments exploring biochemical membrane abnormalities in neuronal and glial membranes isolated from the hippocampal seizure focus would be especially valuable.

A second important goal is the search for polymorphisms of restriction endonuclease patterns in monogenic epilepsies in order to localize the abnormal gene to a specific chromosome. Because of the recent successful applications of positron emission tomography (PET), single-photon-emission-computed tomography and nuclear magnetic resonance computed tomography (NMR-CT), ion transport pathways, neurotransmitter systems and metabolic processes may be constructed within the functioning brains of epileptic patients.

Examples of neurotransmitters are: Glutamate, GABA and glycine, acetylcholine, endorphins, dopamine, serotonin and epinephrine. Glutamate is the major excitatory transmitter in the brain. GABA is the major inhibitory neurotransmitter.

The identification of GABA as a natural inhibitory neurotransmitter led to a search for drugs which might suppress epilepsy through enhancing the inhibiting activity of GABA. Vigabatrin and Tiagabin, for example, are products of this approach. Approaching the opposite way by reducing the activity of the major excitatory neurotransmitter glutamate in the brain, the tendency to develop epileptic seizures might also be minimized. For example Lamotrigine prevents the release of the excitatory neurotransmitter glutamate from nerve endings by interference with the cation channels.

Frequent epileptic seizures interfere with memory and academic performance and their unpredictable nature may have severe repercussions on employment and family life. Thus, treatment of epilepsy is a very important task in society. Diagnosed with epilepsy it can be difficult to obtain health insurance or to get a life insurance. People with epilepsy are not allowed to drive and often have problems getting employment. The addition of these severe social barriers results in depression and social isolation.

Although a wide variety of drugs are available for treatment of epilepsy, many patients with epilepsy still experience uncontrolled seizures. In addition, there is a need for new drugs that can halt epileptogenesis after brain injury.

Treatment of epilepsy is a very challenging area of biomedical research. A long-term drug therapy individualized for each patient and for each type of seizure is used to control the disorder. The type of medication depends on the age of the patient and the frequency of the seizures. Therapy usually starts with a low dose, which is gradually increase until the seizures become controlled. Often, a number of different drugs and concentrations must be tested on the patient before the treatment will be successful. Since all the medications currently available have negative side effects, there is an urgent need to obtain a greater understanding of the disease and to find new ways of controlling or even curing it.

Several so called "seizure related" genes have been isolated from primary cultured cerebral cortical neurons of the mouse after treatment of these cells with pentylenetetrazol (PTZ-17) (Kajiwara, K. et al., Molecular characterization of seizure-related genes isolated

by differential screening, Biochem Biophys Res Commun. 1996 Feb 27; 219(3): 795-9; Shimizu-Nishikawa K, et al., Cloning and characterization of seizure-related gene, sez-6, Biochem Biophys Res Commun. 1995 Nov 2; 216(1): 382-9; Shimizu-Nishikawa K, et al., Cloning and expression of sez-6, a brain-specific and seizure-related cDNA, Brain Res Mol Brain Res. 1995 Feb;28(2):201-10; Kajiwara K, et al., Cloning of SEZ-12 encoding seizure-related and membrane-bound adhesion protein, Biochem Biophys Res Commun. 1996 May 6;222(1):144-8; Kajiwara K, et al., Cloning and characterization of pentylenetetrazol-related cDNA, PTZ-17, Brain Res. 1995 Feb 6;671(1):170-4).

One of these proteins is SEZ-6 (seizure-related gene product 6) (sptrembl|Q62269|Q62269 SEIZURE-RELATED GENE PRODUCT 6 PRECURSOR (TYPE 1). The SEZ-6 protein has its highest expression level in brain and the expression of this family of proteins is upregulated in mice after chemically induced seizures by PTZ-17 application to the cells. sez-6 mRNA expression increases nearly two-fold within 60 minutes after the administration of PTZ. sez-6 encodes a mosaic structure protein, a transmembrane protein with a short cytoplasmic tail and with multiple motifs, five short complement-like repeats or small consensus repeats or SUSHI domains (SCRs) (reviews in Kristensen, T. et al., Fed. Proc., 46, 2463-2469, 1987; Hourcade, D. et al., Adv. Immunol., 45,381-416, 1989) and two N-terminal parts of CUB domains (Bork and Beckmann, The CUB domain. A widespread module in developmentally regulated proteins, J. Mol. Biol., 231, 539-545, 1993) as well as a Threonine rich region (14 out of 33 amino acids are threonine) close to the N-terminal part of the protein.

SCR (short complement-like repeats or small consensus repeats or SUSHI domains) are about 60 amino acid in length and consist of 4 conserved cysteines and one conserved tryptophan. The SCR domains are found in several proteins of the complement system in the selectin family which localises leukocytes to sites of inflammation and is in discussion for anti-inflammatory drug design as well as in some other proteins with different functions (see also below).

The CUB domain, also called "complement C1r/s-like repeats", was first found in C1r, C1s, uEGF, and bone morphogenetic protein and is named after these three proteins. It is an extracellular domain of approximately 110 residues which is found in functionally diverse, mostly developmentally regulated, proteins. Almost all CUB domains contain four conserved cysteines, which probably form two disulfide bridges (C1-C2, C3-C4). The

structure of the CUB domain has been predicted to be a beta-barrel similar to that of immunoglobulins. Proteins with CUB domains are found in developmentally regulated proteins and during embryogenesis and pattern formation

Both motifs play a role in protein-protein interaction, and cell adhesion and could also be important in self-recognition and/or in cooperation with SCR domains.

Antibodies against the C-terminal domain of SEZ-6 can detect a 150 kDa protein and a faint 110 kDa protein in the cerebrum and the cerebellum.

Recently, Northern Blot analyses (Herbst R and Nicklin; MJH sez-6: promoter selectivity, genomic structure and localized expression in the brain, Molecular brain research 44 (1997) 309-322) revealed sez-6 expression not only in the brain (as shown by Shimizu-Nishikawa et al., 1995), but also in testes. The more sensitive method of RT-PCR shows sez-6 expression at low levels even in kidney, liver, heart, lung and thymus. In testes, a shorter transcript was also detected.

In situ hybridization studies with 35S-labelled oligonucleotides on normal rat brain slices reveal sez-6 mRNA expression in the hippocampus, the olfactory tubercule, the piriform cortex and the islands of Calleja. The densest signals collocalized with the regions of greatest concentration of neuronal cell bodies. In the cortex, neuronal layer II, V and VI were labelled the most. The striatum, the reticular nucleus of the thalamus and the lateral sub-nuclei of the amygdala also show a prominent expression of sez-6 mRNA (Herbst R and Nicklin; MJH sez-6: promoter selectivity, genomic structure and localized expression in the brain, Molecular brain research 44 (1997) 309-322).

The group mentioned above isolated a genomic clone with the entire gene and the promoter of sez-6 was identified and analysed. Data suggests that a single gene sez-6 is conserved in the human, mouse, rat, cow and rabbit. However, in the dog, two gene copies of sez-6 can be found. In the mouse, sez-6 is localised on chromosome 11 to the genetic locus sez-6. The promoter seems to be driven by a number of transcription factors, distributed over a GC-rich region from 306 to the start site. The promoter region extends from -121 to the start site and is responsible for neural tissue specific expression. An element between -306 and -121 enhances this specificity. Regions between 3.1 and 5.7 kilobases (kb) further upstream reduce the efficiency of the promoter. They seem to

contain inhibitory elements and elements for constitutive, cell-non-specific gene expression. The efficiency of the promoter in neural cells over fibroblasts is in the range of 80 fold.

The intron-exon structure of sez-6 has also been determined. sez-6 consists of 17 exons, and the transcriptional unit of sez-6 is 49 kb. There is a perfect correspondence between the SCR and the CUB domains and the structure of the gene. Their corresponding nucleotides on the gene lie neatly within the boundaries of exons, implicating a link between the structural domains on the gene and their functional domains in the protein, which domains have probably been assembled via exon reassortment during evolution.

The function of the SEZ-6 protein is still not known, however SEZ-6 could be involved in the RAS-dependent signaling pathway. Ras was originally detected as a homologue to the protein 21 (p21) of the murine Rous sarcoma virus. The ras superfamily includes monomeric proteins that bind GTP and GDP and elicit structural changes that are used as cellular "on" and "off" switches. The short cytoplasmic tail contains a potential binding motif (Asn-Pro-Thr-Tyr) for the phosphotyrosine interaction domain of the signal transduction adaptor molecule Shc (= SH2-containing sequence) Shc is involved in the activation of Ras in response to many growth factors. Shc contains two phosphotyrosine binding domains, an Src homology 2 (SH2) domain in the carboxyl terminus of the protein and a phosphotyrosine binding (PTB) domain in the amino terminus, both of which are known to interact with phosphorylated tyrosines. Both, SEZ-6 type 1 and type 2 protein, contain this sequence which is a potential target for tyrosine phosphorylation by members of the Src family of thyrosine kinases (Src = sarcoma genes, originally identified in the Rous sarcoma virus). The phosphorylated motif could then bind to the N-terminal phosphotyrosine interaction domain of Shc.

The following variants of SEZ-6 have been cloned by RT-PCR:

1. -sptrembl|Q62223|Q62223 SEIZURE-RELATED GENE PRODUCT 6 TYPE 2 PRECURSOR.
2. -sptrembl|Q62224|Q62224 SEIZURE-RELATED GENE PRODUCT 6 TYPE 3 PRECURSOR.

The Type 2 precursor has two additional sequences at the 3'region, leading to a longer and different amino acid sequence at the C-terminus compared to type 1. Type 3 cDNA has the same 3' sequence as type 2, but has one additional insertion, which contains an

in-frame stop codon. Type 3 protein is the shortest seizure-6 related gene product. Because of the occurrence of hydrophobic amino acids near the C-terminal region, type 1 and type 2 proteins are classified as transmembrane proteins, whereas type 3 protein is assumed to be a secreted molecule since it does not contain these amino acids.

In 1995, SEZ-6 was the first protein in the nervous system that contained both SCR and CUB domains together in one protein. It is a novel type of membrane protein with multiple motifs.

Now, another homologous protein called KIAA0927 PROTEIN FRAGMENT, with similar domain structures as SEZ-6, has been identified in the human brain (see: Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,"Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro", DNA RES. 6:63-70(1999)). This protein fragment is highly homologous to SEZ-6.

Furthermore, a protein with the Accession number AL023513 has been identified within the databases of the human genome sequencing project, using FASTA or Blast search from the EBI (European Bioinformatics Institute in Hinxton. Uk) or the NCBI (National Center for Biotechnology Information) WEB page. The human DNA sequence from clone 268D13 on chromosome 22q11.21-12.2 contains a novel mouse sez-6 LIKE gene.

Another cDNA fragment called HFBDF14 of 340bp, also found using FASTA or Blast search is also found to be homologous to sez-6. On amino acid level, using the same reading frame as in sez-6 the level, the homology is 91%. On DNA level the homology is 84%. HFBDF14 is supposed to code for a part of the human counterpart of sez-6.

Furthermore, a number of expressed sequence tags (also named ESTs, short DNA sequences which are derived from cDNAs) from the fetal brain are homologous to the sez-6 DNA. The loci of these two ESTs are Z99410 and R90750 (given by the database of expressed sequence tags (dbEST), maintained at the National Center for Biotechnology, see above). They are both human EST sequences.

Other neural proteins which have either SCR or CUB domains are neurophilin (A5 antigen) and Hikaru genki (Drosophila protein). Both proteins are involved in the formation of a precise neural network.

Until now, most attempts at identifying new cell type specific cell membrane proteins have relied on either nucleic acid based approaches to homology screen cDNA libraries or immunological approaches using antibodies raised to membrane fractions to screen for favourable staining patterns in immunofluorescence. Immunological tools have been employed in the screening of expression libraries to clone proteins of interest. Alternatively, new brain receptor proteins had to be purified in large enough quantities and analysed directly by Edman degradation.

Due to recent advances in biological mass spectrometric techniques the sensitive and unambiguous identification of large numbers of proteins and de-*novo* sequencing of proteins has become possible. For the present invention, a rapid screening of proteins was employed with mass spectrometry from a standard one dimensional SDS-PAGE gel whereby very low amounts of protein (in the picomole range, just detectable by silver staining) could be analysed. Combining general biochemical membrane purification procedures, reverse phase chromatography, MALDI and tandem nanoelectrospray mass spectrometry a novel membrane receptor and two further members of its family from the human brain were identified.

Summary of the invention

In the present invention, a proteomics screening approach for membrane receptors in the brain has been employed to isolate and identify three new genes. The present invention introduces a novel membrane receptor molecule isolated from the human brain, named PSK-1, as well as two additional members of the family PSK-2 and PSK-3. PSK-1 is a highly glycosylated type I transmembrane receptor with several interesting domains. It contains a signal sequence, a threonine rich domain, three CUB- and five Sushi-domains, which are all located extracellularly. The short cytoplasmic tail contains a potential thyrosine phosphorylation site with the common consensus NPXpY, which could link the PSK-1 protein to the Ras-dependent signaling pathway. PSK-2 is a truncated form of PSK-1. It lacks three out of the five Sushi-domains and most strikingly, the whole transmembrane and cytoplasmic part. The signal sequence is intact and the PSK-2

molecule is secreted from the cell. PSK-3 is identical to PSK-1 except for a deletion of 13 amino acids just N-terminal of the transmembrane region. The PSK-proteins show homology to a family of seizure related proteins, which are expressed in the brain. Their level of expression can be increased by chemically induced seizures.

A rapid screening of proteins with mass spectrometry from a standard one dimensional SDS-PAGE gel lead to the surprising detection an isolation of the proteins of the present invention, wherein very low amounts of protein (in the picomole range, just detectable by silver staining) can be analysed. Combining general biochemical membrane purification procedures, reverse phase chromatography, MALDI and tandem nanoelectrospray mass spectrometry a novel membrane receptor was identified, named PSK-1 and two further members of its family from the human brain named PSK-2 and PSK-3, which are potentially involved in the development and maintenance of the proper functions of the nervous system.

The PSK proteins of the present invention are highly glycosylated and predominantly expressed in the brain. PSK-1 and PSK-3 are type I transmembrane receptor molecules, PSK-2 is secreted outside the cells. The PSK molecules have an interesting domain structure and show homology to a family of proteins called seizure related proteins (for example, sez-6, Shimizu-Nishikawa et al., 1995b), which are upregulated in chemically induced seizures in the mouse brain. The PSK protein of the present invention will be interesting candidates for studying epileptic seizures and developing new drug targets against epilepsy. The rapid proteomics screen used in the present invention, has been proven to be a great advantage over previously applied screening techniques and opens up a new era of biological research, rendering it possible to identify single proteins or protein mixtures of whole tissues, cells, cell fractions or specifically purified protein fractions. Already known proteins can mostly be identified by the mass spectrum alone. For unknown proteins, the peptide sequence data is the key to find DNA sequence data, provided from the different sequencing projects, which will guide to the full-length DNA and protein sequence.

One aspect of the present invention are the two new polypeptides PSK-1 and PSK-3, which are receptor molecules in the human brain (The polypeptide molecules themselves were originally isolated from the mouse brain. Peptide masses and peptide sequence tags were identified via mass spectrometry. Then the program PeptideSearch, developed by

the inventor, was applied to search with these peptide masses and peptide sequence tags of the mouse proteins against a non-redundant protein sequence database (nrdb) maintained at the European BioInformatics Institute (EBI, Hinxton Park, England) and against the database of expressed sequence tags (dbEST) maintained at the National Center for Biotechnology Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894).

DNA or protein sequences were searched for which would match these peptide masses and peptide sequence tags. The sequences, which were found were human ESTs (expressed sequence tags). The human ESTs were ordered, sequenced and the complete sequence of the receptor molecules were obtained via RACE-PCR (Rapid amplification of cDNA ends) from human cDNAs). Both receptor molecules most likely play a role in epilepsy.

Another aspect of the present invention is the thrid new polypeptide PSK-2 that is a secreted polypeptide. The secreted polypeptide might act as a ligand on the new receptor molecules of the present invention, or might act as ligand binder with no effect on a cellular signal. The knowledge of thes new isolated polypeptides of the present invention can be used to construct various ways of performing diagnosis and/or prognosis of neurological disorders, eg by the use of specific nucleotide sequences in hybridization assays or by the use of specific polyclonal or monoclonal antibodies. Furthermore, identification of drug targets within the three nucleic acid sequences or fragments thereof, or within the three polypeptide sequences or fragments thereof, disclosed in the present application, will significantly increase the chance to find selective drugs to treat seizures and epilepsy and possibly other neurological disorders.

The new genes show 50% homology to the seizure-related protein *sez-6* (gi1095324, mus musculus, in the NRDB database). The homology between the proteins encoded by the genes disclosed in the present invention and SEZ-6 is 44%. Furthermore, structural features are shared between the proteins encoded by the genes of the present invention and SEZ-6. SEZ-6 has its highest expression level in the brain and the expression level is upregulated when seizures are induced in the mouse during experimental models of epilepsy. Thus, it is believed that the proteins encoded by the genes of the present invention will be involved in the generation or control of epileptic seizures.

Scientific evidence provides that the biochemical and neurological mechanisms underlying ischemia are similar to those of seizures. Thus, a further aspect of the present invention is the identification of new drug targets, within the three nucleic acid sequences and/or the three polypeptide sequences disclosed, for the treatment of ischemia.

The invention comprises a first nucleotide sequence designated *psk-1* (SEQ ID NO:1) which encodes the amino acid sequence designated PSK-1 (SEQ ID NO: 2). PSK-1 is contemplated to be a receptor sequence.

The second nucleotide sequence is designated *psk-2* (SEQ ID NO: 3). It is identical to SEQ ID NO:1 with the exception of a single base change at position 1825. *psk-2* has a T in position 1825, whereas *psk-1* has a G in position 1825. Due to this single base change, *psk-2* has a STOP codon at that position, which is 581 in the amino acid sequence. *psk-2* therefore results in a shorter translated molecule than *psk-1*.

The invention comprises an amino acid sequence designated PSK-2 (SEQ ID NO: 4) which is encoded by *psk-2* (SEQ ID NO: 3). PSK-2 is contemplated to be a secreted ligand or effector molecule. Furthermore, the invention comprises a nucleotide sequence designated *psk-3* (SEQ ID NO: 5). *psk-3* is identical to SEQ ID NO:1 with the exception of a deletion of 39 bases at positions 2365-2403 inclusive of SEQ ID NO:1. *psk-3* encodes an amino acid sequence designated PSK-3 (SEQ ID NO: 6). PSK-3 is identical to PSK-1 with the exception of a deletion of the 13 amino acids in position acids 761 to 773 of PSK-1.

This invention further provides recombinant nucleic acid vectors comprising nucleic acids encoding SEQ ID NOs:1, 3 and 5. These include four different nucleic acid domains of the DNA clone with SEQ ID NO:1 and two domains of the DNA clone with SEQ ID NO:3 cloned into the prokaryotic expression vector pGEX-4T-1. The constructs are named GST1-GST6 and the expression vectors encode the polypeptides pGST1-pGST6.

The present invention also relates to the use of all these sequences or fragments of, as well as antibody reagents against, the above mentioned sequences in diagnosis and/or prognosis, studies of drug design or drug treatment and/or the use of the molecules in prevention or reduction of seizures. A further aspect is the possible functions of the different isoforms of the disclosed sequences.

Legends to figures

The following figures further illuminate the present invention:

Figure Nr.: 1 Flow scheme of the proteomic screening approach from the tissue sequence tags. Brain tissue was homogenised. Nucleic and mitochondrial and microsomal pellet was prepared. After Triton-X 114 extraction treatment the homogenisate was applied to a reverse phase collumn and the different fractions were analysed on PAGE gels. The gels were silver stained and several bands were cut out for MALDI and nanoelectrospray (details see text).

Figure Nr.: 2 Silver stained SDS-PAGE gels of different fractions from a RP-HPLC run. Fractions 11-38 from a total of 50 fractions, which were collected after reverse phase chromatography selection, and loaded onto SDS-Mini Gels. The gels were silver stained and bands were cut out and analysed by nanoelectrospray of the bands which are marked here with circles and numbered 1-9. The marker lanes are on the outer left side of the gel and are indicated by bars from 203 kD to 29 kD.

Figure Nr.: 3 An example of protein identification using nanoelectrospray tandem mass spectrometry.
A. Mass spectrum of the tryptic peptides from an 'in-gel' digest of the band 7 in Figure 2. Marked peaks were fragmented and correspond with peptides from human PSK-1 protein. The arrow indicates one of these peaks with its MS/MS spectrum shown in B. The peaks marked with the * symbol were originally found to represent peptides from translated human EST sequences, the symbol ¤ indicates a peptide which is located in the 5' sequence region of the PSK genes, newly identified via RACE-PCR.
B. Tandem mass spectrum of the doubly charged peptide ion (m2) indicated by arrow in A. The amino acid sequence EEYA is derived from the mass difference of the nested set of peptide fragments. This

sequence, together with the start and end mass of the fragmentation series, was combined into the peptide sequence tag shown:(630,35)-EEYA-(1122,55) and searched in a non-redundant database. The retrieved sequence (VAYEELLDNR) was matched against the tandem mass spectrum. The positions of the assigned series of COOH-terminal (Y-ions) and NH_2-terminal (B-ions) are marked.

Figure Nr.: 4 Schematic representation of the human PSK clones. The coding sequence (CDS) is represented by the black line, the signal sequence is represented by a black arrow and the 3' untranslated region by the grey arrow. The three sequences are nearly identical except for a single base pair change in psk-2, which leads to a stop codon at that position and a deletion of 39 bp in psk-3. The human EST sequences that were used to identify the full length PSK clones are indicated by arrows beneath the clones.

Figure Nr.: 5 Schematic representation of the 3 PSK proteins and their domain structure. The domains are explained in the text. The amino acid range of each domain is listed in Table 3. The three peptide sequence tags, which led to the identification of these proteins are marked as squares in the PSK-1 sequence, within the CUB domains.

Figure Nr.: 6 Northern blot analysis of human psk gene expression.

(A) Multiple tissue Northern blot containing 2mg of poly(A)$^+$RNA from each of the following tissues was hybridised with the EST R90750, including the nucleic acids from base 827-2855: heart (He), brain (Br), placenta (Pl), lung (Lu), liver (Li), skeletal muscle (Mu), kidney (Ki) and pancreas (Pa).

(B) A separate human multiple tissue Northern blot containing 2mg of poly(A)$^+$RNA from the same tissues as in (A) which was hybridised to a 5 prime RACE-PCR cDNA fragment, including the

nucleic acids from the bases 1-449. Molecular size markers are indicated on the left.

Figure Nr.: 7 Northern blot analysis of human psk gene expression. Multiple tissue Northern blot containing 2mg of poly(A)$^+$RNA from each of the following tissues was hybridised to a DNA fragment corresponding to the 3' PSK2 DNA fragment, including the nucleic acids from base 327-2355 (A) and to the Actin gene probe provided from Clontech with the multiple tissue Northern blots (B), indicated as psk and actin: Spleen (Sp), thymus (Th), prostate (Pr), testis (Te), ovaries (Ov), small intestine (sI), colon (Co) and peripheral blood leucocytes (PBL). Molecular size markers are indicated to the left.

Figure Nr.: 8 Northern Blot Analysis of Human psk Gene Expression (Top Panel, A) Multiple Tissue Northern Blot, containing 2 mg of poly(A)+ RNA from each of the following tissues, was hybridised to DNA fragments corresponding to the 3' psk DNA fragment, including the nucleic acids from base 827-2855: Amygdala (Am), caudate nucleus (Cn), corpus callosum (CC), hippocampus (Hi), whole brain (Br), substantia nigra (SN), subthalamic nucleus (Sn), and thalamus (Th). (Bottom Panel, B) The same fragment was used to probe a human multiple tissue Northern Blot, containing 2mg of poly(A) + RNA from each of the following tissues: Cerebellum (Ce), cerbral cortex (CeC), medulla (Me), spinal cord, occipital pole (OP), frontal lobe (FL), temporal lobe (TL), and putamen (Pu). Molecular size markers are indicated to the left.

Figure Nr.: 9 *In situ* hybridsation on a horizontal brain section from an adult mouse (A) and on a sagittal section from a day 16 mouse embryo (B). 16mm frozen sections were hybridized with a radiolabeled 45-mer antisense oligonucleotide from the 5' region of the mouse psk gene homologue. The sense oligonucleotide of the same region, which was used as a negative control, did not give any signal (data not shown).

Figure Nr.: 10 Schematic presentation of the different domains of the PSK-1 protein, which have been chosen for antibody production. The different parts of the sequence were cloned into the bacterial GST-expression system (Pharmacia) and injected as GST:PSK-fusion proteins into rabbits. For the exact amino acid positions see example section. GST:PSK-I contains the last 27 amino acids of CCP1, CCP2 complete, and the first 7 amino acids of CCP3 as well as the two complete CUB domains CUB-2 and CUB-3. GST:PSK-II codes for the CUB-3 domain and the CCP-3-5 domains. GST:PSK-III includes domains from GST:PSK-II as well as the transmembrane and the cytoplasmic domains. GST:PSK-IV contains the CCP-3-5 domains. GST:PSK-V encodes the cytopplasmic domain only.

Figure Nr.: 11 Western blot of 293T cell lysate expressing recombinante V5/HIS-tagged PSK-3 detected with affinity purified anti-GST:PSK-V and anti-V5/HIS antibody.

A: 293T cell lysate expressing V5/HIS-tagged PSK-3 in sense direction, lane 1-4 and in antisense direction lane 5-8, detected with affinity purified anti-GST:PSK-V antibody (directed against the cytoplasmic domain of PSK-3, 1:50 dilution; lane 1, 2, 5, 6) and a second antibody, (Peroxidase-conjugated goat anti rabbit IgG). Lane 3 and 7 show detection with the second antibody (Peroxidase-conjugated goat anti rabbit IgG) and lane 4 and 8 are detected with the first antibody only.

B: 293T cell lysate expressing V5/HIS-tagged PSK-3 in sense (lane 1 and 2) and in antisense direction (lane 3 and 4), detected with anti-V5/HIS antibody. Both antibodies, the one against the C-terminal domain of the protein and the anti-V5/HIS antibody detect a band between 122 and 212 kD. The band was not found in cells trensfected with the antisense construct.

Figure Nr.: 12 Immunocytochemical localization of PSK-3 in 293T cells Cells were fixed in 3% Paraformaldehyde and permeabilized with 0.2% Triton X-100 in PBS. A-C: 293T cells were transfected with psk-3 cDNA, C-terminally tagged with the V5/HIS epitope (see methods). The PSK-3 protein was detected after 48 hr with the anti-V5 and a secondary anti-mouse FITC antibody. The anti-V5 antibody antibody recognizes the V5 epitope at the C-terminus of the tagged PSK-3. The C-terminus is predicted to be cytoplasmic in PSK-1 and -3. A: PSK-3 is localized in the Golgi and in vesicular structures close to the plasmamembrane. B and C are enlarged images of the boxed regions in A. The arrows indicate the position of the Golgi. D: 293T cells were transfected with the psk construct cloned in the opposite direction. An antisense message will be made but no protein and no tag. E: No DNA was transfected. In D and E (negative controls) the anti-V5 antibody is unable to detect a V5 tag or the tagged PSK-3 protein. F-H: 293T cells were transfected with psk-3 cDNA, C-terminally tagged with the V5/HIS epitope. The PSK-3 protein was detected after 48 hr with the anti-GST:PSK-V and a secondary anti-rabbit FITC antibody. The anti-GST:PSK-V antibody recognizes the cytoplasmic C-terminus of the PSK-3 gene and localizes the PSK-3 protein in the Golgi (see arrow in G) and vesicular structures close to the plasmamembrane (see arrow in H) as already described above for the anti-V5 antibody. G and H are enlarged images of the boxed regions in F. I: Phase contrast picture of F.

Figure Nr.: 13 Western blots of secreted PSK-2 and transmembrane PSK-3 protein. 293T cells were transfected with psk-2 and psk-3 cDNA, both tagged with the V5/HIS module. As negative controls, the same cDNAs cloned in the opposite direction into the tagged vector were also transfected. One control dish did not contain any transfected DNA. Lanes 1-5 show an aliquot of the medium of each culture dish harvested after 1 day post transfefction, and lanes 6-10 after 2 days post transfection. Lanes 1-5 and 6-10 respectively: control without DNA, psk-3, psk-3 in opposite direction, psk-2 in opposite direction, psk-2. Lane 11 shows fraction 9 out of total 15 fractions of medium, which was collected from taged psk-2 transfecetd cells and purified over Zn-chelate collumns. In lane 12, an

aliquot of total lysate from psk-3 transfecetd cells is loaded. All samples were loaded on a 7% gel, blotted and detected with the anti-V5/HIS antibody. Lane 5 and 10 and 11 show the secreted PSK-2 protein at a molecular weigth of 120 kD. In lane 12, the PSK-3 transmembrane protein is detected at a molecular weigth of approximately 150 kD. The fuzziness of the bands might be due to a high degree of glycosylation. Secreted PSK-2 can be detected after 24h and in higher amounts after 48h in the medium.

Figure Nr.: 14 Western blot of secreted PSK-2 and transmembrane PSK-3 protein. 293T cells were transfected with psk-2 and psk-3 cDNA, both tagged with the V5/HIS module. Medium was harvested after 48h of transfection and purified with the poly-HIS protein purification kit (ROCHE). The HIS-tagged proteins were eluted from the Zn-chelated collumns in 15x 1ml fractions (see examples). A. 20ml of each eluted fraction (1-15) of the psk-2 purified medium were loaded on a 10% SDS-PAGE gel and detected with the anti-V5-HRP antibody. Fraction 5-10 contain most of the eluted PSK-2 protein with its highest concentration in fraction 9. PSK-2 protein is secreted into the medium after 48h. B. 20ml of each eluted fraction (lane 1-15) of the psk-3 purified medium were loaded on a 7% SDS-PAGE gel and detected with the anti-V5-HRP antibody. PSK-3 can not be detected in any fraction, the protein is absent from the medium. It is the transmembrane form, which does not get secreted into the medium. After purification from the medium, most of the protein elutes in fraction 9 (lane 16), as shown in A. Lane 17 illustrates total cell lysate of cells transfected with tagged psk-3. It shows the tagged psk-3 membrane protein, which has a higher molecular weigth than the secreted form, above 121 kD.

Figure Nr.: 15 Western blot of PSK-2 and PSK-3 protein treated for different periods of time with Peptide:N-Glycosidase F (PNGase F). A. Cells transfected with psk-2, tagged with the V5/HIS module. The medium was harvested after 48h and purified over a Zn-chelate collumn. An aliquot of the

fraction with the highest PSK-2 concentration was taken for the deglycosylation experiment. B. Cells were transfected with psk-3, tagged with the V5/HIS module. An aliquot of the total lysate was taken for the deglycosylation experiment. A and B. Samples were denatured in 0,07M 2-mercaptoethanol/0,3% SDS. PNGase F treatment was carried out for (0min, 5min, 20min, 45min, 4h, 24h, and 48h (the last time period was only carried out for PSK-2)) at 30°C in the presence of 0,26M Tris buffert pH8 and 1,5% CHAPS, a 5 fold excess of nonionic detergent over SDS, to stabilise the enzyme. The enzyme was subsequently inactivated by boiling the samples for 5min. The samples were loaded on a 7% SDS-PAGE gel, blotted and the PSK proteins were detected with the anti-V5-HRP antibody. In A and B both, an aliquot of the pure PSK-2 containing fraction before denaturation with mercaptoethanol/SDS was loaded (last lane). In B, an aliquot of the cell lysate before denaturation (untreated) was also added as control. The fully glycosylated PSK-2 protein runs just above 121 kD, after deglycosylation, the protein shifts clearly below a molecular weigth of 121 kD. PSK-3 protein in its completely glycosylated form runs well above 121 kD. After deglycosylation of 20min, 40min, and 24h, the band of the protein shifts closer to a molecular weigth of 121 kD and migrates at the same hight as the glycosylated form of PSK-2.

Figure Nr.: 16 Immunoprecipitation of V5-HIS tagged PSK-3. Neuro 2A cells were transfected with psk-3 cDNA. Cells were lysated and incubated with anti-V5-HIS antibody and protein A beads over night. The beads were washed, the remaining proteins were eluted in a sample buffert and loaded onto a 18% PAA-gel. The gel was silver-stained. The protein band just below 200 kD, which is marked with*, is precipitated PSK-3. The strongest protein band between 50 and 79 kD (indicated with ¤) is anti-V5/HIS antibody. The protein-bands marked with arrowheads are coprecipitated and potential binding partners of PSK-3.

Detailed description

The present invention relates to three new isolated and identified genes of which two code for novel membrane receptor molecules and one codes for a secreted molecule. The receptor molecules and the secreted molecule show strong homology to a familly of proteins that are termed seizure related proteins and they are potentially involved in the control or generation of seizures such as epileptic seizures or other neurological disorders. The three disclosed nucleotide sequences encode three polypeptides PSK-1, PSK-2 and PSK-3, which are disclosed in the present application.

In the present invention, a systematic screen using mass spectrometry in a biological system has been chosen to find new membrane proteins in the embryonic mouse brain and their human homologues, which will play a role during development of the nervous system, in signaling processes of learning and memory or cell growth and differentiation of the brain. Protein bands of mouse embryonic brain membrane fractions were excised from silver stained SDS-PAGE gels. Using high mass accuracy peptide mass mapping new neural membrane receptors were found, isolated and identified. The proteins of the present invention were subsequentely sequenced by nanoelectrospray mass spectrometry yielding novel proteins in EST database searches, which are known to be related to the nervous system.

The present invention relates to PSK proteins for wich several peptide sequence tags were identified and compared against the available databases of protein and nucleic acid sequence data, including the database of expressed sequence tags (dbEST) maintained at the NCBI. Three human EST sequences were found in the database. Two of the peptide sequence tags relate to the same EST sequence in the database. One peptide tag relates to two EST sequences. The corresponding sequence parts of the ESTs do result in these peptide sequences when translated in a certain reading frame. During the process of sequencing the ESTs and cloning of the full length cDNAs, three different forms of PSK molecules were discovered and isolated. The different DNA sequences comprised in the present invention are named psk-1, psk-2 and psk-3. The corresponding protein sequences, also comprised in the invention, are named PSK-1, PSK-2 and PSK-3. The homology search in the public available databases revealed a protein family of "seizure" proteins, with the highest homology of our proteins PSK-1-3 to the mouse protein seizure 6 (SEZ-6, sptrembl|Q62269). The family of seizure related proteins

consists also of a type I transmembrane receptor, a shorter form, which lacks the transmembrane domain and another variation, which has a different cytoplasmic C-terminus (Shimizu-Nishikawa et al., 1995b). The different forms could play different roles at different stages during development or disease stages. The potentially secreted forms, which lack the transmembrane domain could function as ligands for the transmembrane receptors and induce a signal transduction cascade. The point mutation in the sequence, which leads to the loss of the transmembrane domain could for example only or preferencially exist in a certain disease state or a certain form of epilepsy. Another possibility would be the upregulation of the secreted form during a disease state. PSK-2 could then bind receptors on the plasma membrane and signal back to the cell. Preliminary data of overexpressed PSK-2 in 293T cells show cell surface binding of PSK-2. The deletion in PSK-3 in front of the transmembrane domain could lead to changes in the binding capacity of ligands, either through deletion of the binding domain or through conformational changes, due to missing amino acids. PSK-3 possibly binds different molecules than PSK-1 and they probably regulate different but closely related pathways in the cell. Another possibility is that one of the receptors is unable to signal but accumulates all the ligands during a certain developmental or disease state due to its higher abundance in the cell. Furthermore the PSK molecules could play a role in the apoptotic events during epilepsy or in the abnormal sprouting of neurons during the desease state.

The PSK proteins of the present invention further comprise a similar domain structure with SEZ-6 proteins. The PSK molecules comprises complement control protein modules (CCPs also called short consensus repeat (SCR) and CUB domains (named after **C**1r, C1s, **u**EGF, and **b**one morphogenetic protein). CUB domains are known to be expressed in developmentally regulated proteins. Short consensus repeats (SCRs) are found in members of the selectin superfamily and in twelve of the complement proteins. The selectins belong to a family of cell surface glycoproteins. Like the PSK receptor molecules, they also consist of a transmembrane region and a short cytoplasmic anchor. *In vivo* and *in vitro* studies of selectins have established their importance in a wide variety of human diseases (Gonzalez-Amaro et al., 1999). The selectins have stimulated tremendous interest because of their involvement in a wide array of interactions between leukocytes and endothelial cells. Selectins mediate the initial phase of immunocyte extravasation into inflamed sites. The identification of their carbohydrate ligands provides an opportunity to develop a novel class of anti-inflammatory drugs. In the complement system, the role of SCRs could be in recognition of C3b or C4b as all complement

proteins that interact with these contain SCRs. SCRs are also found in the Interleukin 2 receptor and blood clotting Factor XIII. Focussing on proteins, which are known to interact with these domains, it is presently contemplated that the PSK proteins are involved in inflammation processes and/or neurodegenerative diseases.

Sez-6 is upregulated during chemically induced seizures in cortical neurons. The PSK molecules are most likely involved in similar processes, like epilepsy, which is further supported by the finding that the psk mRNA is predominantly expressed in the brain during embryonic development and in the adult stage. Three of the specific structures within the brain that exhibit these high amounts of psk mRNA molecules (hippocampus, amygdala, caudate nucleus) are located deep within the cerebral cortex. Amygdala and hippocampus are part of the limbic system. The hippocampus is involved in memory storage. The amygdala coordinates the actions of the autonomic and endocrine system and is involved in emotions. The caudate nucleus together with the putamen (which shows only low psk expression) is part of the basal ganglia, which have an important role in the regulation of movement and also contribute to cognition. The thalamus, which is also an area of high psk expression, is part of the diencephalon and is thought to regulate levels of awareness and emotional aspects. It processes and distributes almost all sensory and motor information going to the cerebral cortex. The areas of high psk-mRNA expression correlate very well with areas, which are known to be effected in epilepsy. The in situ hybridization studies on the E16 mouse embryo show a brain restricted and strikingly strong expression throughout the brain, which further suggests an important role of the psk molecules during development, especially during establishment of brain function.

One embodyment of the present invention relates to antibodies which have been made against different domains of the PSK proteins and which now reveal the protein expression of the different forms in different tissues. The transfection studies on 293T cells show that the membrane and the secreted form of PSK proteins can be expressed and detected via a small V5/His tag at the c-terminus of the protein. Immunofluorescense with the anti-V5/His tag antibody and with an antibody against the cytoplasmic domain of PSK-1 show, that the membrane protein is intracellular localized in the Golgi and in vesicles close to the membrane. These are the typical compartments involved in the processing and transport of a membrane protein.

The C-terminal cytoplasmic part of the PSK-1 and PSK-2 proteins shows the common consensus motif NPXpY and is likely involved in the Ras-signaling pathway via

tyrosine phosphorylation. Furthermore the high degree of glycosylation of the PSK proteins could make them an interesting target for drug design. A preferred embodiment of the invention is thus the identification of binding partners and protein complexes associated with the PSK proteins for example will be the subject for further study.

Terms and abbreviations in this document have their normal meanings unless otherwise stated. All nucleic acid sequences are written in the direction from the 5'('stands for prime) end to the 3'end also referred to as 5' to 3'. All amino acid or protein sequences, unless otherwise designated, are written from the amino terminus (N-terminus) to the carboxy terminus (C-terminus).

"Epilepsy" describes a heterogeneous group of disorders bound together by their tendency to produce seizures. It means the condition where seizures recur and need treatment.

The word "base pair" or "bp" refers to DNA or RNA. The single letters A, C, T and G correspond to the 5-monophosphate forms of the deoxyribonucleosides (deoxy)adenine, (deoxy)cytidine, (deoxy)guanine and (deoxy)thymidine, respectively. Furthermore, the abbreviations U, C, G and T may correspond to the 5'-monophosphate forms of the ribonucleosides uracil, cytidine, guanine and thymine, respectively, occurring in RNA molecules. "Base pair" in double stranded DNA refers to a pairing of A with T or C with G. In a heteroduplex of RNA paired with DNA, base pair may refer to a pairing of A with U or C with G.

The codon usage is the standard published and internationally used single or triple letter amino acid code.

"Digestion" or "restriction" of DNA stands for catalytic cleavage of the DNA with a restriction enzyme that is specific for certain sequences in the DNA (sequence specific endonucleases). Reaction conditions and cofactors as well as buffers, substrate amounts and other requirements for these restriction enzymes are used as described in either the manufacturers manual or publications on the subject.

"Ligation" stands for the event of forming phosphodiester bonds between two double stranded nucleic acid fragments.

The term "plasmid" refers to an extrachromosomal (usually) self-replicating genetic element. Plasmids are commonly designated by a lower case "p" followed by letters or numbers. The plasmids used here are either commercially available, publicly available on an unrestricted basis or can be constructed from available plasmids following published procedures. In addition, equivalent plasmids to those described are known and are apparent to the skilled worker in this field.

"Reading frame" means: the nucleotide sequence from which translation occurs read in triplets by the translational machinery consisting of transfer RNA (tRNA), ribosomes and associated factors. Each triplet encodes a particular amino acid. Insertion or deletion of a base pair (known as a frame shift mutation) may result in two different proteins being coded for by the same DNA segment. To prevent this, the triplet codons corresponding to the desired polypeptide are arranged in multiples of three from the initiation codon. This maintains the correct reading frame.

The term "vector" stands for a nucleic acid compound used for the transformation of cells. A vector contains a polynucleotide sequence corresponding to appropriate protein molecules which, when combined with appropriate control sequences, confer specific properties on the host cell to be transformed. Plasmids, viruses and bacteriophage are suitable vectors. Artificial vectors can be constructed by cutting and joining DNA molecules from different sources using restriction enzymes and ligases. The term vector also includes recombinant DNA cloning vectors and recombinant DNA expression vectors.

The term "Recombinant DNA cloning vector" refers to any autonomously replicating agent, also including plasmids and phages comprising a DNA molecule to which one or more additional DNA segments can be or have been added.

The term "Recombinant DNA expression vector" refers to any recombinant DNA cloning vector in which a promoter has been incorporated to control the transcription of the inserted DNA.

The term "Promoter" stands for a DNA sequence which directs transcription of DNA to RNA.

The term "expression vector system" refers to a recombinant DNA expression vector in combination with one or more transacting factors that specifically influence transcription, stability or replication of the recombinant DNA expression vector. The trans-acting factor can be expressed from a co-transfected plasmid, virus or other extrachromosomal element. It may also be expressed from a gene integrated within the chromosome.

The name "transcription" refers to the process of transferring information from a DNA sequence to a complementary RNA strand.

"Translation" refers to the process in which the genetic information of messenger RNA is used to specify and direct the synthesis of a polypeptide chain.

"Transformation" means the introduction of DNA into an organism in a way that it can replicate itself, either as an extrachromosomal element or by chromosomal integration. Methods of transforming bacterial and eukaryotic hosts are well known. Many of these methods, such as nuclear injection, protoplast fusion or calcium treatment with calcium chloride are summarized in J. Sambrook, et al., Molecular cloning: a laboratory manual, (1989).

"Transfection" is the operation of adding an expression vector to a host cell, whereby the cell itself takes up the DNA molecule and either integrates it into the chromosomes (stable transfection) or keeps it separate from the chromosomal DNA in the nucleus as a transient transcribed plasmid (transient transfection). Many different methods of transfection are known, for example calcium phosphate co-precipitation and electroporation. A successful transfection can be monitored by a specific site introduced in the transfected DNA molecule which can be recognised by an antibody. This DNA site can either be a specifically introduced "flag" or "tag" meaning a specific sequence which is easily recognised by an antibody. It could also be an intrinsic part of the expressed gene, recognised by a specific antibody. Another possibility would be the fusion to a fluorescent protein like eg GFP (green fluorescent protein).

The expressions "complementary" and "complementarity" refer to the matter of base pairing of purines and pyrimidines that associate through hydrogen bonding in double stranded nucleic acid. The base pairs guanine and cytosine, adenine and thymidine as well as adenine and uracil are complementary

"Hybridisation" is a process in which a strand of nucleic acids joins with a complementary strand through base pairing. The degree of complementarity of two non-identical, but very similar strands, and their length influence the hybridisation conditions.

"Isolated amino acid sequence" refers to any amino acid sequence, constructed or synthesised, which is locationally distinct from the naturally occurring sequence.

"Isolated DNA" refers to any DNA sequence, constructed or synthesized, which is locationally distinct from its natural location in genomic DNA.

"Isolated nucleic acid compound" refers to any RNA or DNA sequence, constructed or synthesized, which is locationally distinct from its natural location.

A "primer" is a nucleic acid fragment, which functions as an initiating substrate for enzymatic or synthetic elongation.

A "probe" is a nucleic acid compound or fragment, which hybridizes with a nucleic acid compound that encodes either the entire sequence of a gene.

An "oligonucleotide" is a single stranded DNA fragment that can act as a primer or a probe.

The term "PCR" refers to the known polymerase chain reaction The technique uses a thermally stable polymerase and specific reaction conditions combined with precise thermal cycling of the reactants to target a specific piece of DNA and then to use that piece as a template to produce billions of copies.

"RACE-PCR" stands for rapid amplification of cDNA ends and is a method used to find unknown flanking sequences of cDNAs in either 5' or 3' direction.

In the present context, the term "purified polypeptide" means a polypeptide preparation which contains at most 5% by weight of other polypeptide material with which it is natively associated (lower percentages of other polypeptide material are preferred, eg at most 4%, at most 3%, at most 2%, at most 1% and at most ½%). It is preferred that the substantially

pure polypeptide is at least 96% pure, *ie* that the polypeptide constitutes at least 96% by weight of total polypeptide material present in the preparation, and higher percentages are preferred, such as at least 97%, at least 98%, at least 99%, at least 99,25%, at least 99,5% and at least 99,75%. It is especially preferred that the polypeptide is in "essentially pure form", *ie* that the polypeptide is essentially free of any other antigen with which it is natively associated, *ie* free of any other antigen from bacteria belonging to the tuberculosis complex. This can be accomplished by preparing the polypeptide by means of recombinant methods in a non-mycobacterial host cell as will be described in detail below, or by synthesising the polypeptide by the well-known methods of solid or liquid phase peptide synthesis, eg by the method described by Merrifield or variations thereof.

By the term "a polypeptide" in the present application is generally understood a polypeptide of the invention, as will be described later. It is also within the meaning of "a polypeptide" that several polypeptides can be used, ie in the present context "a" means "at least one" unless explicitly indicated otherwise.

The term "homology " indicates a quantitative measure of the degree of homology between two amino acid sequences of equal length or between two nucleotide sequences of equal length. If the two sequences to be compared are not of equal length, they must be aligned to the best possible fit. The sequence identity can be calculated as $\frac{(N_{ref}-N_{dif})100}{N_{ref}}$, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned, and wherein N_{ref} is the number of residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC (N_{dif}=2 and N_{ref}=8). A gap is counted as non-identity of the specific residue(s), ie the DNA sequence AGTGTC will have a sequence identity of 75% with the DNA sequence AGTCAGTC (N_{dif}=2 and N_{ref}=8). Sequence identity can alternatively be calculated by the BLAST program, eg the BLASTP program (Pearson W.R and D.J. Lipman (1988) PNAS USA 85:2444-2448)(www.ncbi.nlm.nih.gov/cgi-bin/BLAST). In one aspect of the invention, alignment is performed with the global align algorithm with default parameters as described by X. Huang and W. Miller. Adv. Appl. Math. (1991) 12:337-357, available at http://www.ch.embnet.org/software/LALIGN_form.html.

The present invention provides new receptor molecules of the human brain which seem to play a role in epilepsy. The possible targets for drug treatment in the field of seizures and epilepsy and possibly other neurological disorders are hereby extended.

The proteins with SEQ ID NOs: 2 and 6 are human receptor molecules which are anchored in the cell membrane, designated as PSK-1 and PSK-3, respectively. The protein with SEQ ID NO: 4 is designated as PSK-2 and is a truncated form of PSK-1. PSK-2 is probably the secreted form of PSK-1 which means neither anchored in the cell membrane – like the receptor molecules PSK-1 and PSK-3 – nor stored in the cytoplasm of a cell. Only a single base pair is different in the DNA sequence SEQ ID NO:3 compared to the DNA sequence of SEQ ID NO:1. This difference is at base 1825, a mutation from a G, as described in SEQ ID NO:1, to a T in SEQ ID NO: 3. Due to this single base pair change (point mutation) the triplet GAG encoding the amino acid glutamic acid (Glu in triple letter code or E in single letter code) changes to TAG which does not encode any amino acid but is a stop codon. The amino acid sequence of PSK-2 stops at this codon, leading to a protein which has 580 amino acids compared to PSK-1 with 853 amino acids. The DNA sequence of SEQ ID NO:5 is identical to the DNA sequence of SEQ ID NO:1 with the exception of a deletion of 39 bases, which equals 13 amino acids. PSK-3 does not contain the amino acids 761 to 773 (both included) which are found in PSK-1.

This invention includes also recombinant nucleic acid vectors comprising nucleic acids encoding SEQ ID NOs:1, 3 and 5.

Furthermore, the following recombinant nucleic acid vectors are disclosed:

GST1: nucleotides 829-1827 of SEQ ID NO:3 (encoding the polypeptide pGST1 having amino acid sequence 249-580 of SEQ ID NO:4)

GST2: nucleotides 829-2646 of SEQ ID NO:3 which has a stop codon at position1825-1827 (encoding the polypeptide pGST2 having amino acid sequence 249-580 of SEQ ID NO:4)

GST3: nucleotides 1450-2445 of SEQ ID NO:1 with an additional stop codon (TAG) inserted at the end (encoding the polypeptide pGST3 having amino acid sequence 456-787 of SEQ ID NO:2)

GST4: nucleotides 1450-2646 of SEQ ID NO:1 (encoding the polypeptide pGST4 having amino acid sequence 456-853 of SEQ ID NO:2)

GST5: nucleotides 1804-2358 of SEQ ID NO:1 also with insertion of an additional stop codon TGA (encoding the polypeptide pGST5 having amino acid sequence 574-758 of SEQ ID NO:2) and

GST6: nucleotides 2518-2646 of SEQ ID NO:1 (encoding the polypeptide pGST6 having amino acid sequence 812-853 of SEQ ID NO:2)

The amino acid sequence of pGST1 is 332 amino acids. The amino acid sequence of pGST2 is 332 amino acids; even if the GST2 clone's DNA template is longer, it contains a STOP codon which results in a sequence of the same length and the same amino acid composition as GST1. pGST3 is also 332 amino acids, pGST4 consists of 398 amino acids, pGST5 consists of 185 amino acids and pGST6 has a length of 42 amino acids. The amino acid sequences of pGST1, pGST2, pGST3 and pGST5 span extracellular domains of the protein SEQ ID NO:1. pGST4 includes part of the extracellular domains, the transmembrane region and the cytoplasmic part. The amino acid sequence of pGST6 spans the intracellular domain of the protein.

SEQ ID NO: 5, from base 79 to base 2604 is cloned into an eukaryotic expression vector referred to as pcDNA3.1/V5/His-TOPO (Invitrogen).

In addition to the deoxyribonucleic acids of SEQ ID NOs: 1, 3 and 5, and GST1-6, this invention also provides ribonucleic acids (RNA) which comprise the RNA sequence. Since GST1-6 are partial clones of SEQ ID NOs: 1 and 3, they are not particularly mentioned with their RNA sequence. Their RNA sequence is also included in this present invention.

This invention also relates to the complementary nucleic acids of all the above mentioned sequences which are designated to have a SEQ ID, or a fragment of either SEQ IDs or a complement thereof, especially GST1-6.

This invention also provides nucleic acids, RNA or DNA, which are complementary to SEQ ID NO:1, 3 and/or 5 and GST1-6 or to their corresponding RNA sequences.

This invention also provides DNA or RNA probes and primers useful for molecular biology techniques. A compound which encodes for SEQ ID NO: 1, 3 and/or 5 or a complementary sequence of SEQ ID NO:1, 3 and/or 5 and/or any of GST 1-6 or fragments thereof, which is at least 18 bp in length and will selectively hybridize to *psk* genes or mRNA of human origin, is provided. Preferably the 18 or more base pair compound is DNA.

The receptors and the potentially secreted molecule of the present invention are believed to be useful in identifying compounds useful in the treatment or prevention of conditions associated with seizures and epilepsy as mentioned above.

The receptors are embedded in the plasmamembrane, with the N-terminal part extracellular and the C-terminal part cytoplasmic (type 1 membrane proteins). The potentially secreted protein could act on the above mentioned receptor molecules as a ligand and possibly introduce a signal transduction cascade either via phosphorylation of the C-Terminal part of the receptor molecules or via heterodimerization of the receptor molecules with a signalling molecule. Another possibility is that it acts on other receptors in the membrane.

The psk genes were originally identified via a mass spectrometric screen of membrane proteins isolated from embryonic mouse brain. The methodology of this screen is described in Schrotz-King, P. et al., Use of mass spectrometric methods for protein identification in receptor research, Journal of Receptor and Signal Transduction Research, 19(1-4), 659-672 (1999). Three peptide sequences were identified and they were compared against the available databases of protein and nucleic acid sequence data, including the database of expressed sequence tags (dbEST) maintained at the National Center for Biotechnology Information (NCBI) (National Center for Biotechnology Information, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894).

The homology search in the public available databases revealed a protein family of "seizure" proteins, with the highest homology of our proteins PSK-1-3 to the protein called seizure 6 (SEZ-6, sptrembl|Q62269|Q62269 SEIZURE-RELATED GENE PRODUCT 6 PRECURSOR (TYPE 1). The SEZ-6 protein has its highest expression level in the brain and the expression of this family of proteins is upregulated in mice after chemically induced seizures.

Furthermore two of the peptide sequences of the present invention matched within the same EST sequence in the database. Said EST could result in these peptide sequences when translated in a certain reading frame. If more than one peptide sequence tags match the same EST sequence in the database it is very likely that this EST codes or partially codes for the protein of interest or the protein.

Peptide No. 3, which was found via mass spectrometry matched a second, different EST sequence in the database.

The loci of the two found ESTs are Z99410 and R90750 which are both human EST sequences.

Further EST searches and alignments were done at:
http://www.hgmp.mrc.ac.uk/ESTBlast/A tool for contig building with Expressed Sequence Tags, created at Glaxo Wellcome by the Bioinformatics Department.

Skilled artisans will be able to synthesise the proteins of the present invention, given the disclosed nucleotide and protein sequences, by a number of different methods. In one embodiment, the amino acid compounds are synthesised by well-known chemical methods, including solid phase peptide synthesis, or recombinant methods. The principles of solid phase chemical synthesis of polypeptides are well known and are found in publications in this field, eg in H. Dugas and C. Penney, Bioorganic chemistry, (1981) Springer Verlag, New York, pgs. 54-92. For example peptides may be synthesized by solid-phase methodology utilizing an Applied Biosystems 430A peptide synthesizer (Applied Biosystems, now called Perkin Elmer Applied Biosystems).

In one embodiment of the present invention, each of the proteins disclosed are produced by recombinant methods. Recombinant methods are used if a high yield is desired. A general method of how to get any desired DNA sequence is provided in J. Brown, et al. or J. Sambrook, et al., in Methods in Enzymology, 68:109 (1979) or Current Protocols in Molecular Biology, published by John Wiley & Sons, Edited by: Fred M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, Kevin Struhl, Series Editor: Virginia Benson Chanda; 0-471-50338-X; Loose-leaf 0-471-30661-4; CD-ROM.

The basic steps in the recombinant production of a desired protein are:

- Construction of synthetic or semi synthetic DNA encoding the protein of interest (for example via polymerase chain reaction (PCR)).
- Integration of this DNA into an expression vector so that the protein can either be expressed alone or as a fusion protein (a fusion protein is a protein formed by expression of a hybrid gene made by combining two gene sequences together. Typically this

is accomplished by cloning a cDNA into an expression vector in frame with an existing gene).

- Transformation of an appropriate eukaryotic or prokaryotic host cell with the mentioned expression vector.
- Culturing the above mentioned transformed or transfected host cell in a manner to express the protein of interest. A cell free *in vitro* system can also here be considered as a source of protein production.
- Recovering and purification of the recombinantly produced protein of interest.

Normally, prokaryotes are used for cloning of the DNA sequences and constructing the vectors for this invention. Prokaryotes can also be involved in the production of the protein of interest.

In addition to E. Coli strains, bacilli, such as Bacillus subtilis, other enterobacteriaceae such as Salmonella typhimurium or Serratia marcescans, as well as various Salmonella species, can be used in the cloning procedure. In addition to these gram-negative bacteria, other bacteria, especially Streptomyces,ssp., may be used in the prokaryotic cloning and expression of the proteins in this invention.

The expression of a foreign gene in E.coli requires the insertion of the gene into an expression vector, usually a plasmid. Generally the vector contains the following elements:

1. sequences encoding a selectable marker that assures maintenance of the vector in the cell.
2. a controllable transcriptional promoter (for example: lac (β-lactamase), trp (tryptophan), or tac), which can produce large amounts of mRNA from the cloned gene, upon induction.
3. Translational control sequences, like a ribosome binding site and an initiator ATG.
4. A polylinker to enable the insertion of the gene in the correct orientation within the vector.

Promoters which are used for prokaryotic hosts include the β-lactamase (vector pGX2907 (ATCC 39344) contains the replicon and the β-lactamase gene) and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter system [vectorpATH1 (ATCC 37695) is designed to facilitate expression of an open reading frame as a trpE fusion protein under the control of a trp promoter] and hybrid promoters, like the tac pro-

moter [isolatable from plasmid pDR540 (ATCC 37282)]. However, other functional bacterial promoters, whose nucleotide sequences are generally known, enable the skilled person in the field of science to ligate them to DNA encoding the proteins of the instant invention, using also linkers and/or adapters to create any restriction site of interest. Promoters for use in bacterial systems will also contain a Shine-Dalgarno/RBS sequence operably linked to the DNA encoding the desired polypeptides. These examples are illustrative rather than limiting.

The proteins of this invention may be synthesised either by direct expression or as a fusion protein, whereby, in the latter, the protein of interest will appear as a translational fusion with another protein or peptide, which may be removable by enzymatic or chemical cleavage. A preferred prokaryotic expression system was the pCR®2.1-TOPO vector from Invitrogen with the LacZα gene.

A second preferred prokaryotic expression vector was the pGex-4T-1 vector from Amersham Pharmacia Biotech, using the *tac* promoter which is induced using the lactose analog isopropyl b-D-thiogalactoside (IPTG). This system is a gene fusion system with Glutathione S-Transferase (GST), an integrated system for the expression, purification and detection of fusion proteins produced in E.coli.

The production of certain peptides as a fusion protein in recombinant systems often prolongs the life span of the fusion protein, increases the yield of the desired peptide and provides in addition a convenient way of purifying the protein of interest. A variety of peptidases (for example Thrombin, Trypsin, Factor Xa) which cleave a polypeptide at a specific site or digest the peptides from the amino or carboxy termini (eg diaminopeptidases) of the peptide chain are known. Also particular chemicals, like cyanogen bromide, will cleave a polypeptide chain at specific sites. A person trained in scientific experimentation will be able to introduce modifications to the amino acid sequence (and to the semi-synthetic or synthetic coding sequence if recombinant means are employed) to incorporate site-specific internal cleavage sites. For reference, see eg P. Carter, " Site specific proteolysis of fusion proteins", Chapter 13 in Protein purification: From molecular mechanisms to large-scale processes, American Chemical Society, Washington, D.C. (1990).

In addition to cloning and expressing the genes of interest in the prokaryotic systems discussed above, the proteins of the present invention can also be produced in eukaryotic

systems. The invention presented here is not limited to the use in a particular eukaryotic host cell. A variety of eukaryotic host cells are available from places like the American Type Culture Collection (ATCC) and are suitable for use with eukaryotic expression vectors of the present invention. The choice of the host cell is to some extend dependent on the particular expression vector used to drive the expression of the PSK proteins encoding nucleic acids of the present invention. Host cells, which amongst others can be used with the present invention, are listed in the following table:

Table 4

Host Cell	Origin	Source
293	Human Embronal Kidney	ATCC CRL1573
HeLa	Human Cervix Epitheloid	ATCC CCL 2
HepG-2	Human Liver Hepatoblastoma	ATCC HB 8065
RPM18226	Human Myeloma	ATCC CCL 155
HS-Sultan	Human Plasma Cell Plastocytoma	ATCC CCL 1484
HCN-1A	Human cortial tissue (Brain)	ATCC CRL 10442
HCN-2	Human cortial tissue (Brain)	ATCC CRL 10742
CV-1	African Green Monkey Kidney	ATCC CCL 70
COS-7	African Green Monkey Kidney Fibroblast Cells, SV40 transformed	ATCC CRL-1651
LLC-MK2	Rhesus Monkey Kidney	ATCC CCL 7
Neuro-2A	Mouse Neuroblastoma, Brain	ATCC CCL-131
C127I	Mouse Fibroblast	ATCC CCL 1616
3T3	Mouse Embryo Fibroblasts	ATCC CCL 92
PC12	Rat Pheochromocytoma (Adrenal Gland)	ATCC CRL-1721
H4IIEC3	Rat Hepatoma	ATCC CCL 1600
BHK-21	Baby Hamster Kidney	ATCC CCL 10
CHO-K1	Chinese Hamster Ovary	ATCC CCL 61

A broad variety of vectors exist for the transformation of such mammalian host cells. But neither the above mentioned host cells nor the specific vectors described below are intended to limit the scope of the present invention.

Today the most useful vectors contain multiple elements like:

1. An SV40 origin of replication useful for amplification to high copy number in COS cells.
2. An efficient promoter element to initiate high level transcription.
3. mRNA processing signals like mRNA cleavage and polyadenylation sequences.
4. Polylinker containing multiple restriction endonucleases sites for insertion of foreign DNA
5. Selectable markers that are used to select cells that have stably integrated the plasmid DNA.
6. Plasmid sequences that permit propagation in bacterial cells.

In addition to the above mentioned properties, most vectors contain also an inducible expression system, regulatable via an external stimulus, and furthermore some vectors are equipped with one or more short recognition sequences or "tags" (for example: GFP-tag (green fluorescent protein), HA-tag (Flu hemagglutinin), FLAG-tag, HIS-tag, c-Myc-tag (from the myc oncogene), Protein A-tag (Immunoglobulin binding domain of Staphylococcus aureus protein A), (for review on tags see: E. Harlow and D. Lane, Using antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory Press1999) which can be translated into a short antigenic site recognized by a specific antibody. These tagged vectors are constructed in a way that the tag or tags are situated at the N- or the C-terminus, translated in frame with the protein of interest. Antibodies against the tag-site enable the detection of the protein in cells as well as the purification of the protein from cells, for example via immunoprecipitation.

Furthermore, vectors containing the green fluorescent protein tag (GFP-tag) can be used to visualize the protein of interest in the cell.

The GFP gene used in the vectors available from Invitrogen is described in Crameri *et al.*, 1996. In this paper, the codon usage was optimized for expression in mammalian cells and three cycles of DNA shuffling were used to generate a mutant form of GFP that has the following characteristics:

1. Excitation and emission maxima that are the same as wild-type GFP (395 nm and478 nm for primary and secondary excitation, respectively, and 507 nm for emission).
2. High solubility in E. coli for visual detection of transformed cells (if expressed from a promoter recognized by *E. coli.*

3. >40-fold increase in fluorescent yield over wild-type GFP.

This GFP protein will subsequently be referred to as SuperGFP to differentiate it from wild-type GFP.

In a preferred embodiment of the invention, Cytomegali Virus promoter (CMV) is used for induced transcription sequences. A number of promoters have been identified and engineered into expression vectors. Examples are sequences inducible by β-interferon, heat shock, heavy metal ions, glucocorticoids and by tetracyclin (Tet-system).

The choice of a certain expression system is dependent on the experimental goal. The transient expression in COS cells is a convenient approach to study general functional activity. If a large amount of protein (>1mg) is required, then stable coamplification in COS cells is a good approach. If the gene is potentially cytotoxic, high level expression can be achieved through vaccinia virus vectors or inducible promoter-vector systems (for example: Tet-system, tetracycline-controlled transactivator (tTA) or the reverse tTA (rtTA) available from CLONTECH Laboratories, Inc.1020 East Meadow Circle, Palo Alto, CA 94303, USA).

The pSV2-type vectors contain segments of the simian virus 40 (SV40) genome that constitute a defined eukaryotic transcription unit-promoter, intervening sequence and polyadenylation site. In the absence of the SV40T antigen, the plasmid pSV2-type vectors transform mammalian and other eukaryotic host cells by integrating into the host cell chromosomal DNA. A wide variety of pSV2-type vectors are available, such as plasmid pSV2-gpt, pSV2-neo, pSV2-dhfr, pSV2-hyg or pSV2-β-globin, in which the SV40 promoter drives transcription of an inserted gene. These vectors can be used with the coding sequence of the present invention. They are available from sources such as the ATCC or the Northern Regional Research Laboratory (NNRL), 1815 N. University Street, Peoria, Illinois, 61604.

PSV2-dhfr (ATCC 37146) contains a murine dihydrofolate reductase (dhfr) gene under the control of the SV40 early promoter. The dhfr gene will be amplified or copied in the host chromosome under certain conditions. This amplification can result in the synthesis of closely-associated DNA sequences, which makes it possible to increase the production of a protein of interest (J. Schimke, Cell 35: 705-713 (1984)).

A wide variety of plasmids are available which can be used in plasmids constructed for expression of the proteins of the present invention in mammalian and other eukaryotic host cells. For example the SV40 promoter, promoters from eukaryotic genes like the estrogen-inducible chicken ovalbumin gene, the interferon genes, the glucocorticoid-inducible tyrosine aminotransferase gene, the thymidine kinase gene and the major early and late adenovirus genes can be readily isolated and modified to express the genes of this invention.

Eukaryotic promoters in tandem can also be used to drive expression of a coding sequence of this invention. In addition a large number of retroviruses that infect a wide variety of eukaryotic host cells are known. The LTRs (long terminal repeats) in the retroviral DNA frequently encode functional promoters and might therefore be used to drive expression of the nucleic acids of the present invention. The invention described in this application is by no means limited to the use of particular promoters exemplified herein.

PRSVcat (ATCC37152) contains portions of a long terminal repeat of the Rous Sarcoma virus, a virus that infects chicken and other host cells. This long terminal repeat contains also a promoter, which can be used in the vectors for this invention (H. Gorman et al., Proceedings of the National Academy of Sciences (USA), 79: 6777, 1982).

PMSVi (NRRL B-15929) comprises the long terminal repeat of the Murine Sarcoma virus, a virus known to infect mouse and other host cells.

The mouse metallothionein promoter has also been well characterized for use. It is present in the plasmid pdBPV-MMTneo (ATCC 37224) which can serve as the starting material of other plasmids of the present invention.

An expression system preferred in this invention is the topoisomerase-activated, pcDNA3.1/V5/His-TOPO vector system (Shuman, S. (1994) Journal of Biological Chemistry 269: 32678-32684, Covered under U.S. patent 5,487,993 and corresponding foreign patents. Other patents pending (assigned to Memorial Sloan-Kettering and licensed exclusively to Invitrogen)) which includes the following features to simplify protein expression, detection, and purification in mammalian cells:

- Strong CMV promoter for high-level, constitutive expression.
- Polyadenylation signal and transcription termination sequence from the bovine growth hormone gene (BGH) to enhance mRNA stability.
- C-terminal V5 epitope tag for detection of recombinant proteins with the Anti-V5 Antibody .
- C-terminal polyhistidine tag for purification using ProBond™
- nickel-chelating resin and detection with the Anti-His (C-term) Antibody
- T7 RNA promoter/priming site for *in vitro* transcription of sense RNA and for sequencing inserts.

Another expression vector system employs one of a series of vectors containing the BK; an enhancer derived from the BK virus, a human papovavirus. The more advanced version of these vectors also contains the adenovirus-2-early region 1A (E1A) gene product. The E1A gene product (which actually consists of two products here collectively summarized as one product) is an immediate-early gene product of adenovirus, a large DNA virus.

Yet another embodiment of the present invention is the utilisation of the phd series of vectors. They comprise a BK enhancer in tandem with the adenovirus late promoter to drive expression of useful products in eukaryotic host cells. E.coli K12 GM48 cells harboring the plasmid phd are available as part of the Northern Regional Research Laboratory under accession number NRRL-18525. The phd series of plasmids functions most efficiently when introduced into a host cell which produces the E1A gene product, cell lines such as AV12-664, 293 cells and others, see also above.

Transformation of the mammalian cells can be performed by any of the known processes including, but not limited to, the protoplast fusion method, the calcium phosphate co-precipitation method, electroporation, liposome-mediated transfection and more, described for example in Current Protocols in molecular biology, Wiley Interscience. Other ways of production are well known to experienced people in the field.

In addition to the plasmids mentioned above, one embodiment of the present invention utilises viruses as appropriate vectors. For example the adenovirus, the adeno-associated virus, the vaccinia virus, the herpes virus, the Rous sarcoma virus and the baculovirus are useful. The baculovirus based expression system taken as an example is an eukaryotic

expression system and thus uses many of the protein modification, processing and transport systems present in higher eukaryotic cells. The baculovirus expression system uses a helper-independent virus which can be propagated to high titers in insect cells adapted for growth in suspension cultures, making it easy to obtain large amounts of recombinant protein. Several alternate methods of expression are described in J. Sambrook et al, see above at 16.3-17.44.

In addition to prokaryotic, mammalian and insect (Drosophila) host cells, microbes such as yeast cultures may also be used in an alternate embodiment of the invention. The imperfect fungus Saccharomyces cerevisiae, or common baker's yeast, is the most commonly used eukaryotic microorganism. For expression in Saccharomyces pombe, eg the plasmid Yrp7 (ATCC-40053) is often used (See for example L. Stinchcomb, et al., Nature, 282:39(1979); J: Kingsman et al., Gene, 7:141(1979); S. Tschemper et al., Gene, 10:157(1980)). This plasmid already contains the *trp* gene, which provides a selectable marker for a mutant strain of yeast lacking the ability to grow in tryptophan.

In combination with the above mentioned considerations, the selection of suitable promoter sequences which can be used with yeast hosts is important. In one embodiment of the invention, the promotor sequence is the promoter for 3-phoshoglycerate kinase (on plasmid pAP12BD (ATCC 53231) and described in U.S. Patent No. 4,935,350) or other glycolytic enzymes such as enolase (on plasmid pAC1 (ATCC 39532)), glyceraldehyde-3-phosphate dehydrogenase (derived from plasmid pHcGAPC1 (ATCC 57090, 57091)), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phsphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase and glucokinase, as well as the alcohol dehydrogenase and pyruvate decarboxylase genes of Zymomonas mobilis (United States Patent No. 5,000,000 issued March 19, 1991).

Inducible yeast promoters have the additional advantage of their transcription being controllable by varying growth conditions. These are for example the promoter regions of alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein (on pCL28XhoLHBPV (ATCC 39475) and described in United States Patent No. 4,840,896) glyceraldehyde 3-phosphate dehydrogenase and enzymes responsible for maltose and galactose (for example GAL1 on plasmid pRY121 (ATCC 37658) utilization). More about vectors and promoters used in

yeast expression can be found in R. Hitzeman et al., European patent publication No. 73.657A. Enhancer elements in yeast such as the GAL UAS from Saccharomyces cerevisiae (found with the CYC1 promoter on plasmid Yepsec1-hI1beta (ATCC 67024)) are also advantageously used with yeast promoters.

In addition to the above-mentioned expression systems, the cloned cDNAs may also be employed in the production of transgenic animals, usually mice or sheep, in which expression or overexpression of the proteins of the present invention can be assessed. The nucleic acids of this invention can also be used in the construction of "knockout" animals, in which the expression of the native cognate of the gene is suppressed.

Persons skilled in the art will recognise that some changes of SEQ ID NOs:1, 3, 5, as well as of *psk* fragments of GST1-6, will fail to change the function of the amino acid compound. For example, some hydrophobic amino acids could be exchanged against other hydrophobic amino acids, or acidic amino acids could be exchanged against acidic amino acids. Those altered amino acid sequences, which confer substantially the same function in the same manner as the here described amino acid compounds with SEQ ID NO:2, 4 and 6 and pGST1-6, are also encompassed within the present invention. Typically, such conservative substitutions preserve the secondary or tertiary structure of the polypeptide backbone, the charge or hydrophobicity of the residue or the bulk of the side chain. The following table exemplifies such conservative substitutions of amino acids. The resulting protein products are functional equivalents of the proteins of SEQ ID NO:2, 4 and 6 and pGST1-6.

Table 5

Original Residue	Exemplified Substitution
Ala	Ser, Gly
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro, Ala
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Mel	Leu, Ile
Phe	Met, Leu, Gyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

There are various ways of introducing these substitutions into the proteins described in this invention, for example during chemical synthesis or via chemical modification of an amino acid side chain after the protein has been prepared.

Another possibility is the alteration of the nucleic acid compounds which encode these proteins. These mutations of the nucleic acid compounds can be generated by either random mutagenesis techniques, for example with chemical mutagens, or by site-specific mutagenesis involving nucleotides. The nucleic acid compounds, which present the same function in the same way, as the exemplified nucleic acid compounds are also included within the present invention.

One particular aspect of the present invention is nucleic acid compounds which comprise isolated nucleic acid sequences which encode SEQ ID NOs: 1, 3 and 5 and GST1-6. The amino acid sequence of this invention can be coded for by a multitude of different nucleic

acid sequences. This is due to that fact that most of the amino acids are encoded by more than one nucleic acid triplet due to the degeneracy of the amino acid code. Since these alternative nucleic acid sequences would encode the same amino acid sequences the present invention also includes these alternate nucleic acid sequences.

The nucleic acids encoding the PSK molecules may also be produced synthetically. This synthetic synthesis of nucleic acids is well known in the art (see eg: E. L. Brown, R. Belagaje, M. J. Ryan and H. G. Khorana, methods in Enzymology, 68: 109-151, 1979). The DNA segments corresponding to the invention can be generated using conventional DNA synthesizing apparatus such as the Applied Biosystems Model 380A or 380B DNA synthesizers (commercially available from Applied Biosystems, Inc., 850 Lincoln Center Drive, Foster City, CA 94404) which employ phosphoramidite chemistry. There is also a possibility of using the more traditional phosphotriester chemistry to synthesize the nucleic acids of this invention. (See for example M. J. Gait, ed., Oligonucleotide synthesis, A Practical Approach, (1984).

In one embodiment of the invention, the synthetic *psk* genes are designed to possess restriction endonuclease cleavage sites at either end of the transcript to facilitate isolation from and integration into expression and amplification plasmids. The restriction sites are chosen in a way that the coding sequences of the *psk* genes are properly orientated to achieve proper in-frame reading and expression of the PSK molecules. A variety of other such cleavage sites can be incorporated depending on the particular plasmid constructs employed and can be generated by well-known techniques.

In another aspect, the desired DNA sequences can be generated using the polymerase chain reaction (PCR) as described in U.S: Patent No. 4,889,818.

It is contemplated that truncations of the nucleic acid and amino acid sequences of the present invention could be undertaken by experienced people in the field. This means selective cloning of certain areas or domains and their separate expression or in conjunction with other nucleic acid sequences. These alterations which use parts of the present invention are also enclosed within this present invention

One embodiment of the invention relates to the preparation of the disclosed ribonucleic acids, or fragments thereof, by using the polynucleotide synthetic methods discussed

above. An alternative embodiment relates to the enzymatical preparation using RNA polymerases to transcribe a DNA template complement thereof.

One method of choice for preparing the nucleic acids of the present invention is to employ RNA polymerase from the bacteriophage T7 or the bacteriophage SP6. Both these RNA polymerases are highly specific and require the insertion of bacteriophage-specific sequences at the 5'end of the message to be read. See J. Sambrook et al., as above, at 18.82-18.84.

"Selectively hybridise" as used above can refer to two situations:

1 The nucleic acid compound described above hybridises to the human *psk* genes or mRNAs under more stringent hybridisation conditions than the same nucleic acid compound would hybridise to an analogous *psk* gene of another species, for example murine.

2 Secondly, the *psk* probes hybridise to the *psk* genes or mRNAs under more stringent hybridisation conditions than other related compounds, including nucleic acid sequences encoding for example the *sez-6* gene.

Probes and primers of the described kind can be prepared enzymatically as described above. Preferentially, these probes and primers are synthesised using chemical means as described above. Probes and primers of defined structure can also be bought commercially.

The present invention also includes recombinant DNA cloning vectors and expression vectors comprising the nucleic acids of the present invention. Many vectors encompassed in this invention are described above. The preferred nucleic acid vectors are those which are DNA. The most preferred recombinant DNA vectors in this invention are the TOPO Type vectors from Invitrogen for prokaryotic and eukaryotic expression as well as for expression in Drosophila and the GST expression vectors from Pharmacia. Both systems are mentioned above.

Any plasmid containing the genes of the present invention are readily modified to construct expression vectors that produce PSK proteins, in a variety of organisms, including, for example, E.coli, Sf9 (as host for baculovirus), Spodoptera and Saccharomyces. The

current literature contains techniques for constructing AV12 expression vectors and for transforming AV12 host cells. Eg United States Patent No. 4,992,373 describes these techniques.

One of the most widely applied techniques to alter a nucleic acid sequence is the oligonucleotide-directed site-specific mutagenesis. B. Comack, Current Protocols in Molecular Biology 8.01-8.5.9., (F. M. Ausubel, et al., eds., 1991). In this technique an oligonucleotide, whose sequence contains the mutation of interest, is synthesized as described above. This oligonucleotide is then hybridized to a template containing the wild-type sequence. Preferentially, the template is single stranded. Plasmids which contain regions such as the f1 intergenic region are used preferentially since this region allows the generation of single-stranded templates when a helper phage is added to the culture to harbor the "phagemid" (a phagemid is a plasmid containing a phage).

The oligonucleotide anneals to the template and DNA-dependent DNA polymerase is then used to synthesize the second strand complementary to the template DNA using the oligonucleotide as a primer.

The product is a DNA heteroduplex which contains a mismatch due to the mutation of the oligonucleotide. The DNA gets replicated in the host cell and two types of plasmids originate from this heteroduplex, the wild type and the newly constructed mutant. Applying this technique makes it easy to introduce convenient restriction sites which allows the coding sequence to be placed immediately adjacent to any transcriptional or translational regulatory elements.

Protocols used for E.coli can be followed to construct analogous vectors for other organisms. Therefore the regulatory elements can be substituted using well-known molecular biology techniques.

The present invention provides host cells which contain the nucleic acids supplied with this invention. So far these are E.coli cells (Top10F'cells, from Invitrogen) and 293T cells (from human embryonic kidney), but transfection of the *psk* genes (in different eukaryotic expression vectors) into several eukaryotic cell lines, especially neuronal cell lines, will follow. The present invention also provides 6 constructs comprising different regions and different functional domains of the SEQ ID NO:1 and 3 which are cloned into prokaryotic

expression vectors (pGex-4T-1) fused in frame with the GST gene in a system called the GST gene fusion system (see above). The SEQ ID NOs of these constructs are mentioned above.

One aspect of the invention relates to antibodies against the amino acid sequences disclosed in the present application. In a particular aspect, the invention relates to antibodies against the amino acid sequences encoded by the GST constructs, GST1, GST2, GST3, GST4, GST5 and/or GST6. The antibodies are contemplated to be beneficial as diagnostic tools or as ligands that will block or activate the function of the herein described molecules. The antibodies will also be used to find the interaction partners of the herein described molecules and to develop functional assays.

Not only the protein fragments mentioned above but also the full length or other fragments or even smaller peptide sequences could be used as antigens to raise antibodies against these sequences. With the word "antibodies", the following is included: antibodies, fragments of antibodies (such as, but not limited, to Fab, Fab', Fab2, and Fv fragments) and chimeric, humanized, veneered, CDR-grafted or resurfaced antibodies that are able to bind antigens of similar nature as the parent antibody molecule from which they derive. Also single chain polypeptide binding molecules are included in this current invention. Thereby the term "antibody" is not limited by the way they are produced, for example whether they are produced *in situ* or not, encompassing those antibodies produced by recombinant DNA technology, including but not limited to expression in bacteria, yeast, insect cell lines or mammalian cell lines.

The procedures for making monoclonal and polyclonal antibodies are well known in the art and can be found in several manuals and books for example in E. Harlow and D. Lane, Antibodies, A Laboratory Manual, (Cold Spring Harbor Laboratory), 1988; or J. Goding, Monoclonal Antibodies, Principles and Practice, (Academic Press, 1983); C. Milstein, Handbook of Experimental Immunology, (Blackwell Scientific Pub., 1986)

Chimeric antibodies are described in U.S Patent No. 4,816,567 (March 28, 1989, S. Cabilly, et al.) which encloses methods and vectors for the preparation of the chimeric antibodies and which is incorporated herein by reference. U.S. Patent No. 4,816,397 (issued March 28 1989, M. Boss et al.) shows the simultaneous co-expression of heavy and light

chains of antibodies in the same host cell. It is an alternative approach to produce genetically engineered antibodies and is with its entire content incorporated herein by reference.

A further refinement of this method was the genetic engineering of the reactive monoclonal antibodies, European Patent Publication No. 0 239 400, September 30, 1987 by Winter. Thereby the complementarity determining regions (CDRs) of a human antibody get replaced with the CDRs of a murine monoclonal antibody. The specificity of the human antibody converts to the specificity of a murine antibody.

The single chain antibody technology joins the binding region of heavy and light chains with a polypeptide sequence to generate a single polypeptide having the binding specificity of the antibody from which it derived. This technology is well known (see for example R. E. Bird et al., Science 242:423-426 (1988); PCT Publication No. WO 88/01649, published 10 March, 1988; U. S. Patent 5,260,203, issued November 9, 1993) and the entire content is incorporated herein by reference.

The techniques mentioned provide numerous ways to generate molecules which have the binding characteristics of the parental antibody but affording a less immunogenic format.

These antibodies are used in diagnostics, therapeutics or in diagnostic/therapeutic combinations. By " diagnostics" is meant the testing related to either the *in vitro* or *in vivo* diagnosis of disease states or biological status in mammals, preferably in humans.

"Therapeutics" or "therapeutic/diagnostic combinations" used in this context means the treatment or the diagnosis and treatment of disease states or biological status by the *in vivo* administration of the antibodies of the present invention to mammals, most preferably humans

The antibodies of the present invention are especially preferred in the diagnosis and/or treatment of conditions associated with an excess or deficiency of the PSK receptors or the potentially secreted form. A "family" of antibodies against different forms and different domains of the PSK proteins makes it possible to develop a variety of assay systems for detecting binding partners of these PSK molecules, for example via immunoprecipitation techniques, or a radioactive labelled anti-PSK-molecule-antibody bound to solid phase PSK receptor molecule could be displaced by adding a potential antagonist.

Furthermore, to find the active interaction sites the protein can be immobilized and the different binding partners could be differentially eluted by competing with various peptides of the protein.

Antibodies against the hypervariable region of anti PSK antibodies, (so called anti-idiotypic antibodies) would resemble the original epitope, the PSK receptor or the secreted molecule. Such antibodies would be very useful in evaluating the effectiveness of compounds which are potential antagonists, agonists or partial agonists of the PSK receptors and of the potentially secreted molecule (see for example: Wasserman et al., Proceedings of the National Academy of Sciences (USA), 79: 4810 (1982); or Cleveland, et al., Nature (London), 305:56 (1983)).

A variety of other assay systems are also adaptable to detect agents which bind the PSK receptors (for examples see: E. Harlow and D. Lane, Using Antibodies, Cold Spring Harbor Laboratory Press, 1999; J. Langone and H. Vunakis (eds.), Methods in Enzymology, Vol. 73, Part B; the contents are incorporated herein by reference). (See example 8).

The part of the mouse homologue was RACE-PCR cloned out of a mouse Marathon-Ready™ cDNA from Clontech (Mouse 17-day Embryo Swiss-Webster/NIH #7460-1) and is about 600 nucleic acids in size.

The nucleic acid compounds of this invention may also be used to hybridize to genomic DNA which has been digested with one or several restriction enzymes and run on an electrophoretic gel. This hybridization of a probe, the probe being labelled either radioactive or non-radioactive onto restricted DNA, wherein the DNA has been blotted to a membrane, is called Southern hybridization and is well known to people in the field (see for example Current Protocols in Molecular Biology, mentioned above).

Southern blotting in context with the nucleic acids of the present invention could be used to search for individuals with mutations in these receptors by the well-known technique of restriction fragment length polymorphism (RFLP). These procedures are described in U.S. Patent 4,666,828., issued May 19, 1987, the entire contents are incorporated herein by reference.

Due to the 44% homology to SEZ-6 and especially due to the striking similarity in their domain structure, we conclude that the PSK proteins could play an essential role in epilepsy.

The secreted form could be the activator or inhibitor of the PSK receptor molecules and activate or inhibit a signal transduction cascade. Another possible role of the secreted form of the PSK could be as a soluble ligand binder which compete with the membrane bound receptor molecules of PSK. Furthermore, the appearance of different forms of one molecule like with the PSK molecules speaks strongly for the involvement at a certain disease state (see for example: R. Vidal et al., A stop-codon mutation in the BRI gene associated with familial British dementia, Nature 24 June, 1999).

Furthermore, to test these hypothesis an assay is on the way to be developed. The assay will include a method which can determine whether a substance is a functional ligand of the PSK molecules. The binding activity has to be monitored by physically detectable means, for example radio-labelling or enzymatic or fluorescent labelling of the antibodies against certain domains and competition with chemical substances including the ones which are known to or supposed to play a role in epilepsy treatment.

To undertake such a screen, the PSK molecules will be prepared using recombinant DNA technology and will possibly be immobilized on a column or on beads. The PSK molecules will then be incubated together with antibodies against themselves and a test compound is added to compete with the antibody. Unbound molecules will be washed away and this eluent will be scored for radioactivity or other added chemical labels. If the antibody is labelled radioactively or with a fluorescence marker, the presence of label in the eluate shows that the test compound has bound to the PSK molecules by competing with the antibody. This method can also be performed the opposite way that the test compound is labelled instead of the antibody.

An important issue in this assay is that the conformation of the proteins is conserved during the assay. For membrane receptors this would be possible in a hydrophobic environment, employing natural or synthetic detergents, membrane suspensions, micelles and the like. Preferred detergent complexes include the zwitterionic detergent CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate) and sodiumdeoxycholate.

A compound with a dissociation constant (K_i) of less than 10 nM is generally considered an excellent candidate for drug therapy. On the other hand a compound, which has a lower affinity, but which is selective for the particular receptor, can be a better candidate.

A different assay of estimating the efficiency of certain chemical drugs or compounds in binding the PSK molecules would be:

Transfecting of a mammalian host cell with an expression vector containing the DNA encoding the PSK molecules.

- Culturing the host cell under conditions that the PSK molecules are expressed.
- Adding a test compound onto the transfected cells.
- Measuring the change in a physiological condition, like up- or downregulation of the PSK molecules or other molecules in the cells, which molecules are involved in certain signaling processes.

The present invention further relates to the generation of cell lines expressing any of the genes presented in this invention in a regulated manner, with the goal of transplanting/implanting these cells into mammals and testing them for their efficacy of treatment of neurological disorders or ultimately for the treatment of neurological disorders.

Additionally, gene transfer of the *psk* genes of the present invention into mammals is made to test for their efficacy in treatment of neurological disorders.

In another aspect of the present invention, DNA chip array assays with the nucleic acids bound to a reusable solid support can be made, allowing for improvements in automation and data analysis. The DNA chip technology allows biomedical researchers to study entire genes and gather information about expression and gene mutation patterns. DNA chip technology requires hybridization of an unknown nucleotide sequence to an ordered array of known DNA immobilized onto a solid surface, eg a glass slide or silicon chip. The hybridization results are detected by fluorescence and analyzed by pattern recognition. DNA chip assays are also referred to as biochips, oligonucleotide array assays and hybridization array assays.

Applications of these array methods are:

Gene Expression Patterns

To detect changes in patterns of gene expression that may occur during neurological disorders and other disease states, an array assay can be used to monitor gene expression in different human cells and tissues.

Gene Mutation Patterns

DNA array analysis has been widely used to detect genetic mutations.

Human Diagnostics

The DNA chip array assay has a tremendous potential in human diagnostics for monitoring changes in gene expression within specific cell types, tissue types, physiological states and genetic make-up of the patient.

Sequencing by Hybridization

To detect a gene expressed at levels as low as one copy per cell. A series of short, overlapping probes are hybridized and bioinformatically reconstructed to recreate the DNA sequence.

Furthermore, the polypeptides of the present invention can be used in The ProteinChip System. The ProteinChip System is based on SELDI-TOF-MS (Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry) technology. The system uses small arrays or plates with chemically or biologically treated surfaces to interact with proteins. Unknown proteins are affinity captured on treated surfaces, desorbed and ionized by laser excitation and detected according to molecular weight. Known proteins are analyzed using on-chip functional assays. The Implications: This technology is closing the existing information gap between genomics and proteomics. Applications include protein composition studies, (differential protein display), target and marker protein discovery and identification, purification design and monitoring, receptor ligand identification, phosphorylation assays, protein-DNA interactions, immunoglobin capture, protein modifications, peptide mapping and sequencing, etc.

The ProteinChip Array contains hundreds of multiple, addressable locations of chemically or biologically defined protein binding surfaces arranged as wells of 1mm diameter or smaller. The arrays contain various kinds of functional groups – from antibodies and re-

ceptors to chromatographic surfaces – to capture proteins. Minute quantities of samples such as serum, urine or tissue extracts, can be incubated directly on the chip without having to purify the proteins first. After the chip has had a chance to bind molecules in the sample and unbound material is washed away, it is treated with short bursts of laser light, releasing and ionizing the proteins which are accelerated to a detector. The time-of-flight required for proteins to reach the detector is calculated to determine the molecular weight of each released protein. The result is graphically displayed as a protein profile of all captured proteins from a particular sample. The ability to use multiple array surfaces amplifies the ability to rapidly define complex protein compositions by simultaneously investigating a variety of purification conditions.

Examples

Example 1

The screening approach

Preparation of plasma membrane proteins of mouse E16 brain

Total brain tissue from CD1 mice at embryonic day 16 (E16) were prepared in ice cold PBS (phosphate buffered saline) and homogenized in 25mM MES-buffer (4-Morpholineethanesulfonic acid) at pH 6.5 containing 5mM DTT, 2mM EDTA, 0.3M sucrose and a cocktail of protease inhibitors (Boehringer Mannheim). The homogenate was incubated on ice over night in 125U/ml Benzonase (Merck). Post nuclear supernatant (PNS) was obtained by spinning the homogenate at 800 x g at 4°C. To reduce the amount of mitochondrial proteins, the PNS was loaded onto a 1.2M sucrose layer, buffered with 25mM MES, pH 6.5, 5mM DTT and 2mM EDTA also containing protease inhibitors and centrifuged at 10.000 x g at 4°C.

The resulting supernatant contains now most of the soluble proteins. The pellet, also called microsomal fraction, consists of all remaining membranes, like plasma membrane and intracellular components (ER, Golgi, endosomes and lysosomes). Many membrane proteins are highly hydrophobic and poorly soluble in water. Therefore they require high concentrations of solubilising agents such as urea or detergents to remain in solution. Triton-X-114 (TX-114) is a non-ionic detergent that solubilises proteins into protein-detergent micelles at temperatures below the critical micelle concentration point (cmc-value) of 22 0C. At temperatures above the cmc-value, micelles merge and the solution separates into two phases. Hydrophobic proteins partition into the detergent phase, hydrophilic proteins into the aqueous phase. The microsomal pellet was extracted twice with TX-114 via phase separation, which was performed by centrifugation through a sucrose cushion as described by Bordier (1981). The membranes accumulate and concentrate in the TX-114 detergent pellet. The membranes were reconstituted to a detergent concentration of 0.25-0.5% TX-114, precipitated in 12% TCA final concentration and stored at –20 degrees C until required.

Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC)

For reverse phase chromatography, a MZ-Microcolumn (length 100mm, ø 1mm, silica based VYDAC, 300 Å pore size, C8 Carbon chains, 5µm particle size) (MZ-Analysentechnik) was used. The column was attached to a solvent delivery system (140B, Applied Biosystems), an automated fraction collector for micro-preparative peptide-HPLC (Gilson/Abimed 231XL sampling injector), developed on the base of a pipette robot Gilson 221XL (Gilson, Abimed) and a programmable absorbance detector (783A, Applied Biosystems). Special peak detection software was used to take the information of the analog entry of the fraction collector and to guide the automated and exact fractionation into a microtiter plate.

The TX114 extracted, TCA precipitated membranes were dissolved in 65% formic acid and separated by RP-HPLC using a gradient of formic acid-water (buffer A: 65% formic acid / 35% H2O) and formic acid-acetonitrile (buffer B: 65% Formic acid / 35% acetonitrile) as the mobile phase. The buffers A and B were chosen according to Lee and colleges (6). Samples were injected with 40% buffer B and the gradient was started with 40% buffer B for 10min isocratic and then run for 90min up to 95% buffer B, with a flow rate of 30µl/min and 38 bar pressure. At 95% and 100% buffer B fractions were collected for further 10min. The peak detector was set to 280nm and a sensitivity of 0.01. Fractions between 50 and 150µl were collected by an automated fraction collector into a 96 well plate. The fractions were dried in a speed vac, reconstituted in Laemmli sample buffer, heated to 50°C for 30min and run on a 10% SDS-PAGE gel. SDS-PAGE was performed according to the method of Laemmli (1970), using a mini Protean II gel apparatus (Bio-Rad, Munchen, Germany). All samples were run under reducing conditions.

Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC), an elegant way to get membrane proteins solubilized and separated.

The extreme hydrophobicity of this membrane fraction made it difficult to further purify or fractionate the proteins over ion-exchange or size-exclusion chromatography

The TCA precipitated membranes were dissolved in 65% formic acid and separated by RP-HPLC using a gradient of formic acid-water (buffer A) and formic acid-acetonitrile (buffer B) as the mobile phase. Samples were injected with 40% buffer B and the gradient was started with 40% buffer B for 10min and then run for 90min up to 95% buffer B, with a flow rate of 30µl/min and 38 bar pressure. At 95% and 100% buffer B fractions were collected for further 10min. The peak detector was set to 280nm and a sensitivity of 0.01.

An automated fraction collector collected fractions between 50 and 150µl into a 96 well plate. The high volatility of the solvent was a further advantage in the purification, since the solvent could be removed without any loss of protein material. The fractions were dried, heated in Laemmli sample buffer to 50°C for 30min and run on a 10% SDS-PAGE gel. All samples were run under reducing conditions, the gels were silver stained and several bands were cut out for mass spectrometric analysis.

Identification of proteins by MALDI and Nanoelectrospray

SDS-PAGE gels were silver stained, bands were excised from the gel, in-gel reduced, S-alkylated and in-gel digested with an excess of trypsin (Boehringer Mannheim) overnight at 37°C as described in Wilm et al. (1996) and Shevchenko et al. (1996). The digested proteins were further processed according to the fast evaporation method of Vorm et al. (1994) with the modifications introduced by Jensen et al. (1996). Mass spectra from these peptide mixtures were obtained on a modified Matrix Assisted Laser Desorption Ionization 'Time-of-Flight' (MALDI TOF) mass spectrometer equipped with delayed ion extraction (REFLEX, Bruker-Franzen, Bremen, Germany). Spectra were calibrated on matrix ion peaks of known mass. For peptide sequencing by nanoelectrospray tandem mass spectrometry, gel pieces were further extracted. The resulting peptide mixture was concentrated, desalted and eluted into a nanoelectrospray spraying needle as described previously (Wilm et al., 1996 and Wilm and Mann, 1996). Peptide sequence tags (Mann and Wilm, 1994) were assigned in peptide fragmentation spectra obtained on an API III triple quadrupole or a quadrupole time-of-flight mass spectrometer (Perkin-Elmer Sciex, Toronto, Canada). The program PeptideSearch, developed by our group, was applied to search with peptide masses and peptide sequence tags against a non-redundant protein sequence database (nrdb) maintained at the European BioInformatics Institute (EBI, Hinxton Park, England) and against the database of expressed sequence tags (dbEST) maintained at the National Center for Biotechnology Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894).

Identification of brain receptor molecules by nanoelectrospray

The use of MALDI and Nanoelectrospray techniques in protein research allows the efficient microanalysis of proteins also in large-scale experiments. About 50 silver stained bands of different intensity and size of several RP-HPLC fractions were excised 'in-gel'-digested with Trypsin and analyzed by MALDI and / or Nanoelectrospray.

For MALDI, 1-2% of the tryptic digest was applied to the probe of the mass spectrometer. Since unmodified trypsin was used as the protease, known autodigestion products of trypsin were used to 'internally mass calibrate' the spectrum. The mass-to-charge ratio of the peptides released from the gel was measured with high mass accuracy. A mass accuracy of better than 30 parts per million (ppm) can usually be achieved by this procedure. The nrdb database was searched for proteins with a mass range between 50 and 300 kd corresponding to the molecular weight of the proteins excised from the gel. No restriction was made during the search concerning species of origin. The complete set of peptide masses from each protein band was compared to the tryptic peptide masses predicted for each protein. As a general criterion for positive identification by mass maps alone, at least five peptides are required to match calculated tryptic peptides for an individual protein in the database, within 50 ppm deviation in mass between measured and calculated values. Furthermore, the matching peptides should account for at least 15 percent of the sequence. Other criteria, such as the percentage of the total number of the observed peaks that can be assigned to a putative match, and the similarity in molecular weight of the unknown protein to the putative match are also considered. Post-translational modifications have to be taken into account as the relative mobility of a protein on a gel can differ from the calculated mass of the corresponding protein found in the database. Multiple proteins, which comigrate in one band can also be identified. Therefore the peptide masses, which were assigned, to the first identified protein are removed from the complete list of masses and the database gets scanned again applying the remaining masses. Following the initial identification of a protein, a second path search was performed. In this step, incomplete tryptic cleavage and peptide modifications that could alter the peptide masses were calculated for the identified protein and compared to the measured masses. Such modifications could be for example oxidized methionine or S-acryamidocysteine. modified peptides, which were identified in the second pass search were added to the list generated in the first path search to increase the number of matching peptides and sequence coverage.

In the samples where none or only tentative identifications could be made by MALDI peptide mapping sequence information was generated via electrospray tandem mass spectrometry. For this purpose the peptide mixture was further extracted. The peptides were concentrated, desalted and eluted into a nanoelectrospray spraying needle. Peptides were electrosprayed into a tandem mass spectrometer, which selects a peptide ion based on its mass and yields a mass spectrum of fragments of that ion. The

nanoelectrospray ion source is a miniaturized form of electrospray, which allows analysis of extremely small quantities of proteins and peptides (8, 12). The fragment or tandem mass spectrum contains sequence information, which can be extracted by subtracting the mass of series of fragments, which differ, by one amino acid molecular weight from each other.

In this way a partial sequence is obtained which can be submitted together with the mass spectrometric information to a search program. The resulting 'peptide sequence tag' is then searched in sequence databases for a matching pattern. Finally, fragment masses calculated for the retrieved sequences were compared to the complete fragmentation spectrum to confirm the match. In cases where searching in protein sequence databases did not yield any results, the database of expressed sequence tags was searched (dbEST), which contains at least fragmentary information of most human and murine sequences. Figure 1 illustrates the screening approach in a flow scheme caricature

Identification of the PSK-molecules, a new family of type I membrane receptors and secreted proteins

An example of the screening procedure described above is shown in Figure 2. Nine proteins from one RP-HPLC preparation were analyzed by nanoelectrospray tandem mass spectrometry. More abundant proteins were excised that were of high molecular weight (above 80 kD) to gain the maximum from this strategy.

Table 1 shows the results, which have been obtained by nanoelectrospray on the excised proteins from Figure 2. It includes the peptide sequences, which were found by searching the NRDB or EST databases with peptide sequence tags. Samples 1-4, 6, 8 and 9 are known proteins, CAML1, NR-CAM, sample 3 equals sample 2, Mannose Receptor, Herpes Virus entry protein, CD 166 and Contactin and/or Axonin precursor, respectively. They are all either transmembrane or membrane associated and known to play important roles in the nervous system. Sample 5 is a new protein, which is currently under investigation and will be described elsewhere. Sample 7 from the mouse brain membrane fractions is also a new protein, which was named PSK-1, a protein of about 150 kd. The peptide tags, which are corresponding to PSK-1 are marked in bold letters. The corresponding EST sequences in the database originated from human brain. The MS and a MS/MS spectrum of this PSK-1 protein is given in Figure 3 A and B as an example for the nanoelectrospray identifications. The peak hight of the measured peptides is given as relative intensity in percent. The location of the peaks in the spectrum is dependent on their mass/charge (m/z) value. A shows the MS spectrum. The peptide peaks marked with

the * symbol could be correlated with translated human EST sequences. The ¤ symbol indicates a peptide, which did not find a hit in the database, but could be correlated later to the 5' sequence of the PSK genes, which we identified via RACE-PCR. One doubly charged peptide ion was chosen for further fragmentation (indicated by an arrow). The MS/MS spectrum of this peptide is shown in B. Peptide fragments, illustrated here as peaks of Y- and B- ions, differ by the mass of one amino acid allowing therefore the readout of a partial amino acid sequence, which is indicated as EEYA. This sequence, together with the start and end mass of the fragmentation series, Y5 at 630.35 m/z and Y9 at 1122.55 m/z was combined into the peptide sequence tag shown: (630.35)-EEYA-(1122.55) and searched in a non-redundant database. The retrieved sequence (VAYEELLDNR, see also Table 1) was matched against the tandem mass spectrum. The positions of the assigned series of COOH-terminal (Y-ions) and NH2-terminal ions (B-ions) are marked. Three different forms of this protein were cloned in our laboratory and named PSK-1, -2 and -3 proteins (see below).

Table 1

Sample Nr.	Expected molecular weight from gel (kD)	Peptides found with Peptide Sequence Tags in Protein or EST databases	ES Results	Species	Accession Nr.	Molecular weight calculated from the sequence (kD)
1.	150	LDCQVQGRPQPEITWR AQLLVVGSPGPVPHLELSDR WMDWNAPQIQYR	CAML1	Mouse	P11627	140
2.	120	NALGAVHHTISVTVK LSPYVNYSFR YIVSGTPTFVPYLIK	NG-CAM related cell adhesion molecule precursor	Rat	P97686	133
3.	120		Same as sample 2	Rat	P97686	133
4.	180	SGWQWAGGSPFR IFGFANEEK SWGQASLECLR	Mannose Receptor	Mouse	Q61830	165

5.	130	Not named	EST (not named) **New Protein**	Mouse	Homology to O15031 (human)	205
6.	90	ITQVTWQK VLVATCTSANGKPPSVVSWETR GEAEYQEIR YDEDAKRPYFTVDEAEAR	Herpes Virus Entry Protein	Human	O75465	57
7.	150	**VSLDEDNDRLMVR** **ILLQVEILNVR** **EGDMLTLFDGDGPSAR** **VAYEELLDNR**	**Several ESTs, (see methods and results part)** **PSK 1-3 (new) proteins**	Human	AJ245820, AJ245821, AJ245822	92 62 90
8.	90	LDVPQNLMFGK SVQYDDVPEYK APFLETDQLK EIDPGTQLYTVTSSLEYK SSNTYTLTDVR ESLTLIVEGKPQIK	CD166 (ALCAM)	Mouse	U95030	65
9.	120	MPLLVSHLLLISLTSCLGDFTWHR ASPFPVYK GPPGPPGGLRI	Contactin precursor	Mouse	P12960	113

		IVESYQIR DGEYVVEVR FAQLNLAAEDPRL VISDTEADIGSNLR NGEPLASQNRVEVLAGDLR GPPGPPGGVVVR VSASNILGTGEPSGPSSK TTGPGGDGIPAEVHIVR	Axonin-1 precursor	Rat	P22063	113

Table 1. Nanoelectrospray results from nine silver stained SDS-PAGE gel bands of Figure 2.

The table includes from left to right, the number of the gel bands, the expected molecular weight of the excised gel band, the peptides obtained by nanoelectrospray, the identified protein name, the species, in which the protein is known and the accession number in the EMBL/NCBI databases, and the calculated molecular weight from the sequence. Differences the expected molecular weight of the gel-excised band and the predicted size from the sequence arise due to posttranslational modifications or the charge of the protein.

cDNA screening and RACE-PCR cloning

One novel protein candidate was revealed with three peptide tags three EST clones in the **human** EST database (shown below).

The following peptide sequence tags were used to identify the following ESTs:

Peptide TAG:	EST (Accession Number):	EST translated:
1.(743.44)VQL(1083.63)L(1196.77)	H10539 and HSZ99410	ILLQVEILNVR
2.(1147.53)DLS(1462.69)	R90750	VSLDEDNDRLMVR
3.(921.41)LTL(1248.63)	HSZ99410	EGDMLTLFDGDGPSAR

The I.M.A.G.E. EST clones (H10539 and R90750) were ordered from Research Genetics Inc. and completely sequenced. The EST clone R90750 contains a stop codon within the reading frame, the EST clone H10539 misses this stop codon due to one base exchange but is otherwise identical in its overlapping sequence. Both clones contained the polyadenylation site and the polyA tail. Oligonucleotides were designed within these sequences and were used to obtain 5 prime RACE- (rapid amplification of cDNA ends) PCR products from mouse brain and human brain Marathon-Ready cDNA (Clontech). RACE products were purified or cloned directly into the T/A-type PCR cloning vector pCRII-TOPO (Invitrogen) and transformed into TOP10F' cells (Invitrogen). The cells were grown overnight and the plasmids isolated with BioRad Quantum Prep (MiniPrep) or Quiagen Plasmid MidiKit and sequenced on the automatic DNA Sequencer ABI PRISM 310 Genetic Analyzer from Perkin Elmer.

The following two nested antisense primers were designed based on the human EST sequence matching bp 977-1001 (5'-CAGGTGAGGTTGGGCCTTACGGCTC-3') and bp 925-948 (5'-CAGGGTGGCATTGTGGATGGTGCC-3') in the *psk-1* gene and were used to perform RACE-PCR with the mouse and human brain Marathon-Ready cDNA (Clontech). From the new amplified mouse sequence of 553 bases (comparable to bp 292-844 of the human sequence) the oligos for in situ hybridization studies on mouse tissue were chosen (see below). From the new human sequence another RACE-PCR was performed with two sets of nested antisense primers;

set A: matching bp 925-948 as shown above together with a primer matching bp 420-447 (5'-GCTCCCCAGATCTGGAGACTCCACATAC-3') and

set B: matching bp 420-447 as above and matching bp 361-390 (5'-CAGAACTGGGCTGGTCACCGTAGTGGTAAC-3') in the *psk-1* gene. Both primer sets

gave the same 5 prime sequence with an ATG start codon and a Kozak sequence, which was assumed to be the 5 prime start site of the *psk* genes
The overlapping regions were compared to establish the correspondence between ESTs and the new sequences.

Furthermore a PCR with the human brain Marathon cDNA was performed whereby one primer was a sense primer chosen within the new 5 prime end with the presumptive ATG Start-codon (bp 79-108, 5'-GTCGCCATGGGGACTCCCAGGGCCCAGCAC-3') and an antisense primer was chosen within the EST sequence part (bp 2578-2608, 5'-CATACAGCGGGTTGCTGAAGTCCGACTCCAC-3'). A nested PCR on that sequence with the primer set matching in sense direction bp 79-108 as above and in antisense direction bp 420-447, located on the previous RACE product, see above gave a 361 bp band. This is the correct size for the two overlapping 5' RACE products.

The results from PCR amplifications were derived from at least 4 independent PCR products sequenced in both directions. The overlapping sequences were assembled with DNASTAR software SeqMan (DNASTAR Inc.).

The human EST sequences corresponding to the peptide tags of the PSK molecule were ordered from Research Genetics and completely sequenced (see methods part). EST clone with the accession number R90750 contains a stop codon within the reading frame, the EST clone H10539 misses this stop codon due to one base exchange but is otherwise identical in its overlapping sequence. Both clones contained the polyadenylation site and the polyA tail. EST clone R90750 extended further 5'. Oligonucleotides were designed within these sequences 5' RACE PCR was performed to get the start site of the genes. The obtained 5' end was verified by defining the total mRNA length and the tissue distribution of this 5'fragment compared to the EST by Northern hybridization (see below) and PCR with primers in the new 5'region and the EST part (see methods). Furthermore a peptide sequence tag, which was found in the sample before but could not find a hit in the databases (nrdb or EST) could now be assigned to the new 5' sequence of the PSK clones.

The complete sequences were assembled and submitted to the EMBL database at the European Bioinformatics Institute in Hinxton, UK. The longest clone with the stop codon at nucleotide position 2644 is named psk-1 (accession number: AJ245820). The clone derived from EST R90750, which has the stop codon further 5 prime at nucleotide

position 1825 is named psk-2 (accession number: AJ245821). A PCR across the whole gene on human brain Marathon-Ready cDNA with primers of the 5' and the 3' end revealed another clone with a small deletion of 39 bases (at position 2365-2403) of the psk-1 gene, which we named psk-3 (AJ245822). The DNA sequences were named in small characters the corresponding protein sequences were named in uppercase PSK-1, PSK-2 and PSK-3. Figure 4 shows a schematic representation of the human *psk*-clones in relation to their EST sequences.

The PSK proteins show homology to a family of seizure related proteins

Looking for homologous sequences to these EST hits the first protein neighbour sequence with the highest homology was a protein called seizure-related gene product 6 precursor (TYPE 1), Sez-6 from mouse (sptrembl|Q62269|). The SEZ-6 protein has its highest expression level in the brain and the expression of this family of proteins is upregulated in mice after chemically induced seizures by pentylenetetrazol (PTZ-17) application to the cells (Kajiwara et al., 1996). The *sez-6* mRNA expression increases nearly two-fold within 60 minutes after administration of PTZ. *Sez-6* encodes a mosaic structure protein, a transmembrane protein with a short cytoplasmic tail and with multiple motifs, five SCRs (reviewed in Kristensen et al., 1987; Hourcade et al., 1989) and two N-terminal parts of CUB domains (Bork and Beckmann, 1993) as well as a Threonine rich region (14 out of 33 amino acids are threonine) close to the N-terminal part of the protein. Several so called "seizure related" genes have been isolated from primary cultured cerebral cortical neurons of the mouse after treatment of these cells with PTZ-17 (Kajiwara et al., 1995, 1996a, 1996b; Shimizu-Nishikawa et al., 1995a, 1995b).

Example 2

Cloning of GST1-GST6

Six nucleotide fragments were cloned into the prokaryotic expression vector pGEX-4T-1. These include four different nucleic acid domains of the DNA clone with SEQ ID NO:1 and two domains of the DNA clone with SEQ ID NO:3 and the constructs are named GST1-GST6. The cloning sites were EcoRI at the 5' end and Not I at the 3' end.

GST1: nucleotides 829-1827 of SEQ ID NO:3 (encoding the polypeptide pGST1 having amino acid sequence 249-580 of SEQ ID NO:4)

GST2: nucleotides 829-2646 of SEQ ID NO:3 which has a stop codon at position1825-1827 (encoding the polypeptide pGST2 having amino acid sequence 249-580 of SEQ ID NO:4)

GST3: nucleotides 1450-2445 of SEQ ID NO:1 with an additional stop codon (TAG) inserted at the end (encoding the polypeptide pGST3 having amino acid sequence 456-777 of SEQ ID NO:2)

GST4: nucleotides 1450-2646 of SEQ ID NO:1 (encoding the polypeptide pGST4 having amino acid sequence 456-853 of SEQ ID NO:2)

GST5: nucleotides 1804-2358 of SEQ ID NO:1 also with insertion of an additional stop codon TGA (encoding the polypeptide pGST5 having amino acid sequence 574-758 of SEQ ID NO:2)and

GST6: nucleotides 2518-2646 of SEQ ID NO:1 (encoding the polypeptide pGST6 having amino acid sequence 812-853 of SEQ ID NO:2)

All were cloned 5' into the EcoRI site and 3' into the NotI site. In all of these 6 GST clones the first base after the EcoRI site is a start of the sequence and base number one of the first amino acid coding base triplet.

pGEX-4T-1 is derived from pGEX-2T and contains a thrombin recognition site. pGEX-4T-1 is together with twelve other pGEX vectors part of the Glutathione S-Transferase (GST) Gene Fusion System which is available from Amersham Pharmacia Biotech.

The GST Gene Fusion System

The Glutathione S-Transferase (GST) Gene Fusion System combines the expression, purification and detection of fusion proteins produced in Escherichia coli (E. coli). The system consists of three major components: the pGEX plasmid vectors, various options for GST purification and a variety of GST detection products. A series of site-specific proteases complements the system. The pGEX plasmids are designed for inducible, high-level intracellular expression of genes or gene fragments as fusions with Schistosoma japonicum GST.

GST occurs naturally as a 26 kilo Dalton (kDa) protein and can be expressed in E. coli with full enzymatic activity. Fusion proteins which possess the complete amino acid sequence of GST also demonstrate GST enzymatic activity and can undergo dimerization similar to that observed in nature.

Fusion proteins are easily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B contained in the GST Purification Modules. Cleavage of the desired protein from GST is achieved using a site-specific protease whose recognition sequence is located immediately upstream from the multiple cloning site on the pGEX plasmids. Fusion proteins can be detected using a colorimetric assay or immunoassay provided in the GST Detection Module, or by Western blotting with anti-GST antibody. The system has been used successfully in many applications such as molecular immunology, the production of vaccines and studies involving protein-protein and DNA-protein interactions (for reference see Amersham Pharmacia Biotech).

Thirteen pGEX vectors are available from Amersham Pharmacia Biotech. Nine of the vectors have an expanded multiple cloning site (MCS) that contains six restriction sites. All of the GST gene fusion vectors offer:

1. A tac promoter for chemically inducible, high-level expression.
2. An internal lac Iq gene for use in any E. coli host.
3. Very mild elution conditions for release of fusion proteins from the affinity matrix, thus minimizing effects on antigenicity and functional activity.
4. Precision, thrombin or factor Xa protease recognition sites for cleaving the desired protein from the fusion product.

Example 3

Alignment of *psk-1*, *psk-2*, and *psk-3*

The following EST sequences were ordered from Research Genetics, 2130 Memorial Parkway SW, Huntsville, AL 35801 and sequenced: R90750 and H10539 as well as R55720 (loci), their cDNA IDs are respectively: 167208, 46853 and 41030.

Clone R90750 contains the nucleotides from base 827 to base 2855 of the above mentioned sequence SEQ ID NO:3 and is 2029 bases long. Of these 298 bases are published.

Clone H10539 contains the nucleotides from base 1456 to base 2855 of the above mentioned sequence SEQ ID NO:1 and is 1400 bases long. Of these 466 bases are published.

Both sequences are identical in their overlapping area with the exception of a single base change. At nucleotide position 999 of R90750 is a T whereby in H10539 at this position there is a G. The T results in a TAG Stop codon whereby the G results in the continuation of the open reading frame. These sequences are contained within the SEQ ID NO:1 and 3, which results with a G at position 1825 in SEQ ID NO:1 and a T at the same position in SEQ ID NO:3.

SEQ ID NO:5 with the deletion of 13 amino acids was partially cloned via PCR from Clontech Marathon cDNA of mouse embryonic day 17 cDNA. Two PCR primers were used: 5p-GCCCAGAGATGGGCTACCTG-3p as forward primer and 5p-TGGGCGGCAGGATGCTGGTTTATTTACTGTAGG-3p as reverse primer (with KlenTaq LA Polymerase Mix, a PCR enzyme mix designed for long and accurate amplification of PCR products. The KlenTaq LA Polymerase Mix contains a primary polymerase (Klen-Taq-1 polymerase) and a minor amount of a proofreading polymerase to increase the efficiency, yield, and sensitivity of all PCR assays, provided by Clontech).

The 5 prime part identical for all three *psk* genes SEQ ID NO:1, 3 and 5 was obtained via RACE-PCR. For SEQ ID NO:5 the 5'end was ligated to the PCR product mentioned above via a Bgl II (AGATCT) site at position 435-440 of SEQ ID NO:5 which was overlapping in both clones.

Example 4

Cloning of *psk* genes

The E.coli strain **TOP10F´** is particularly useful for the prokaryotic expression of foreign proteins, but other strains can also be used and are mentioned together with their relevant genotypes in the following table:

***E.coli* strains:**

TOP10: (Invitrogen)
This strain is used for general cloning and blue/white screening without IPTG. This strain cannot be used for single-strand rescue of DNA.

F *mcr*A D(*mrr-hsd*RMS-*mcr*BC) F80*lac*ZDM15 D*lac*C74 *rec*A1 *deo*R *ara*D139 D(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str R) *end*A1 *nup*G

TOP10F': (Invitrogen)

This strain overexpresses the Lac repressor (*lac*I q gene). For blue/white screening, IPTG has to be added to the plates to obtain expression from the *lac* promoter. This strain contains the F episome and can be used for single-strand rescue of plasmid DNA containing a f1 origin.

F' *lac*I q Tn*10* (Tet R)} *mcr*A D(*mrr-hsd*RMS-*mcr*BC) F80*lac*ZDM15 D*lac*C74 *rec*A1, *deo*R *ara*D139 D(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str R) *end*A1 *nup*G_

JM109

e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17(rK- mK+) supE44 relA1 D(lac-proAB) [F' traD36 proAB lacIqZDM15

RR1 HB101 RecA+

X1776

tonA53 dapD8 minA1 glnV44 (supE44) D(gal-uvrB)40 minB2 rfb-2 gyrA25 thyA142 oms-2 metC65 oms-1 (tte-1) D(bioH-asd)29 cycB2 cycA1 hsdR2

LE392 e14-(McrA-) hsdR514 supE44 supF58 lacY1 or D(lacIZY)6 galK2 galT22 metB1 trpR55

294 endA thi hsdR

DH5α (F$^-$) Hanahan, 1983)

supE44, ΔlacU169 (Φ80 lacZ□M15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1

HB101^f (Boyer and Roulland-Dussoix, 1969)

Δ(gpt-proA)62, leuB6, thi-1, lacY1, hsdS$_B$20, recA, rpsL20, (Str^r), ara-14, galK2, xyl-5, mtl-1, supE44, mcrB$_B$

XL1-Blue (Stratagene)

recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, [F´proAB, $lacI^qZ\Delta M15$, Tn 10 (tet^r)]

XL1-Blue MRF´ (Stratagene)

Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi1, recA1,gyrA96, relA1, lac, [F´proAB, $lacI^qZ\Delta M15$], Tn10 (tet^r)]

INVαF´ (Invitrogen)

F´ endA1, recA1, hsdR17, (r_K^-, m_K^+), supE44, thi-1, gyrA96, relA11, φ80lacZΔM15, Δ(lacZYA-argF)U169 deoRλ$^-$

These strains are commercially available from various suppliers for molecular and cellular biology products, such as STRATAGENE EUROPE, Gebouw California, Hogehilweg 15, 1101 CB Amsterdam Zuidoost, The Netherlands, Life Technologies A/S , 4000 Roskilde Denmark, Invitrogen, 9704 CH Groningen, The Netherlands, Roche Diagnostics, Mannheim Germany or from the American Type Culture Collection (ATCC), 12301, Rockville Maryland.

These bacterial strains can be used interchangeably, except otherwise noted. The genotypes listed show many of the desired characteristics for choosing a bacterial host and are not meant to limit the invention in any way. The genotype designations are in accordance with standard nomenclature (see literature mentioned above). A preferred strain of E.coli used in the cloning and expression of the genes of this invention is TOP10F´, which is available from Invitrogen and which is generally employed in the procedure of PCR cloning with the TOPO TA Cloning® Kit, provided from the same company.

The literature mentioned below is a selection from the rapidly increasing variety available on this subject:

1. Jendrisak, J. et al., In: Guide to Molecular Cloning Techniques, Academic Press, 359-371, 1987.
2. Maniatis, T., et al., In: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, Cold Spring Harbor

3. Palmer, B.R., Marinus, M.G., The dam and dcm strains of Escherichia coli, a review, Gene, 143, 7-8, 1994.

4. Dila, D., Sutherland, E., Moran, L., Slatko, B., Raleigh, E.A., J. Bacteriol, 172, 4888-4900, 1990.

5. Current Protocols in Molecular Biology, published by John Wiley & Sons, Edited by: Fred M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, Kevin Struhl, Series Editor: Virginia Benson Chanda; 0-471-50338-X - Looseleaf 0-471-30661-4 - CD-ROM.

Example 5

The TOPO™ TA Cloning® Kit

The TOPO™ TA Cloning® Kit is the first and only PCR cloning kit specifically designed for cloning Taq-amplified PCR products with about 95% recombinants in only 5 minutes. With TOPO™ TA Cloning, there is no post-PCR modification, no special primers and no extra steps. You simply add 1 µl of your PCR product to the activated pCR®2.1-TOPO vector, incubate for 5 minutes on your bench top and transform. TOPO™ TA Cloning is made possible by a unique enzyme, topoisomerase I. When combined with the pCR®2.1-TOPO vector, the topisomerase I activates the DNA, making it ready for rapid ligation with a PCR product with compatible overhangs.

Convenient features of the pCR®2.1-TOPO vector include:

1. 3´ T-overhangs for direct ligation of Taq-amplified PCR products
2. EcoR I sites flanking PCR product insertion site for easy excision of insert
3. Kanamycin and ampicillin resistance genes for your choice of selection in E. coli.
4. Easy blue/white colony screening for identification of recombinants

(See: Shuman, S., Journal of Biological Chemistry, 269: 32678-32684, 1994).

Example 6

Alignments of PSK polypeptides

The polypeptide alignments were performed using the alignment algorithm described by Thompson J.D., Higgins D.G., Gibson T.J.;"CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice."; Nucleic Acids Res. 22:4673-4680(1994) which is

available at the ClustalW WWW Service at the European Bioinformatics Institute (http://www2.ebi.ac.uk/clustalw). The parameters set for the alignments were:

Alignment:	-align
OutputFormat:	aln
OutOrder:	aligned
ktup:	1
Window:	0
Score:	percent
PairGap:	0.05
Matrix:	blosum
GapOpen:	10
EndGaps:	10
GapExt:	0.05
GapDist:	0.05

Example 7

Structure predictions suggest: The PSK proteins consist of different domains and can be transmembrane as well as secreted

The *psk*-cDNAs were translated into their proper reading frame, which was originally given by the peptide sequence tags as a result of the nanoelectrospray analysis. Amino acid sequence motifs can indicate particular structural or functional elements. The use of generally available bioinformatic tools from the bioinformatic group at the EMBL in Heidelberg, for example the SMART program (Schultz et al., 1998) or from the EMBL Hinxton Outstation as well as from the National Center for Biotechnology Information (NCBI) and other bioinformatic sites revealed several functional domains in the proteins of the present invention (for review on bioinformatic tools and sites on the internet see M. Patterson and M. Handel (eds.) Trends Guide To Bioinformatics, Trends supplement 1998, Elsevier Trends Journals). A summary of the amino acid start and end points of the different domains listed below is given in Table 2 and a graphical view of these domains is illustrated in Figure 5.

Table 2

PSK Proteins	Domains	Amino Acid Start Point	Amino Acid End Point
PSK-1 ,2 and 3	Signal Sequence	1	27
PSK-1 ,2 and 3	Threonine Rich Domain	83	98
PSK-1 ,2 and 3	CUB-1	103	216
PSK-1 ,2 and 3	CUB-2	279	389
PSK-1 ,2 and 3	CUB-3	457	568
PSK-1 ,2 and 3	CCP-1	220	275
PSK-1 ,2 and 3	CCP-2	394	453
PSK-1 and 2. PSK-2 includes only the first 7 amino acids.	CCP-3	574	639
PSK-1 and 3	CCP-4	635	694
PSK-1 and 3	CCP-5	701	758
PSK-1	Transmembrane Domain	788	808
PSK-3	Transmembrane Domain	775	795
PSK-1	Cytoplasmic Domain	809	853
PSK-3	Cytoplasmic Domain	796	840

Table 2. List of the different protein domains within the PSK-Proteins according to the SMART output (Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P., 1998).

The amino acid start and end points of the different domains are listed. Note that in the PSK proteins only the N-terminal halves of the defined CUB domains are present containing only two instead of four conserved cysteine residues. The transmembrane domain of PSK-1 and PSK-3 consists of the same amino acids. In PSK-3 13 amino acids are deleted in front of the transmembrane domain.

SEQ ID 2 consists of several domains:
One domain of PSK-1 is a signal peptide. That domain starts at position 1 of the query sequence and ends at position 27. (GTPRAQHPPPPQLLFLILLSCPWIQG). According to SMART and to the " SignalP V1.1 World Wide Web Prediction Server Center for Biological Sequence Analysis" the signal peptide gets cleaved between pos. 27 and 28: IQG-LP. (See: Henrik Nielsen, Jacob Engelbrecht, Søren Brunak and Gunnar von Heijne: Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering, 10, 1-6 (1997)).

Another domain of PSK-1 is a threonine rich domain starting at amino acid position 83 of SEQ ID NO:2. 12 out of 17 amino acids are threonine.

PSK-1 and 3 contain five complement control protein (CCP) modules (also known as short consensus repeats or short complement-like repeat (SCRs) or SUSHI repeats). The CCP modules contain approximately 60 amino acid residues, are beta-sheet type proteins and have been identified in several proteins of the complement system (A missense mutation in seventh CCP domain causes deficiency of the b sub-unit of factor XIII). Four invariant cysteines are characteristic of a SCR. According to the SMART output the CCP modules also include one tryptophan. Short consensus repeats (SCRs) are found in members of the selectin superfamily and in twelve of the complement proteins. These SCRs are very similar to the hypervariable region of antibodies. These domains are conserved over several species and are found in secreted, transmembrane, membrane associated, cytoplasmic and nuclear proteins. (Ripoche et al., 1988; Norman et al., 1991; Hashiguchi et al., 1995).

Four invariant cysteines are characteristic of a SCR. Short consensus repeats (SCRs) are found in members of the selectin superfamily and in twelve of the complement proteins. The selectins belong to a family of cell surface glycoproteins. Like the PSK receptor molecules, they consist of a transmembrane region and a short cytoplasmic anchor. These

protein modules seem to make up functional molecules in a patchwork arrangement. *In vivo* and *in vitro* studies of selectins have established their importance in a wide variety of human diseases. The selectins have stimulated tremendous interest because of their involvement in a wide array of interactions between leukocytes and endothelial cells. Selectins mediate the initial phase of immunocyte extravasation into inflamed sites. The identification of their carbohydrate ligands provides an opportunity to develop a novel class of anti-inflammatory drugs.

In the complement system, the role of SCRs could be in recognition of C3b or C4b as all complement proteins that interact with these contain SCRs. SCRs are also found in the Interleukin 2 receptor and blood clotting Factor XIII.

These SCRs are very similar to the hypervariable region of antibodies and are an example of divergent evolution of a primordial gene. These domains are conserved over several species and are found in secreted, transmembrane, membrane associated, cytoplasmic and nuclear proteins. (See: J. Ripoche, et al., The complete amino acid sequence of human complement factor H, Biochem. J., 249, 593-602, (1988); D. G. Norman et al., Three-dimensional structure of a complement control protein module in solution, J Mol Biol 219, 717-725, (1991); T. Hashiguchi, et al. Molecular and cellular basis of deficiency of the b subunit for factor XIII secondary to a Cys430-Phe mutation in the seventh Sushi domain, J. Clin. Invest., 95, 1002-1008, 1995.)

According to the SMART output, the CCP modules, including the four cysteines and one tryptophan, start and end at the following amino acid positions:

Module	Start	End
CCP-1	220	275
CCP-2	394	453
CCP-3	574	639
CCP-4	635	694
CCP-5	701	758

PSK-1 and 3 further contain three N-terminal parts of complement C1r/s-like repeats or CUB domains. The domain was first found in **C**1r, C1s, **u**EGF, and **b**one morphogenetic protein and is named after these three proteins. This domain is found mostly among

developmentally regulated proteins such as the dorso-ventral patterning protein tolloid, bone morphogenetic protein 1, a family of spermadhesins which contain only this domain, complement subcomponents Cls/Clr and the neuronal recognition molecule A5 (Bork and Beckmann, 1993). The CUB domain consists of approximately 110 amino acids, which suggests an antiparallel beta-barrel similar to those in immunoglobulins.

The CUB domain is also conserved within several species and is found in secreted, transmembrane, membrane associated, cytoplasmic and nuclear proteins (See: P. Bork, et al., The CUB domain. A widespread module in developmentally regulated proteins, J. Mol. Biol., 231: 539-545, 1993)

According to the SMART output, the CUB domains start and end at the following amino acid positions:

Domain	Start	End
CUB-1	103	216
CUB-2	279	389
CUB-3	457	568

Note that in the PSK proteins only the N-terminal halves of the defined CUB domains are present containing only two instead of four conserved cysteine residues.

Yet another domain of PSK-1 is a transmembrane domain (transmembrane helix) located between amino acid position 788 to 808 of SEQ ID NO:2.

Furthermore, the short cytoplasmic tail contains a potential tyrosine phosphorylation site: Asn-Pro-Leu-Tyr or NPLY, which is at amino acid position 838-841 in PSK-1 and and 825-828 in PSK-3. This particular tyrosine phosphorylation site is also known as a binding motif for proteins with a phosphotyrosine binding (PTB) domain and a Src homology 2 (SH2) domain. For example, Michael and colleagues (1996) describe the importance of the sequence N-terminal of Tyrosine 960 in the juxtamembrane region of the insulin receptor for the interaction with insulin receptor substrate (IRS-1). This N-terminal sequence with the common consensus NPXpY appears in sites recognized by both IRS-1 and Shc PTB domains (phosphotyrosine-binding domain of the Shc protein). The adaptor protein Shc contains a phosphotyrosine binding domain (PTB) and a Src homology 2 domain (SH2), both of which are known to interact with phosphotyrosine residues. IRS-1

itself serves as a link between the activated insulin receptor and a variety of effector proteins that bear Src homology-2 (SH2) domains. Therefore, the PSK-1 and PSK-3 could be involved in the Ras signaling pathway.

The N-terminal part of the PSK-1 and 3 proteins from amino acid position 28 to the transmembrane domain (760 amino acids in PSK-1 and 747 amino acids in PSK-3) is most likely located outside the cell (extracellular domain), followed by the transmembrane part from amino acid (21 amino acids), followed by a short cytoplasmic tail of 45 amino acids.

Yet another domain of PSK-1 is the nine potential N-glycosylation sites (identified using the program ProfileScan from the ISREC-Server (http://www.ch.embnet.org/software/PFSCAN_form.html) at the following amino acid positions:

106-109, 152-155, 177-180, 262-265, 285-288, 303-306, 403-306, 447-450, 572-574.

An additional domain of PSK-1 is the nine potential Protein kinase C (PKC) and 12 Casein kinase II (CK2) phosphorylation sites as well as another Tyrosine phosphorylation site, but since these are all in the potential extracellular domain, their phosphorylation is rather unlikely. The PKC sites are at amino acid positions: 3-5,191-193, 200-202, 305-307, 383-385, 519-521, 684-686, 688-690, and 724-726. The CK2 sites are at amino acid positions: 108-111, 154-157, 232-235, 326-329, 350-353, 405-408, 511-514, 538-541, 573-576, 588-591, 634-637, 647-650. The Tyrosine site as mentioned under point 8 above is at amino acid position 61-68. It has a different consensus than the above-mentioned Tyrosine phosphorylation site. The consensus pattern is according to PROSITE: PDOC00007 documentation: [RK]-x(2)-[DE]-x(3)-Y or [RK]-x(3)-[DE]-x(2)-Y, Y is the phosphorylation site.

Substrates of tyrosine protein kinases are generally characterized by a lysine or an arginine seven residues to the N-terminal side of the phosphorylated tyrosine. An acidic residue (Asp or Glu) is often found at either three or four residues to the N-terminal side of the tyrosine (Patschinsky T., Hunter T., Esch F.S., Cooper J. A., Sefton B.M., Proc. Natl. Acad. Sci. U.S.A. 79:973-977(1982), Hunter T., J. Biol. Chem. 257:4843-4848(1982), Cooper J.A., Esch F.S., Taylor S. S., Hunter T., J. Biol. Chem. 259:7835-7841(1984)).

There are a number of exceptions to this rule such as the tyrosine phosphorylation sites of enolase and lipocortin II.

PSK-1 also contains 14 potential N-myristoylation sites at amino acid positions (2-7, 120-125, 128-133, 165-170, 239-244, 263-268, 281-286, 402-407, 414-419, 458-463, 556-561, 595-600, 705-710, 785-790) as well as a potential amidation site at amino acid position 314-317.

Furthermore, PSK-1 contains a prokaryotic membrane lipoprotein lipid attachment site located at amino acid position 12-22, within the above mentioned signal sequence. For reference see PROSITE: PS00013.

SEQ ID NO:4

The protein with SEQ ID NO:4 stops at amino acid 580. Its sequence is identical to the first 580 amino acids in SEQ ID NO:2. It is therefore a truncated version of SEQ ID NO:2.

SEQ ID NO:4 contains the signal peptide, the three CUB domains and the first two out of five CCP modules of SEQ ID NO:2. It has no C-terminal transmembrane domain.

The signal peptide suggests that SEQ ID NO:4 is a secreted protein, which could bind to SEQ ID NO:2 in an autocrine manner and activate this receptor molecule for cell signalling.

SEQ ID NO:6

SEQ ID NO:6 is identical to SEQ ID NO:2 except that it does not contain 13 amino acids between position 762 and 774 (these two positions included). The deletion is in-between the last CCP domain which ends at amino acid position 758 and the transmembrane domain which starts at amino acid position 788.

This could be important for potential cleavage of the extracellular part of the receptor molecule from its membrane-spanning domain. Potential cleavage or recognition sites for proteases could be inserted or deleted with these amino acid changes which could allow the protein to be processed differently to SEQ ID NO:2.

All the above mentioned domains at SEQ ID NO:2 are contained within SEQ ID NO:6.

Since the PSK molecules consist of complement control protein modules (CCPs also called short consensus repeat (SCR) and CUB domains (named after **C**1r, C1s, **u**EGF, and **b**one morphogenetic protein) we can focus also on proteins which are known to interact with these domains. It is presently contemplated that the PSK proteins are involved in inflammation processes and/or neurodegenerative diseases and therefore directly and/or indirectly provoke epileptic seizures.

Background on proteins interacting with SCR domains:

SCR domains are reported in interactions of complement proteins with C3b or C4b and of selectins on leukocytes for example with endothelial cells. They also interact with viruses and bacterial pathogens.

Examples:

1. CR1 (C3b receptor, CD35) is a polymorphic transmembrane single-chain polypeptide (Mr 190 000-280 000) containing variable numbers of short consensus repeats in the extracellular region. It is expressed on erythrocytes, where it is responsible for the clearance of circulating immune complexes and other blood cells. It also stimulates the dissociation of classical and alternative pathway C3 convertases. Its deficiency is associated with the disease SYSTEMIC LUPUS ERYTHEMATOSUS.

2. CD55, for example, is a short consensus repeat (SCR) domain containing widely expressed polymorphic glycoprotein which functions as a complement regulator by inhibiting assembly and promoting destruction of C3 and C5 convertases. As a key regulator of complement, CD55 is implicated in the hyperacute rejection of xenografts from pigs into primates. It is also commonly hijacked as a receptor by viruses (eg medically important echoviruses and coxsackieviruses) and bacterial pathogens (eg certain pathogenic strains of Escherichia coli). Here, crystallization of a virus-binding fragment expressed in yeast, consisting of two of the four extracellular SCR domains of CD55, is reported.

3. Enterovirus 70 (EV70), like several other human enteroviruses, can utilize decay-accelerating factor (DAF [CD55]) as an attachment protein. Using chimeric molecules composed of different combinations of the short consensus repeat domains (SCRs) of DAF and membrane cofactor protein (CD46) we show that

sequences in SCR1 of DAF are essential for EV70 binding. Of the human enteroviruses that can bind to DAF, only EV70 and coxsackievirus A21 require sequences in SCR1 for this interaction. (Karnauchow TM, et al. Short consensus repeat domain 1 of decay-accelerating factor is required for enterovirus 70 binding. J Virol. 1998 Nov;72(11):9380-3. PMID: 9765493; UI: 98440612.)

4. Factor H, a secretory glycoprotein composed of 20 short consensus repeat modules, is an inhibitor of the complement system. Previous studies of inherited factor H deficiency revealed single amino acid substitutions at conserved cysteine residues on one allele arginine for cysteine 518 (C518R) and on the other tyrosine for cysteine 941 (C941Y) (for reference see: Ault, B. H., Schmidt, B. Z., Fowler, N. L., Kashtan, C. E., Ahmed, A. E., Vogt, B. A., and Colten, H. R. (1997) J. Biol. Chem. 272, 25168-25175). Mutant factor H was retained in the endoplasmic reticulum and degraded relatively slowly as compared with most other mutant secretory and membrane proteins that are retained in the endoplasmic reticulum. This impaired secretion of the naturally occurring C518R and C941Y mutant factor H proteins is due to disruption of framework-specific disulfide bonds in factor H short consensus repeat modules. (Schmidt BZ, et al., Disruption of disulfide bonds is responsible for impaired secretion in human complement factor H deficiency. J Biol Chem. 1999 Apr 23;274(17):11782-8. PMID: 10206995; UI: 99223500.)

Background on proteins interacting with CUB domains:

Neuropilins bind secreted members of the semaphorin family of proteins. Neuropilin-1 is a receptor for Sema III. Here, we show that neuropilin-2 is a receptor for the secreted semaphorin Sema IV and acts selectively to mediate repulsive guidance events in discrete populations of neurons. neuropilin-2 and semaIV are expressed in strikingly complementary patterns during neurodevelopment. The extracellular complement-binding (CUB) and coagulation factor domains of neuropilin-2 confer specificity to the Sema IV repulsive response, and these domains of neuropilin-1 are necessary and sufficient for binding of the Sema III semaphorin (sema) domain. These results provide insight into how interactions between neuropilins and secreted semaphorins function to coordinate repulsive axon guidance during neurodevelopment. (Giger RJ, et al., Neuropilin-2 is a receptor for semaphorin IV: insight into the structural basis of receptor function and specificity. Neuron. 1998 Nov;21(5):1079-92. PMID: 9856463; UI: 99072283.

Deletion mutants demonstrated that the N-terminal CUB domain and the EGF-like domain of C1r together are responsible for the calcium dependent C1r-C1s interaction. It seems very likely that these two modules form the calcium-binding site of the C1r alpha-fragment and participate in the tetramer formation (Gal P, et al., Structure and function of the serine-protease subcomponents of C1: protein engineering studies. Immunobiology. 1998 Aug;199(2):317-26. Review. PMID: 9777415; UI: 98450590).

PSK-3 is identical to PSK-1 except that it does not contain 13 amino acids between position 762 and 774 inclusive (AYEELLDNRKLEV). The deletion lays between the last CCP domain which ends at amino acid position 758 and the transmembrane domain which starts at amino acid position 788 in the PSK-1 protein. This could be important for potential cleavage of the extracellular part of the receptor molecule from its membrane-spanning domain, which could allow the two proteins to be processed differently. Several protease cleavage sites are removed after deletion of these 13 amino acids. Proteases like V8, trypsin, pancreatic elastase, endopeptidase, clostripain, chymotrypsin, armillaria mellea, they all cleave within these 13 amino acids. At the deletion point two amino acids are joined together in PSK-3, V and T, which lead to a cleavage site for pancreatic elastase. Potential cleavage or recognition sites for other yet unknown proteases, which are involved in diseases and only occur at a certain disease state could also be inserted or deleted within that region. PSK-1 and PSK-3 are classified by their domain structure with an N-terminal extracellular domain, a single transmembrane domain and the intracellular cytoplasmic C-terminus as a TYPE 1 transmembrane protein.

Other sites identified in the PSK-1 protein with the program ProfileScan (from the ISREC-Server) are:
Nine potential N-glycosylation sites at amino acid positions: 106-109, 152-155, 177-180, 262-265, 285-288, 303-306, 403-306, 447-450, 572-574. Nine potential Protein kinase C (PKC) and 12 Casein kinase II (CK2) phosphorylation sites as well as another Tyrosine phosphorylation site, but since these are all in the potential extracellular domain, their phosphorylation is rather unlikely. The PKC sites are at amino acid positions: 3-5,191-193, 200-202, 305-307, 383-385, 519-521, 684-686, 688-690, and 724-726. The CK2 sites are at amino acid positions: 108-111, 154-157, 232-235, 326-329, 350-353, 405-408, 511-514, 538-541, 573-576, 588-591, 634-637, 647-650. The Tyrosine site is at amino acid position 61-68. It has a different consensus than the above-mentioned Tyrosine phosphorylation site. The consensus pattern is according to PROSITE: PDOC00007

documentation: [RK]-x(2)-[DE]-x(3)-Y or [RK]-x(3)-[DE]-x(2)-Y, Y is the phosphorylation site. Substrates of tyrosine protein kinases are generally characterized by a lysine or an arginine seven residues to the N-terminal side of the phosphorylated tyrosine. An acidic residue (Asp or Glu) is often found at either three or four residues to the N-terminal side of the tyrosine (Patschinsky et al., 1982; Hunter 1982, Cooper et al.,1984). There are a number of exceptions to this rule such as the tyrosine phosphorylation sites of enolase and lipocortin II. PSK-1 also contains 14 potential N-myristoylation sites at amino acid positions (2-7, 120-125, 128-133, 165-170, 239-244, 263-268, 281-286, 402-407, 414-419, 458-463, 556-561, 595-600, 705-710, 785-790) as well as a potential amidation site at amino acid position 314-317. Furthermore, PSK-1 contains a prokaryotic membrane lipoprotein lipid attachment site located at amino acid position 12-22, within the above mentioned signal sequence. The membrane anchor domain is conserved in bacteria. For reference see PROSITE: PS00013.

PSK-2 stops at amino acid 580. Its sequence is identical to the first 580 amino acids in PSK-1 and 2. It is therefore a truncated version of these proteins. PSK-2 contains the signal peptide, the Threonine rich domain, the three CUB domains and the first two out of five CCP modules of PSK-1 and 2 as well as the first 7 amino acids of CCP-3. It has no C-terminal transmembrane domain (see Table 2). The signal peptide suggests that PSK-2 is a secreted protein, which could bind to PSK-1 or 3 in an autocrine manner and activate these receptor molecules for cell signaling.

Example 7

Northern blotting

For human gene expression studies, human multiple tissue Northern blots (Clontech, 7760-1, 7755-1,7750-1) were probed with the human EST-sequence (R90750) and the human RACE-PCR product of the 5prime end (base 1 to 449). The probes were obtained as gel-purified restriction fragments: for the EST, a 2.1kb NotI-EcoRI fragment and for the RACE-product an EcoRI fragment with the very 5prime 449bp was used.

As a control the actin probe (2 kb human β-actin cDNA) provided with the Clontech blots was used. It should generate a signal at approximately 2 kb. Specialized tissues such as pancreas and testes will produce lower β-actin signal intensities. In both, heart and

skeletal muscle there are two forms of β-actin mRNA, a 2 kb and a 1.6-1.8 kb form. The probe hybridizes to the α and γ form of actin.

The probes were labeled and hybridization was performed as described in the ECL Random-Prime Labeling and Detection System (RPN3030 from Amersham/Pharmacia). Blots were washed for 20 min at 65C in 1xSSC/0.1%SDS and for 20 min in 0.5xSSC/0.1%SDS and rinsed for 1min in 0.1xSSC/0.1%SDS. Filters were stripped inbetween subsequent hybridizations by boiling 3-4 times for 15min each in 0.1%SDS/H_2O.

Tissue distribution of *psk* mRNA:

Different multiple tissue Northern blots (MTN) from Clontech were used to study the expression pattern of the *psk* mRNAs in the different human tissues. In human adult tissue, Northern analysis shows the highest levels of expression of *psk* mRNAs in the adult brain followed with slightly decreased expression in the pancreas (see Figure 17, top panel, A). Other tissues such as heart, placenta, lung and kidney show only very small amounts of *psk* mRNAs. The *psk* mRNA expression in liver and skeletal muscle lies here under the detection limit. An actin probe provided from Clontech as a positive control together with the multiple tissue Northern blots was used on the same blot to demonstrate that equal amounts of mRNA were loaded on each lane (see Figure 17, top panel, B). As described in the Clontech manual for MTN blots, the β-actin control probe generates a singal with the approximate length of 2 kb. Specialized tissues like pancreas and testes will produce lower β-actin intensities. In heart and skeletal muscle there are two forms of β-actin mRNA, a 2-kb and a 1.6-1.8-kb form. The difference in size is due to probe hybridization to either the a or g form of actin. Depending on the stringency other isoforms of actin my also be detected in some tissues. This holds also true for the *psk* mRNAs, different forms or other members of this gene family can be seen on the MTN blots.

Figure 7 (panel, A) shows mRNA expression of the *psk* genes in prostate, testes, ovary, small intestine and colon. The *psk* mRNA expression in spleen, thymus and peripheral blood leukozytes lies here under the detection limit. Actin is again used as a control on the same blot (see Figure 7, panel, B).

Figure 6 A shows again the MTN blot shown in Figure 17A, but this time in comparison to a second MTN blot from Clontech with the mRNA from the same tissues (Figure 6 B). In

A, the 3 prime *psk-2* DNA fragment (bases 827-2855) was used in the hybridization, in B a 5 prime *psk-2* DNA fragment (from bases 1-449) was used for hybridization. The 5 prime and 3 prime probe show a similar pattern with highest mRNA expression in the brain and the pancreas.

Further analyses (Figure 8 A and B) show that the *psk* mRNAs are expressed in several areas of the brain which are involved in epileptic seizures or other neurodegenerative deseases: the occipital lobe, the frontal and the temporal lobe. *psk* mRNAs are also highly expressed in the cerebellum as well as in the hippocampus, thalamus, the amygdala and the caudate nucleus.

Further areas of *psk* mRNAs expression are the corpus callosum, the substanca nigra, the subthalamic nucleus, the cerebral cortex, the medulla and the putamen. In the spinal cord the messages of the *psk* genes are on this Northern blot under the detection limit. The probe which was used in the hybridization included parts of the extracellular domain, the transmembrane region, the intracellular domain and the 3 prime untranslated region. It included the nucleic acids from base 827 to base 2855 of the above mentioned sequence SEQ ID NO: 3 and is 2029 bases long. The Northern blots used were multiple tissue Northern blots provided from Clontech containing 2μg of poly(A)+ RNA from each tissue mentioned.

Northern Hybridization was done according to standard molecular biology techniques, in this case acoording to the "Amersham Fluorescein labelling and detection kit" on Multiple Human Tissue Northern Blots from Clontech (CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303, USA).

To detect the mRNA expression in the adult mouse brain, antisense oligos of 45 bases were end-labeled with Terminal Deoxynucleotidyl Transferase (TdT) and hybridized on equal amounts of mouse brain mRNA (Figure 18). The oliogonucleotides chosen have the following sequences:

Oligo NO. 4 antisense:
5'-GCCGTAGCCAGGGTAGACATGGATGCTGTAAGTGCAGTCCAGGAG-3',
this is the antisense sequence to the bases: 469-513. The corresponding sense sequence (reverse complement) was used as a negative control in the hybridization procedures.

Oligo NO. 2 antisense:
5'-GCCGATAAGCTCAAAGCCCTCATAGCAGAAGAAGCGCAGAGACTC-3', this is the antisense sequence to the bases: 2251-2295. The corresponding sense sequence (reverse complement) was used as a negative control in the hybridization procedures.

Oligo NO. 1 antisense:
5'-GGTCACTTCCAGTTTTCGGTTGTCCAGGAGCTCCTCATAGGCAAC-3', this is the antisense sequence to the bases: 2362-2406. The corresponding sense sequence (reverse complement) was used as a negative control in the hybridization procedures.

Oligo NO. 3 antisense:
5'- GAGGCAGAGACCGGGTCCCGTATTTCCCTCTGCCCGAATGAGGAG-3', this is the antisense sequence to the bases: 2687-2731. The corresponding sense (reverse complement) sequence was used as a negative control in the hybridization procedures.

Oligo NOs. 1 and 2 are from the 3' end of the coding sequence of the *psk-1* and *psk-2* genes and oligo NO. 1 spans over the deletion of the *psk-3* gene. The deleted bases are 2364- 2402.

Oligo NO. 3 is from the 3' untranslated region of the PSK genes.

Oligo NO. 4 is from the 5' end of the *psk* genes, from part of the N-terminal region of SEQ ID NO: 1, which is except for five exchanged bases identical to the mouse sequence in this region. The mRNA from mouse adult brain tissue was prepared with standard procedures known to the skilled person in the field.

In all Northern hybridization experiments there were at least two mRNA species detected on the Northern blot, and the resulting mRNA bands were between the 2.4 and 4.4 kb in size.

<u>The psk genes are localized on Chromosome 16</u>

Using the increasing amount of databases on all different aspects of the human genome, the UniGene-Homo sapiens database was searched with the EST accession numbers R 90750 and H10539 (see above) for the chromosomal localization of that DNA. According

to the mapping information, the Whitehead map, the psk genes are localized on the human chromosome 16.

Tissue distribution of *psk* mRNA expression

Northern blot analysis showed, that the *psk* mRNA is predominantly expressed in brain and pancreas tissue and only very minor amounts are expressed in kidney, placenta and lung, these results are identical using the 5' RACE-cDNA of the psk molecules or the EST with the accession number: R90750 (Figure 6 A and B). Strong hybridization signals are also detected in prostate and testis. Lower amounts of psk mRNA were found in ovary, small intestine and colon (Figure 7). Since the strongest expression level of the psk-molecules was found in the brain, different areas of the human brain were closer analyzed (Figure 8 A and B). Highest expression levels were present in cortical structures such as amygdala, caudate nucleus and hippocampus and in the thalamus, which is part of the diencephalon. Furthermore expression was detected in the substanca nigra, the subthalamic nucleus and the corpus callosum. Figure 8 shows that the cerebellum also consists of a considerable amount of psk mRNAs as well as the cortex with its different lobes, the occipital lobe, the frontal lobe and the temporal lobe. Medulla and putamen show only low amounts of expression. Spinal cord shows no hybridization signal. Compared to the *sez*-6 molecule the psk molecules have also their highest expression level in the brain. Sez-6 was reported to be upregulated after chemically induced seizures by application of PTZ-17. The areas of psk-expression found in the present study correlate with brain regions, which are known to be involved in epileptic seizures, like the hippocampus and the different cortical lobes (for example Chan S et al., 1996).

Example 8

In situ hybridization studies

In situ hybridization studies were essentially performed as described in Wisden et al. (1991) and Laurie and Schrotz (1994).

A short summary of the procedure will be given below:
45 bp oligonucleotides were synthesized to the PSK c-DNAs corresponding mouse sequence, which was obtained via RACE_PCR with human primers on embryonic mouse cDNA (see above). The resulting mouse sequence was 553 bases (homologous to bp

292-844 of the human sequence). A part of this cDNA corresponding to base 303-347 was chosen for the 45 bases long in situ sense (5'-CCTCAGGCCTGAGGGAGGAGAGGAGGAAACCACCACCACCATCAT-3') and antisense primer (5'-ATGATGGTGGTGGTGGTTTCCTCCTCTCCTCCCTCAGGCCTGAGG-3').

The oligonucleotides were 3'-end labeled for 10 min at 37 C with terminal deoxynucleotidyl transferase (Roche) using a 30:1 molar ratio of 5'-(α-35S)dATP (NEN, 1460 Ci/mmol, <1mM DTT) to oligonucleotide. Unincorporated nucleotides were removed by purification with Bio-Spin-6 chromatography columns (Bio Rad).

The labeled oligonucleotide was applied onto 16μm frozen tissue sections of adult mouse brain and E16 mouse embryo (NMRI mice). The sections had been fixed in 4% Paraformaldehyde and stored in absolute Ethanol at 4C until use. For hybridization the sections were remove from Ethanol and air-dried. The labeled oligonucleotide was dissolved in hybridization buffer with a final concentration of 0.06 fmol or 1pg/μl or 1500dpm/ μl and applied to the sections.

Hybridization was carried out in a humid chamber at 42°C. Posthybridization washes were carried out in 1 x SSC for 20 min at 55 C and in 1xSSC and 0.1 x SSC for 1min each at room temperature. After dehydration in ethanol, slides were air dried and exposed on Kodak autoradiography X-OMAT S film.

Signal specificity was confirmed by using the sense oligonucleotide, which gave no signal on the autoradiogramm. As a positive control we used a gene called brain factor-1 or Bf-1 (Tao and Lai, 1992), which is only expressed in the telencephalon. The oligonucleotides were taken from the rat sequence corresponding to bp 1291-1336 in sense (5`-CACCTCCACCGGCCTCACCTTCATGGACCGCGCCGGCTCCCTCTA) and antisense (5`-TAGAGGGAGCCGGCGCGGTCCATGAAGGTGAGGCCGGTGGAGGTG) direction.

In situ hybridization studies in mouse show a similar distribution of the mouse psk mRNA in the brain as described above in human tissue using northern hybridization. Figure 9A shows high levels of expression in the hippocampus, the cortex (especially in the outer

layers of the cortex) and in the cerebellum. Looking at the distribution of mRNA during development, psk mRNA seems to be restricted to the brain and the developing eye at embryonic stage 16 in the mouse (Figure 9B).

Example 9

Functional Assays with the PSK family:

Construction of fusion proteins and preparation of antisera

Five different regions of the PSK genes were cloned into the prokaryotic expression vector pGEX-4T-1 (Pharmacia) using the *tac* promoter which is induced using the lactose analog isopropyl β-D-thiogalactoside (IPTG). This system is a gene fusion system with Glutathione S-Transferase (GST), an integrated system for the expression, purification and detection of fusion proteins produced in E.coli. The following regions were cloned in frame with the GST:

GST:PSK-I: bp 829-2646 of psk-2, which has a stop codon at position1825-1827 (coding for the amino acid sequence 249-580 of the PSK proteins)

GST:PSK-II: bp 1450-2445 of psk-1 with an additional stop codon (TAG) inserted at the end (coding for the amino acid sequence 456-787 PSK-1)

GST:PSK-III: bp 1450-2646 of psk-1 (coding for the amino acid sequence 456-853 of PSK-1)

GST:PSK-IV: bp 1804-2358 of psk-1 also with insertion of an additional stop codon TGA (coding for the amino acid sequence 574-758 of PSK-1) and

GST:PSK-V: bp 2518-2646 of psk-1 (coding for the amino acid sequence 812-853 of PSK-1)

The proper insertion of the inserts was verified by sequencing. The fusion proteins were produced in TOP10F' (Invitrogen). After transformation the cells were grown at 30C to an optical density of 0.8 at 600nm. Then the expression of fusion protein was induced by the addition of 0.01mM isopropyl-β-d-thiogalactopyranoside (IPTG) for 1.5 hr at 30 C. The cells were pelleted by centrifugation at 6000g for 20min and resuspended in PBS (phosphate buffered saline) with protease inhibitors (Complete, protease inhibitors cocktail tablets, ROCHE). The suspended cells were lysed by a 6 times freeze and thaw cycle. Triton X-100 was added to a final concentration of 1% and the lysate was gently mixed at 4C for 30min. The lysate was cleared by centrifugation at 12.000xg for 10min at 4C and the supernatant was incubated with glutathione-conjugated agarose beads

(equilibrated in PBS buffer). The amount of beads was chosen depending on protein yields and culture volume according to the Pharmacia protocol (200µl of bed volume (pure beads without buffer) / 400ml culture). The solution was incubated for 30min at room temperature on a rocking platform. The supernatant was discarded and the beads were washed three times in PBS. The beads were eluted in Laemmli sample buffer and the eluate was run on a preparative 10%SDS-PAGE gel. Only one side of the gel was Coomassie stained and after detection of the fusion protein band the same area was cut out of the unfixed part of the gel and used as an immunogen for making antibodies. Alternatively the GST-fusion proteins were eluted overnight from the glutathione-conjugated agarose beads in 10mM reduced glutathione in 50mM Tris-HCl (pH8). The supernatant with the fusion protein was remove by centrifugation at 500g for 5 min and used as an immunogen in rabbits for making antibodies.

The gel excised and eluted fusion proteins (approximately 10µg/injection) were injected into rabbits to generate polyclonal ascites fluid (Antibodies book!!!) The specificity of the antibodies was tested by immunoblotting against both purified fusion protein and tissue culture cell lysate of 293T cells (human kidney cells expressing the Large T-antigen) transfected with the V5/His tagged psk-3 cDNA (in the pcDNA3.1/V5/His-TOPO vector, Invitrogen). The cell lysates were also tested for the expression of the tagged protein with the anti-V5 antibody (Invitrogen) (see below).

Affinity purification of the anti-PSK antibodies

5ml of the polyclonal Anti-GST:PSK rabbit sera were preincubated with Glutathione Sepharose beads. The beads had been preincubated with GST, which was prepared from bacteria expressing the pGEX-4T-1 vector. This preincubation was done to eliminate antibodies against GST from the serum (GST protein from 400ml bacterial culture bound to 400ml bed volume of beads was used in the preincubation experiment). Then 100mg Cyanogen Bromide-(CNBr-) activated Sepharose (Pharmacia) were washed 3x with 1mM HCl and spun for 3min at 5000 rpm in an Eppendorf centrifuge. The volume of the Sepharose was 250-300µl. 400-600µl of the cleaved or uncleaved PSK-GST fusion protein (0.5mg/ml) were taken and 0.2M $NaHCO_3$ pH 8.3 was added to a final concentration of 0.1M and 5M NaCl to a final concentration of 0.5M. The PSK protein was added to the activated beads and rotated over night at 4C. The sample was spun for 3min at 5000rpm and the beads were added to 0.2M Glycine pH8 for 2hr at room temperature. The beads were washed in 1ml of coupling buffer (0.1M $NaHCO_3$, pH 8.3/ 0.5M NaCl),

then in 1ml of 0.1M Na-acetate, pH 4.5/0.5M NaCl and finally again in 1ml coupling buffer. Then they were washed twice in PBS pH 7.4. The beads were now coupled with the antigen and were incubated with 7ml of serum over night at 4C. Then the whole mixture was loaded onto a column. The serum was allowed to drip out of the column and the beads in the column were washed with 5ml of 20mM Na_2HPO_4, pH7.3/150mM NaCl. Five Eppendorf tubes, each with 60µl Neutralization buffer (0.5M Na-phosphate, pH 7.7) were set up and the antibody was eluted with 250µl elution buffer (0.1M Glycine, pH 2.77), into each tube, in total five times. All fractions of affinity purified anti-PSK antibodies were tested in different dilutions on Western blots of PSK transfected 293T cell lysates.

Immunoblotting

10% SDS-PAGE Mini-gels (Bio Rad) were blotted onto Nitrocellulose in a semi dry blotter (Hoefer Semiphor) in a Tris/Glycine/Methanol buffer (48mM Tris/39 mM Glycine/0.037% SDS/20% Methanol pH 8.3) for 45min-1.5 hr at 0.8mA/cm^2 of gel. The blot was rinsed in water and processed according to the protocols for the specific antibodies used.

Tagging of all the psk cDNAs with the V5/HIS-TAG

The *psk* cDNAs were PCR amplified on their corresponding cDNAs with the primers listed below and cloned directly into the PCR cloning vector for expression in eukaryotic systems: pcDNA3.1/V5/His-TOPO vector (Invitrogen).

For the 5' end of all three psk-V5/HIS cDNAs the same primer was used, starting six bases before the ATG start codon:

5'-GTCGCCATGGGGACTCCCAGGGCCCAGCAC-3'. For the 3' end the primers were used according to the sequences of the individual psk cDNAs but without the stop codons. For psk-1-V5/HIS and psk-3-V5/HIS the same 3' primer was used: 5' GATGGAAACTTCATACTCCCGCGTATC-3' and for psk-2-V5/HIS the 3'-primer 5'-CGGAGGTGGCAGCTCGGGGCACGTGTC-3' was chosen. The constructs were all verified for in frame insertion and proper direction via sequencing with the T7 primer and a specific primer in the tag-sequence corresponding to the bases. The cDNA inserted in the opposite direction into the pcDNA3.1/V5/His-TOPO vector was also obtained for each psk clone and used as one negative control in the transfection experiments.

Transfection and lysis of cells

For transfection of 293T cells (human kidney cells expressing the Large T-antigen), they were grown to 60%-70% confluency. The Calcium phosphate transfection method (Invitrogen) was chosen and 250ng of DNA/cm^2 were added to the tissue culture dish in a buffer consisting of a final concentration of 120mM $CaCl_2$ in Hepes buffered saline (HBS), forming a Calcium phosphate-DNA precipitate. The precipitate was added dropwise to the media to the cells and the cells were incubated for 48 hr at 37C in a humidified CO_2 incubator. To harvest the cells they were washed in PBS and 1ml of cold NP-40 lysis buffer (150mM NaCl, 1% NP-40, 50mM Tris pH8) was added to a 10cmØ dish on ice. The cells were incubated on ice for 30 min while gently rocking or shaking, to achieve a good mixing. The cell lysate was transferred to a 1.5ml tube and centrifuged at 10.000xg for 10min at 4C. The supernatant was removed into a fresh tube and both pellet and supernatant were frozen until further analysis.

Pull downs with the GST tagged proteins:

The PSK genes and certain functional domains of the PSK genes are cloned into the bacterial Glutathione S-Transferase (GST) Gene Fusion System from Pharmacia (vector pGEX-4T-1). The fusion proteins are expressed, purified and bound to a solid support or matrix (glutathione-matrix). As a control only GST will be bound to the matrix. Cell lysates of embryonic and/or adult brain tissue extracts (or from other tissues, like pancreas) or cell culture lysates are added to the fusion protein and to GST only and incubated with it, and proteins which are unspecifically bound are washed away with different salt or pH concentrations. The remaining proteins, which are still bound even after extensive washings, indicating a strong interaction between these proteins, are eluted and run on a SDS-PAGE-gel. One- and/or two-dimensional separation is performed to see the different proteins which are bound to the PSK proteins. The different domains of the PSK genes will pick up different binding partners. These protein bands or spots are excised from the gel and further processed for Mass Spectrometric analysis to find out the identity of these proteins. Mass spectra from these peptide mixtures are obtained on a modified Matrix Assisted Laser Desorption Ionization 'Time-of-Flight' (MALDI TOF) mass spectrometer equipped with delayed ion extraction (REFLEX, Bruker-Franzen, Bremen, Germany). For peptide sequencing by nanoelectrospray tandem mass spectrometry, gel pieces are further extracted. The resulting peptide mixture is concentrated and desalted in a two needle assembly consisting of a 100nl R2 Poros microcolumn (C8 type, reverse-phase packing,

PerSeptive Biosystems) which is then eluted with 2 x 0.5µl of 60% methanol, 5% formic acid into a nanoelectrospray spraying needle as described previously (Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; and Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nanoelectrospray mass spectrometry, Nature, 379(6564):466-469, 1996; Wilm, M.; and Mann, M. Analytical properties of the nanoelectrospray ion source, Anal. Chem., 68, 1-8, 1996). Peptide sequence tags are assigned in peptide fragmentation spectra obtained on an API III triple quadrupole or a quadrupole time-of-flight mass spectrometer (Perkin-Elmer Sciex, Toronto, Canada) (Mann, M.; and Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags, Anal. Chem., 66, 4390-4399, 1994). The program PeptideSearch (Petra Schrotz-King) is used to search with peptide masses and peptide sequence tags against a non-redundant protein sequence database (nrdb) maintained at the European BioInformatics Institute (EBI, Hinxton Park, England) and against the database of expressed sequence tags (dbEST) maintained at the National Center for Biotechnology Information (NCBI).

The candidates identified as potential binding partners will give an idea as to in which kind of signaling process our PSK molecules might be involved. To determine whether the observed interactions between the PSK molecules and the bound molecules are physiologically relevant, the expression pattern and the subcellular localization of the binding partners are analyzed via Western Blot and Immunofluorescense. Another possibility would be to add the interacting molecules together in an *in vitro* assay and analyse for changes in eg. cAMP-levels, Calcium levels, Membrane potential or Enzyme reactions.

Positive controls are cell lines, which are transiently or stably transfected with PSK molecules. Negative control are untransfected cells.

Detection of tagged PSK- proteins with the anti-V5-HRP Antibody

For the anti-V5-HRP antibody from Invitrogen: 2% of the total cell lysate from a confluent 10cm dish transfected with V5/HIS-tagged PSK protein or with the tagged PSK protein in the opposite orientation (negative control) was loaded per lane on the gel. The blot was incubated for 1hr at room temperature in 10ml of blocking buffer (TBS: 20mM Tris-HCl/ 500mM NaCl, pH7.5, with 5%milk powder, w/v). The nitrocellulose was washed twice in 20ml TBST (TBS + 0.05%Tween-20, w/v) for 5min with gentle agitation and transferred for 2hr at room temperature into 10ml blocking buffer containing Anti-V5-HRP-antibody,

diluted 1:5000. The membrane was washed twice for 5min in TBST and once for 5min in TBS. Detection of the bound antibody was performed with the ECL Western Blotting Detection Kit from Amersham/Pharmacia. Alternatively the tagged protein can also be detected with the Anti-V5 Antibody (1:5000) as a first antibody and with an anti mouse-HRP (1:4000) (Amersham/Pharmacia).

Detection of tagged or endogenous PSK- proteins with the Anti-GST:PSK-V Antibody

For the Anti-GST:PSK-V Antibody: Preparative SDS-PAGE mini gels (with a one well comb across the gel, 2D comb) were blotted onto Nitrocellulose as described before. The blots were cut into 0.4mm stripes for incubation with different antibody fractions or dilutions. All following incubations were performed at room temperature, unless otherwise stated. The stripes were blocked for 45min at room temperature in 1xPBS (Phosphate Buffered Saline)/ 0.2% Tween-20/ 5% milk powder and incubated for 1hr in fresh blocking buffer supplied with the anti-GST:PSK-V Antibody (diluted 1:50 up to 1:250). The nitrocellulose was washed 3 x 10 min in PBS/0.2% Tween-20 and further incubated for 1hr with the second antibody, Peroxidase-conjugated goat anti rabbit IgG (Dianova) in a 1:5000 dilution in blocking buffer. Then the membrane was rinsed twice and washed for 4 x 10 min in PBS/0.2% Tween-20. Detection was performed as described in the ECL Western Blotting Detection Kit from Amersham/Pharmacia.

Purification of V5/His tagged PSK-2 protein from the medium of transfected 293T cells

293T cells were transfected with psk-2-V5/His tagged cDNA. Fresh medium was added after 24 hr of transfection and collected 24 hr later. The medium was purified over a Zn-chelate column according to the poly-His protein purification kit (Roche). According to the recommended protocol, three elution buffers with increasing imidazole concentration (10mM, 50mM and 500mM) were used. For each buffer 5 x 1ml fractions were collected and aliquots were analysed via Western blotting with the V5/His antibody.

The predicted transmembrane protein PSK-3 can be overexpressed in 293T cells and be detected on Western blot by antibodies against different domains of the protein.

Psk-3 was C-terminally tagged with the V5/His tag (see methods) and overexpressed in 293T cells. We detected the overexpressed protein on immunoblots of cell lysate with an antibody against the V5 epitope (Invitrogen) and with an antibody against the cytoplasmic domain of PSK-1 and 3 named anti-GST:PSK-V (see methods). We had raised antisera against fusion protein containing GST and different domains of the PSK proteins including various extracellular parts, the cytoplasmic domain and the transmembrane domain with extracellular and cytoplasmic domain together in one sequence (for details see methods). All antibodies were affinity purified and tested on western blots and by immunofluorescense. Figure 10 shows the different domains on the PSK-1 protein, which were chosen as immunogens. Figure 11 A and B illustrates one of the anti-GST:PSK antibodies, GST:PSK-V, in comparison to the anti-V5 antibody on the PSK-3 transfected cell lysates. The anti GST:PSK-V antibody is directed against the cytoplasmic tail of the PSK-1 and 3 protein, the anti-V5 antibody is directed against the C-terminally located V5 tag within the eukaryotic expression vector (pcDNA3.1/V5/His-TOPO vector, Invitrogen), which was used for eukaryotic expression cloning (see methods). Both antibodies detect a band between 122 and 212 kd. The 840 amino acids of PSK-3 would only code for a molecular weight of 90.8kd and without the signal sequence the molecular weight would only be 87.8 kd. The tag is only an additional sequence of 45 amino acids (4.8 kd). We therefore conclude that the transmembrane forms of the PSK proteins have a number of post-translational modifications. This is supported by the fact that the protein that was originally cut out of the silver stained gel for mass spectrometric analysis had a molecular weight between 115 and 203 kd. Furthermore it has nine potential N-glycosylation sites as mentioned above and a number of other potential modifications. The control lanes were loaded with cell lysate of 293T cells overexpressing PSK-3, cloned into the same vector, but in the opposite direction, so that only antisense mRNA but no protein should be made. No protein was detected in these control lanes.

PSK-3 localizes intracellularly in the Golgi and in vesicular structures at the plasma membrane in transfected 293T cells

Immunofluorescense studies were carried out to see if the PSK-1 and 3 molecules, which were isolated as and predicted to be membrane proteins also have the characteristic expression pattern for these kind of proteins. Proteins, which are destined to go to the plasma membrane get translated into the ER and transported via the Golgi to the plasma membrane. Apart from a bright cell surface stain, membrane proteins can accumulate in the Golgi and in transport vesicles between ER and Golgi as well as in vesicles between

Golgi and plasmamembrane. During recycling or degradation processes of plasmamembrane proteins they will also pass endosomes, located at the cell periphery just below the plasmamembrane and or lysosomes. Transfected V5/His tagged PSK-3 was detected with anti-V5/HIS and with anti GST:PSK-V antibody. Both antibodies detect the same cellular distribution of the antigen (Figure 12 A-C and F-H). The cells have to be permeable for the antibodies, since both antibodies bind to the predicted cytoplasmic C-terminal region of the PSK-3-V5/HIS protein. The PSK-3 molecule exhibits a typical fluorescent pattern for plasmamembrane proteins, which are in transit through the cell: a bright Golgi and a vesicular staining along the plasmamembrane in permeabilized 293T cells. In the negative controls the construct with the antisense orientation was transfected as well as no DNA. Cells with the antisense construct and cells without transfected DNA show no fluorescense (Figure 12 D and E). That means that with the anti GST:PSK-V antibody against the cytoplasmic domain of PSK-1and 3 also no endogenous PSK protein can be detected in the 293T (kidney) cells. This result supports the observation made from the Northern blot data (see above), that there is also very little PSK mRNA expressed in kidney. The anti V5 antibody can only detect the transfected, tagged version of the PSK proteins.

PSK-2 protein gets secreted out of the cells

To test weather the potential signal sequence is real, and if the PSK-2 protein gets secreted we transfected 293T cells with the V5/His tagged psk-2 cDNA and collected medium after 24 and 48 hr after transfection. An aliquot of the medium (20 µl out of 5ml) was loaded on a 7% SDS-PAGE gel, blotted and detected with the anti-V5/His antibody. Figure 13 shows the medium after day 1 and day2. As controls we also analysed the medium of untransfected cells (lane 1 and 6), the medium of cells transfected with the cDNA of the transmembrane protein PSK-3 (lane 2 and 7), the psk-3 molecule cloned in opposite direction (lane 3 and 8) and the psk-2 molecule cloned in opposite direction (lane 4 and 8). All these controls were negative. No band could be detected. Lane 5 shows the secreted PSK-2 molecule after 1 day transfection in the medium, lane 10 after 2 days of transfection. The amount of protein detected in the medium after 2 days has dramatically increased. Lane 11 and 12 show the positive controls: the PSK-2 molecule, detected from an aliquot of a highly enriched fraction of poly-His column purified medium of psk-2-V5/His transfected 293T cells (see methods) and the PSK-3 molecule from an aliquot of total cell lysate of psk-3-V5/His transfected 293T cells. The experiment proves that PSK-2 has a functional signal sequence, which allows the molecule to be secreted from the cell. PSK-3

cannot be detected in the medium. The transmembrane domain anchors it in the plasmamembrane and the compartmental membranes within the cell.

Furthermore a purification of the highly enriched PSK-2 protein in the medium after 48hr of transfection was performed, using the poly-His protein purification kit (Roche). The PSK-2 protein from the medium was bound to a Zn-chelate column and eluted with three different buffers with increasing imidazole concentration (see methods). For each buffer 5 x 1ml fractions were collected, which resulted in 15 fractions. Figure 14A. illustrates that most of the PSK-2 protein elutes in fractions 5-10, which correspond to the elution buffer with an imidazole concentration of 50mM. As a control the same experiment was performed for the medium of 293Tcells transfected with psk-3-V5/His cDNA (Figure 14B, lane 1-15). PSK-3 does not get secreted and cannot be enriched from the cell culture medium. As a positive control an aliquot of fraction 9 from Figure 14A. was loaded (lane 16), together with an aliquot of total cell lysate of 293T cells transfected with tagged psk-3. The difference in molecular weight between the secreted and the transmembrane form can clearly be seen.

Example 10

Deglycosylation with Peptide:N-Glycosidase F (PNGase F)

For the analysis of the secreted molecule PSK-2, 293T cells were transfected with psk-2-V5/His cDNA. The medium was collected after 48 hr and purified with the poly-His protein purification kit (see above). We chose an aliquot of 30µl out of a 1ml fraction (fraction 9) of elution buffer 2 (most of the PSK-2 protein eluted in elution buffer 2). For the PSK-3 protein 30µl of total cell lysate from transfected 293T cells were chosen. Each sample was denatured in 0.07M 2-mercaptoethanol /0.3% SDS. Then from each sample 1/6 was taken out as 0min control without PNGase after denaturation. To the rest of each sample 5U of recombinant PNGase F (Roche) were added and the samples were incubated at 30 C in the presence of 0.26M Tris buffer pH 8 and 1.5% CHAPS, a 5 fold excess of nonionic detergent over SDS, to stabilize the enzyme.. An aliquot of the samples was taken out at 5 different timepoints (5min, 20min, 45min, 24hr, 48hr). The enzyme in these aliquots was inactivated by boiling the sample for 5 min. The samples were loaded on a 7% SDS gel, blotted and the PSK proteins were detected with the anti-V5-HRP antibody as described above.

The PSK molecules are highly glycosylated

In order to obtain some more insight as to the difference between calculated and actual molecular weight of the PSK proteins, we treated the PSK-2 and PSK-3 protein with PNGase F. There are multiple N-glycosylation sites predicted for the PSK proteins, that can be digested with PNGase F.

We transfected 293T cells with psk-2 and psk-3, tagged with the V5/HIS module. The medium of the psk-2 transfected cells was harvested after 48 hr and purified over a Zn-chelate column. An aliquot of the fraction with the highest PSK-2 concentration was taken for the deglycosylation experiment. An aliquot of the total cell lysate was taken for PSK-3 deglycosylation. The samples were denatured in 0.07M 2-mercaptoethanol /0.3% SDS and PNGase F treatment was carried out for different timepoints (0min, 5min, 20min, 45min, 4hr, 24hr and 48 hr) at 30 C in the presence of 0.26M Tris buffer pH 8 and 1.5% CHAPS. A 5 fold excess of nonionic detergent over SDS is required to stabilize the enzyme. Then the enzyme was inactivated by boiling the samples for 5 min. The samples were loaded on a 7% SDS-PAGE gel, blotted and the PSK proteins were detected with the anti-V5-HRP antibody. Figure 15 A shows, that the fully glycosylated PSK-2 protein runs just above 121 kD whereby after deglycosylation the PSK-2 protein shifts clearly below a molecular weight of 121kD. The shift occurs stepwise between 5min and 48 hr of incubation, indicating a possibility of 5 glycosylated sites in the PSK-2 protein. PSK-3 protein in its complete glycosylated form runs above 121kD (Figure 15 B). After deglycosylation of 45 min, 4hr and 24hr the protein band clearly shifts closer to 121 kD and migrates at the same hight as the glycosylated form of PSK-2. The deglycosylation on total cell lysate is not as efficient as on the purified protein fraction. The experiment shows that the PSK proteins are highly glycosylated.

Example 11

Immunocytochemistry

Cells were plated onto tissue culture treated glass slides, with 2 cambers (Falcon *Culture slides* #354112) and transfected the following day at a density of 60-70% as described above. The usual volume added to these slide chambers is 2ml unless stated otherwise. After two days the cells were washed 3x in PBS and fixed in 1ml 3% Paraformaldehyde solution (pH 7.4, 0.1mM $CaCl_2$ /0.1mM $MgCl_2$) for 20min at room temperature. The cultures were washed 3 x in PBS, quenched with 50mM NH_4Cl/PBS for 10min and washed again 3x in PBS. The slides were stored in PBS over night at 4C. The next day

cells were opened with 0.2% Triton X-100 (w/v) in PBS for 5min, washed 3 x in PBS and again washed in PBS/1% BSA for 5min. 50 to 100µl first antibody (anti-V5-Antibody/mouse monoclonal (1:500)) in PBS/1%BSA was added into each chamber and incubated in a humid environment for 20min at room temperature. The cultures were washed 2 x in PBS/1%BSA for 5min and stained with 50-100µl second antibody (anti-mouse-FITC, DAKO (1:20)) in PBS/1%BSAfor 20min at room temperature. Another wash 3x 5min in PBS/1%BSA was performed followed by 2 x 5min in PBS only. The chambers were removed from the slides and the slides were dipped into distilled water. Excess water was removed from the edges with filter paper, a drop of mounting medium (Kaiser's glycerol gelatin, Merck) was added on the slides and the slides were covered with coverslips. The slides were allowed to set for 45 min before they were observed under the microscope (Leica DM*RA* with the DC200 camera).

Transfection into eukaryotic cells and immunoprecipitation:

Further experiments to characterize the PSK proteins are to express tagged and non-tagged psk genes in eukaryotic cells for example mouse Neuro-2A or HCN-1A or HCN-2 human cortical tissue (brain) cells, to prepare cell lysates and to immunoprecipitate the expressed tagged or non-tagged genes with their interacting partners bound. In case of the tag an anti tag antibody is used, the non-tagged proteins precipitate by specific polyclonal antibodies against the psk genes. The interacting partner is analyzed by mass spectrometry and indicates a role in a certain signal transduction event.

Electroporation and Immunoprecipitation of Neuro2A cells

Ten 10cm dishes of Neuro2A cells were grown to 70% confluency. Each dish was electroporated with 40□g of psk-3 cDNA (BioRad; Gene Pulser II Electroporation System). Medium was changed the following day and the cells were harvested two days after transfection in 350ul lysisbuffer per dish (1%NP-40, 150mM NaCl, 50mM Tris pH 7.5 and protease inhibitors). The lysate was cleared by centrifugation and a total of 31.2ul Protein A beads and 12.5 ug anti-V5 antibody per ml lysate were added. The Immunoprecipitation was incubated at 4 C over night on a rotating device. Then the beads were pelleted, washed with lysisbuffer, the remaining proteins were eluted by boiling in Laemmli sample buffer and loaded on a 12% PAA-gel. The gel was silver stained and bands were cut out and analysed via mass spectrometry.

Immunoprecipitation of PSK-3, overexpressed in Neuro2A cells

The binding of other proteins to the PSK receptor molecules will tell about the function of the PSK receptors and will make it possible to analyse the signal transduction pathway, in which these PSK molecules are involved.

Therefore, the mouse neural cell line Neuro-2A was transfected with V5/His-tagged psk-3 cDNA. The cells were harvested after 2 days and PSK-3 protein was immunoprecipitated via the anti-V5 antibody. The precipitate was separated on a 12% PAA-gel, which was silver stained. Arrowheads in Figure 16 indicate several bands which are coprecipitated with the PSK-3 protein (PSK-3 is indicated with a *). The most prominent band around 60 kD is the anti-V5 antibody. The analysis of these bands will show us which proteins receive signalling information from the PSK receptor molecules.

Chimeric proteins

Further studies to characterize the function of the PSK is the cloning of a chimera protein between the insulin receptor and for example PSK1. The insulin receptor is known to have a Tyrosine phosphorylation site in its cytoplasmic domain which gets autophosphorylated upon activation with insulin. Then several different signal transduction events follow. To see if the potential Tyrosine phosphorylation site in PSK1 is also phosphorylated upon stimulation, the transmembrane and the C-terminal cytoplasmic parts of PSK1 are cloned together with the extracellular part of the insulin receptor. Transfection of this construct into eukaryotic cells, expression and stimulation of these cells with insulin, compared with several control cells (a) untransfected, (b) with the insulin receptor, (c) with the extracellular part of the insulin receptor, (d) with the transmembrane domain and the intracellular part of the PSK1 molecule, reveals if there is an intracellular phosphorylation event and a signaling pathway which can be associated with the C-terminal part of PSK1.

Localization studies: immunofluorescense and *in-situ* hybridization:

Antibodies which have been made against different domains of our protein and antibodies against a tag which is fused to our protein are used to study the subcellular localization of the PSK proteins in different embryonic and adult tissues or cell cultures. Colocalization studies with other molecules or pulse chase experiments will give an insight in the role of the PSK proteins.

In-situ hybridization studies will give information on the temporal and spatial distribution of the PSK mRNAs in the developing and the adult organism.

References

Bordier C (1981) Phase separation of integral membrane proteins in Triton X-114 solution. J Biol Chem 256:1604-1607.

Bork P, Beckmann G (1993) The CUB domain. A widespread module in developmentally regulated proteins. J Mol Biol 231:539-545.

Chan S, Chin SS, Kartha K, Nordli DR, Goodman RR, Pedley TA, Hilal SK (1996) Reversible signal abnormalities in the hippocampus and neocortex after prolonged seizures. AJNR Am J Neuroradiol. Oct;17(9):1725-31.

Davis JQ, Lambert S, Bennett V (1996) Molecular composition of the node of Ranvier: identification of ankyrin-binding cell adhesion molecules neurofascin (mucin+/third FNIII domain-) and NrCAM at nodal axon segments. J Cell Biol 135:1355-1367.

Eck MJ, Dhe-Paganon S, Trub T, Nolte RT, Shoelson SE (1996) Structure of the IRS-1 PTB domain, bound to the juxtamembrane region of the insulin receptor. Cell 85:695-705.

Furley AJ, Morton SB, Manalo D, Karagogeos D., Dodd J, Jessell TM (1990) The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. Cell 61:157-170.

Gennarini G, Cibelli G, Rougon G, Mattei MG, Goridis C (1989) The mouse neuronal cell surface protein F3: a phosphatidylinositol-anchored member of the immunoglobulin superfamily related to chicken contactin. J. Cell Biol. 109:775-788.

Gonzalez-Amaro R, Sanchez-Madrid F (1999) Cell adhesion molecules: selectins and integrins. Crit Rev Immunol. 19(5-6):389-429

Hashiguchi T, Ichinose A (1995) Molecular and cellular basis of deficiency of the β subunit for factor XIII secondary to a Cys430-Phe mutation in the seventh Sushi domain. J Clin Invest 95:1002-1008.

Hourcade D, Holers VM, Atkinson JP (1989) The regulators of complement activation (RCA) gene cluster. Adv Immunol 45:381-416.

Jensen ON, Podtelejnikov A, Mann M (1996) Delayed extraction improves specificity in database searches by MALDI peptide maps. Rapid Commun Mass Spectrom 10:1371 - 1378.

Kajiwara K, Sugaya E, Kimura M, Katsuki M, Nagasawa H, Yuyama N, Tsuda T, Sugaya A, Motoki M, Ookura T, et al (1995) Cloning and characterization of pentylenetetrazol-related cDNA, PTZ-17. Brain Res Feb 6 671(1):170-174.

Kajiwara K, Nagawawa H, Shimizu-Nishikawa S, Ookuri T, Kimura M, Sugaya E (1996a) Molecular characterization of seizure-related genes isolated by differential screening. Biochem Biophys Res Commun Feb 27 219(3):795-799.

Kajiwara K, Nagasawa H, Shimizu-Nishikawa K, Ookura T, Kimura M, Sugaya E (1996b) Cloning of SEZ-12 encoding seizure-related and membrane-bound adhesion protein. Biochem Biophys Res Commun May 6 222(1):144-148.

Kanki JP, Chang S, Kuwada JY (1994) The molecular cloning and characterization of potential chick DM-GRASP homologs in zebrafish and mouse. J. Neurobiol. 25:831-845.

Kristensen T, D'Eustachio P, Ogata RT, Chung LP, Reid KB, Tack BF (1987) The superfamily of C3b/C4b-binding proteins. Fed Proc 46:2463-2469.

Laurie DJ, Schrotz PCU (1994) Processing vertebrate embryonic and early postnatal tissue for *in situ* hybridization. In: *In situ* hybridization protocols for neurobiology (Wisden W, Morris BJ, eds), chapter two. Academic Press.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680.

Mann M (1996) A shortcut to interesting human genes: Peptide Sequence Tags, ESTs and Computers. TIBS 21:494–495.

Mann M, Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal Chem 66:4390-4399.

Moos M, Tacke R, Scherer H, Teplow D., Frueh K, Schachner M (1988) Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. Nature 334:701-703.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10:1-6.

Norman DG, Barlow PN, Baron M, Day AJ, Sim RB, Campbell ID (1991) Three-dimensional structure of a complement control protein module in solution. J Mol Biol 219:717-725.

Patschinsky T, Hunter T, Esch FS, Cooper JA, Sefton BM (1982) Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. Proc Natl Acad Sci USA. 79:973-977.

Hunter T (1982) Synthetic peptide substrates for a tyrosine protein kinase. J. Biol. Chem. 257:4843-4848.

Cooper JA, Esch FS, Taylor SS, Hunter T (1984) Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. J. Biol. Chem. 259:7835-7841.

Ripoche J, Day AJ, Harris TJ, Sim RB (1988) The complete amino acid sequence of human complement factor H. Biochem J 249:593-602.

Schrotz-King P, Wilm M, Andersen JS, Ashman K, Podtelejnikov AV, Bachi A, King A, Mann M (1999) Use of mass spectrometric methods for protein identification in receptor research. Journal of Receptor and Signal Transduction Research, 19(1-4), 659-672.

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proc Natl Acad Sci USA 95:5857-5864.

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Shevchenko An, Boucherie H, Mann M (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA 93 (25):14440-14445.

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal Chem Mar 1 68(5):850-858.

Shimizu-Nishikawa K, Kajiwara K, Kimura M, Katsuki M and Sugaya E (1995a) Cloning and expression of *sez-6*, a brain-specific and seizure-related cDNA. Brain Res Mol Brain Res Feb 28(2):201-210.

Shimizu-Nishikawa K, Kajiwara K and Sugaya E (1995b) Cloning and characterization of seizure-related gene, *sez-6*. Biochem Biophys Res Commun Nov 2 216(1):382-389.

Tao W, Lai E (1992) Telencephalon-restricted expression of Bf-1, a new member of the HNF-3/fork-head gene family, in the developing rat brain. Neuron 8:957-966.

Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. Anal Chem 68:1-8.

Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, and Mann M (1996) Femtomole sequencing of proteins from polyacrylamide gels by nanoelectrospray mass spectrometry. Nature 379(6564):466-469.

Wisden W, Morris BJ, Hunt, SP (1991) *In situ* hybridization with synthetic DNA probes. In: Molecular Neurobiology: A Practical Approach (Chad J, Wheal H, eds), pp 205-225. New York: Oxford UP.

Vorm O, Roepstorff P, Mann M (1994) Matrix surfaces made by fast evaporation yield improved resolution and very high sensitivity in MALDI TOF. Anal Chem 66:3281-3287.

Claims

1. A polypeptide, which has a homology of at least 70% to an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 4, and 6 or a polypeptide fragment of at least 6 amino acids with a homology of at least 80% to an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, and 6 wherein the polypeptide fragment has the same functional characteristics as the polypeptide or the polypeptide fragment has the same immunological effect as the polypeptide, wherein the functional characteristics are determined by testing the polypeptide and the polypeptide fragment in an assay, and the immunological effect is determined by exposing the polypeptide and the polypeptide fragment to the same immunological test.

2. A polypeptide or polypeptide fragment according to claim 1, wherein the homology is at least 75%, such as 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%.

3. The polypeptide with SEQ ID NO: 2.

4. The polypeptide with SEQ ID NO: 4.

5. The polypeptide with SEQ ID NO: 6.

6. A polypeptide or polypeptide fragment according to claim 1 or 2, wherein the polypeptide or polypeptide fragment has been modified compared only by conservative substitutions.

7. A polypeptide fragment according to any of the preceding claims which has a length of at least 12 amino acids, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 20, at least 22, at least 24 and at least 30 amino acid residues.

8. A polypeptide fragment according to any of the preceding claims, which has a length of at most 500 amino acid residues, such as at most 400 amino acid residues, at most 300, at most 200, at most 100, at most 80, at most 60, at most 50, at most 40, at most 30, at most 20 and at most 15 amino acid residues.

9. A polypeptide or polypeptide fragment according to any of the preceding claims in substantially pure form.

10. A fusion polypeptide comprising at least one polypeptide fragment according to any of the preceding claims and at least one fusion partner, said fusion partner being selected from the group consisting of GFP, GST, Myc, HIS, Flag and V5.

11. A polypeptide or polypeptide fragment according to any of the preceding claims coupled to a carbohydrate or a lipid moiety.

12. A polypeptide according to any of claims 1-11 which is glycosylated and/or phosphorylated.

13. A substantially pure polypeptide according to any of the preceding claims for use as a pharmaceutical.

14. Use of a compound, capable of interacting with the polypeptide or polypeptide fragment of claims 1-9 for the preparation of a medicament for the treatment or prophylaxis of epilepy or ischemia.

15. An isolated and purified fragment having a polynucleotide sequence encoding a polypeptide comprising the amino acid sequence of a member of the *psk*-gene family according to claim 1.

16. An isolated and purified polynucleotide sequence encoding a polypeptide comprising the amino acid sequence of a polypeptide according to any of claims 1-13.

17. An isolated and purified polynucleotide sequence comprising the nucleotide sequence as shown in SEQ ID NOs: 1, 3 and/or 5.

18. An isolated and purified polynucleotide sequence which is at least 15 bases long and hybridizes under moderately stringent conditions with a nucleic acid fragment which has a nucleotide sequence as shown in SEQ ID NOs:1, 3 and/or 5, said polynucleotide sequence being located within nucleic acids 1-826 of SEQ ID NO:1.

19. A nucleic acid fragment according to any of claims 15-19, which is a DNA fragment.

20. A nucleic acid fragment according to any of claims 15-19, which is a RNA fragment.

21. A nucleic acid fragment according to any of claims 15-20 for use as a pharmaceutical.

22. An expression vector comprising the polynucleotide of any of claims 15-18.

23. An expression vector according to claim 22 which is replicable.

24. An expression vector according to claim 23, which is selected from the group consisting of a virus, a bacteriophage, a plasmid, a cosmid, a Bacterial Artificial Chromosome (BAC), a Yeast Artificial Chromosome (YAC), a Mouse Artificial Chromosome (MAC) and a microchromosome.

25. An expression vector according to claim 24 containing a suitable promotor.

26. A host cell comprising the expression vector according to any of claims 22-25.

27. A host cell according to claim 26 which is a microorganism, preferably a bacterium, a yeast, a fungus, a protozoan, an insect cell, a plant cell, a mammalian cell or a cell line.

28. A host cell according to claim 27 which is a bacterium of the genus *Bacillus*, eg *B. subtilis*, *Escherichia*, eg *E. coli*, *Pseudomona, Streptomyces* or *Salmonella*.

29. A method of producing the polypeptide defined in any of claims 1-13, comprising cultivating or breeding a host cell according to any of claims 26-28 under conditions leading to expression of said polypeptide, and recovering the polypeptide.

30. A method of producing the polypeptide defined in any of claims 1-13, the method comprising the following steps:

a) inserting a DNA fragment as defined in any of claims 15-18, optionally in suitably modified form, in an expression vector,

b) transforming a suitable host cell with the vector produced in step a),

c) cultivating the microorganism produced in step b) under suitable conditions for expressing the polypeptide,

d) harvesting the polypeptide from the culture, and

e) optionally subjecting the polypeptide to posttranslational modifications.

31. A method according to claim 29 or 30, wherein the DNA fragment coding for the polypeptide is modified by substitution, addition, insertion, deletion or rearrangement of one or more nucleotides in the fragment.

32. A method according to any of claims 29-31, wherein the organism is subjected to mutation.

33. A method according to any of claims 29-32, wherein the polypeptide is isolated from the culture by a method comprising immunoprecipitation, gel filtration, ion-exchange chromatography or one or more affinity chromatography steps employing eg immobilized antibodies reactive with said polypeptide, and/or size chromatography steps.

34. A method of producing a polypeptide as defined in any of claims 1-13, the method comprising performing a liquid or solid phase peptide synthesis.

35. A method for detecting a polynucleotide encoding a member of the *psk*-gene family in a biological sample containing nucleic acids, for the diagnosis of seizure related conditions or other neurodegenerative conditions in a mammal, the method comprising the steps of:

(a) hybridizing the polynucleotide according to claim 19 or 20 to at least one of the nucleic acids in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding a member of the *psk*-gene family in the biological sample.

36. The method of claim 35 wherein the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to hybridization.

37. A method for detecting a polynucleotide encoding a member of the *psk*-gene family in a biological sample containing nucleic acids by amplifying the nucleic acids in the sample by the polymerase chain reaction using amplification primers capable of hybridizing to a polynucleotide encoding a member of the *psk*-gene family.

38. A composition for diagnosing seizure related conditions or other neurodegenerative such as Alzheimer, Rasmussen's Encephalitis, Parkinson's disease, multiple sclerosis, cerebrovascular disorders (stroke syndromes) like ischaemia, Huntington's disease or schizophrenia in a mammal, comprising a nucleic acid fragment according to any of claims 15-18, optionally in combination with a means for detection.

39. A method for preventing or treating a condition in a mammal caused by deficiency or impaired function of the protein with an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4 and 6 or any subsequence thereof, comprising introducing a cDNA encoding an active form of said amino acid sequence

40. A method for preventing or treating a condition caused by PSK upregulation and overexpression in a mammal, comprising introducing a cDNA encoding an inactive form of PSK, in the mammal.

41. A method of detecting a polypeptide of the *psk*-gene family in a tissue comprising treating the tissue with an antibody that is capable of binding to a polypeptide of any of claims 1-13, for a time sufficient to allow a complex to form between said antibody and any PSK present in the tissue, and visualizing the presence of the complex, if any.

42. A method according to claim 41, wherein the substance is a monoclonal antibody.

43. A method according to claim 42, wherein the antibody is an antibody that distinguishes between various forms of PSK polypeptides.

44. A method for quantification of a polypeptide of the *psk*-gene family in biological material comprising specifically binding, to the PSK, antibodies against the PSK and detecting the presence of bound antibodies or assessing the amount of bound antibodies.

45. A method according to claim 44 in which the assessment of the amount of bound antibodies is performed by a method of the ELISA type.

46. A method according to claim 44 in which the assessment of the amount of bound antibodies is performed by a radioimmune assay.

47. A method according to any of claims 44-46, wherein the biological material is tissue, and the method is used for the diagnosis of seizure and/or assessment of prognosis in seizure and/or epilepsy.

48. A method for the detection of soluble antibodies against PSK and/or autoimmune antibodies

49. A method for the detection of ligands of the PSK-family receptor.

50. A method according to any of claims 47-49, wherein the tissue is a sample of blood.

51. A method according to claim 47, wherein the tissue is a sample of serum.

52. A method according to claim 47, wherein the tissue is a sample of plasma.

53. A method according to claim 47, wherein the tissue is a sample of urine.

54. A method according to claim 47, wherein the tissue is a tissue section.

55. A method according to any of claims 50-54 in which the method is an assay for assaying on an extract from a seizure affected tissue or a suspected seizure affected tissue.

56. The use according to any of claims 41-55 in which the assay is used in a prognostic *in vitro* assay.

57. The use according to any of claims 41-55 in which the assay is used in a diagnostic *in vitro* assay.

58. Polyclonal antibodies which are reactive with the polypeptide as defined in any of claims 1-5.

59. Monoclonal antibodies which are reactive with the polypeptide as defined in any of claims 1-5.

60. An antibody according to claim 58 or 59 provided with a detectable label.

61. A polypeptide according to any of claims 1-5 provided with a detectable label.

62. A polypeptide according to claim 61 or an antibody according to claim 60, wherein the label is selected from the group consisting of enzymes, fluorophores, radioactive isotopes and complexing agents such as biotin.

63. A polypeptide according to any of claims 1-5 coupled to a solid support.

64. A polypeptide or an antibody according to claim 63, wherein the support is selected from the group consisting of plates, strips, beads, particles, film and paper.

65. An antibody according to claim 58 or 59, coupled to a solid support.

66. An antibody according to claim 65 wherein the support is selected from the group consisting of plates, strips, beads, particles, film and paper.

67. A polypeptide according to claim 64 or an antibody according to claim 66, wherein the solid support comprises a polymer, eg selected from the group consisting of plastics, eg latex, polystyrene, polyvinyl chloride, polyolefin, nylon or polyvinylidene difluoride, cellulose, silicone and silica.

68. A method of determining the presence of a peptide in a sample, wherein the sample is incubated with a monoclonal and/or polyclonal antibody against PSK coupled to a solid support and subsequently with a polypeptide as defined in any of claims 1-4 provided with a label, or wherein the sample is incubated with a polypeptide as defined in any of claims 1-5 coupled to a solid support and subsequently with a monoclonal and/or polyclonal antibody against PSK provided with a label.

69. A method for inhibiting seizures in a mammal, by inhibiting the binding of a PSK receptor to its ligand in the mammal, by administration of a substance which binds to PSK and thereby inhibits the binding of the ligand to a cell-surface PSK, and hence inactivating the function of PSK.

70. A method according to claim 69, wherein the substance is a monoclonal antibody.

71. A method according to claim 69, wherein the substance is an inactivated ligand of PSK.

72. A method for determining whether a substance inhibits the activation of the function of PSK by inhibiting the binding of a ligand to a PSK in the mammal, the method comprising one or more of the following steps:

1) a screening assay in which the possible inhibition of ligand/PSK interaction by the substance is determined by adding the substance to a system comprising immobilized PSK and solubilized ligand, ligand bound to PSK being detected by being labelled or by means of a labelled anti-PSK antibody, or adding the substance to a system comprising immobilized ligand and solubilized PSK, PSK bound to ligand being detected by being labelled or by means of a labelled anti-PSK antibody,

2) an assay in which the possible inhibition of ligand/PSK interaction by the substance is determined by adding the substance to a system comprising PSK and radiolabelled ligand or a derivative thereof, cross-linking any PSK bound to ligand and detecting any cross-linked product by SDS page and autoradiography,

3) an assay in which the possible inhibition of binding of ligand to PSK on the surface of cultured cells is determined by adding the substance to a system comprising ra-

diolabelled ligand or a derivative thereof and cells carrying PSK and detecting any ligand or derivative binding to PSK by gamma counting of the cells.

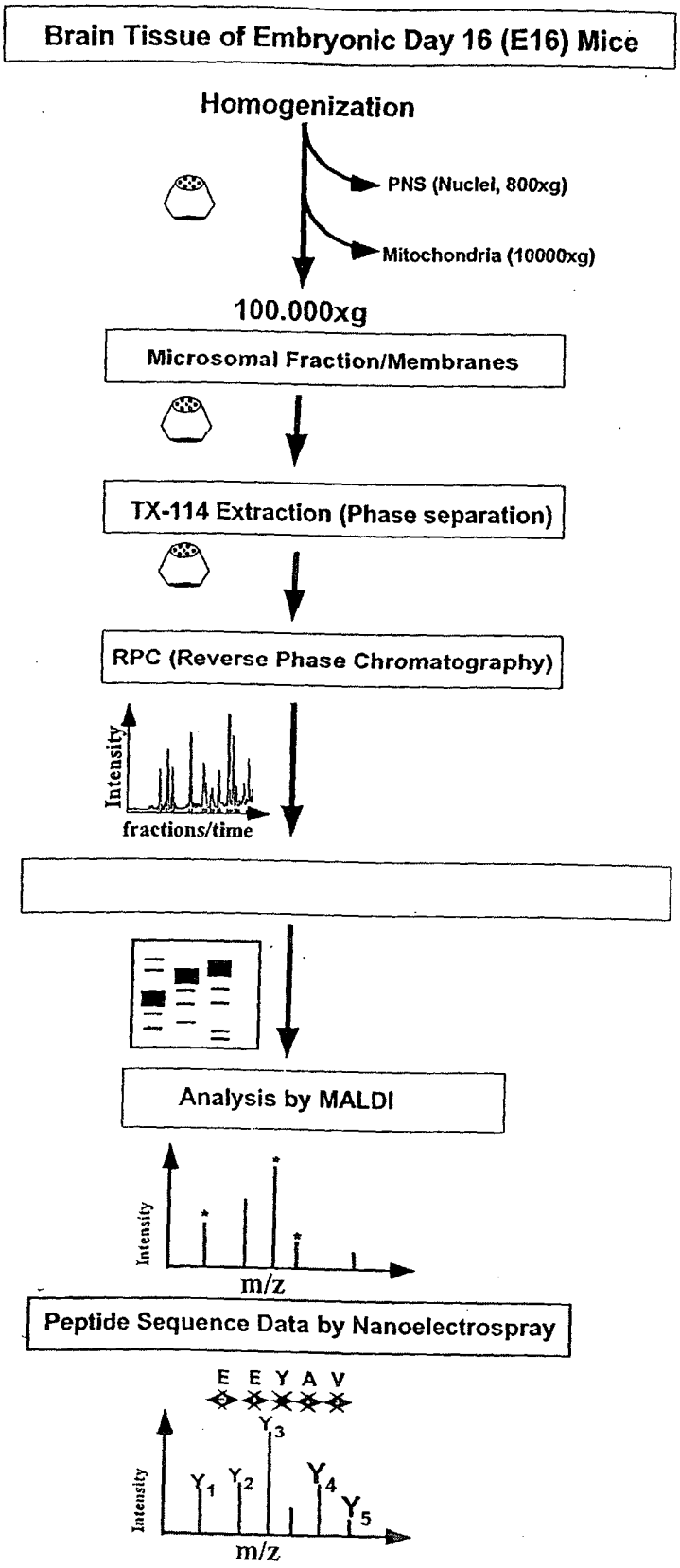

Brain Tissue of Embryonic Day 16 (E16) Mice
Homogenization
PNS (Nuclei, 800xg)
Mitochondria (10000xg)
100.000xg
Microsomal Fraction/Membranes
TX-114 Extraction (Phase separation)
RPC (Reverse Phase Chromatography)
Intensity
fractions/time
Analysis by MALDI
Intensity
m/z
Peptide Sequence Data by Nanoelectrospray
E E Y A V
Intensity
m/z

Fig. 1

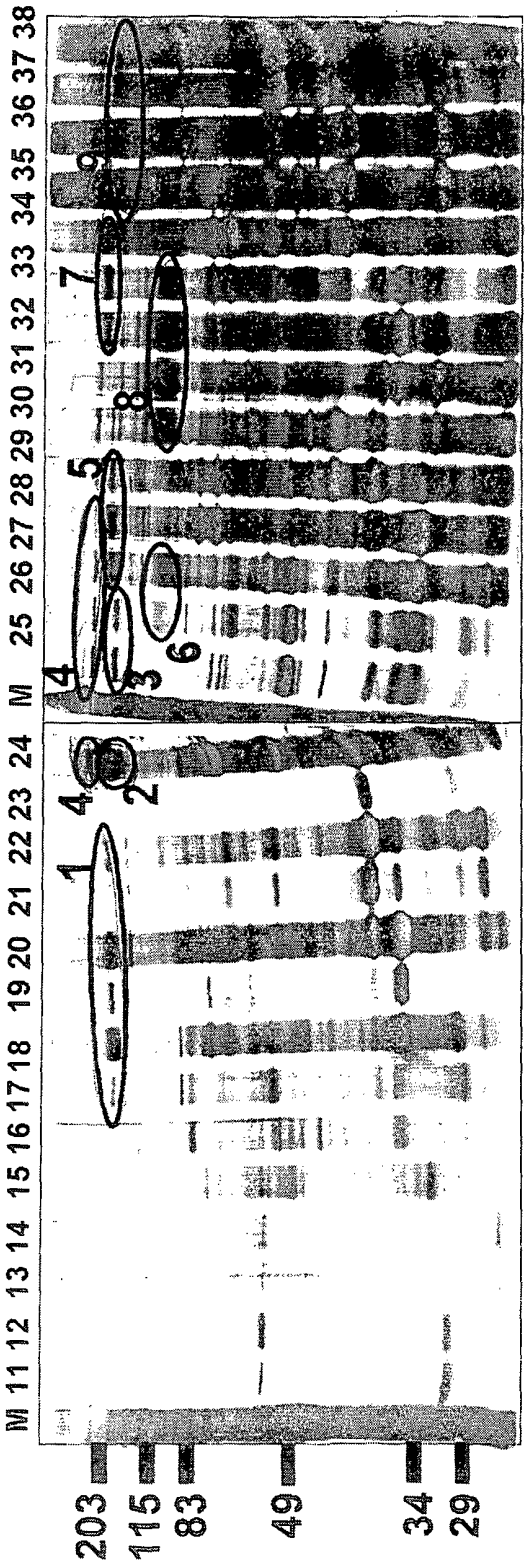

M 11 12 13 14 15 16 17 18 19 20 21 22 23 24 M 25 26 27 28 29 30 31 32 33 34 35 36 37 38
1
2
3
4
5
6
7
8
9
203
115
83
49
34
29

Fig. 2

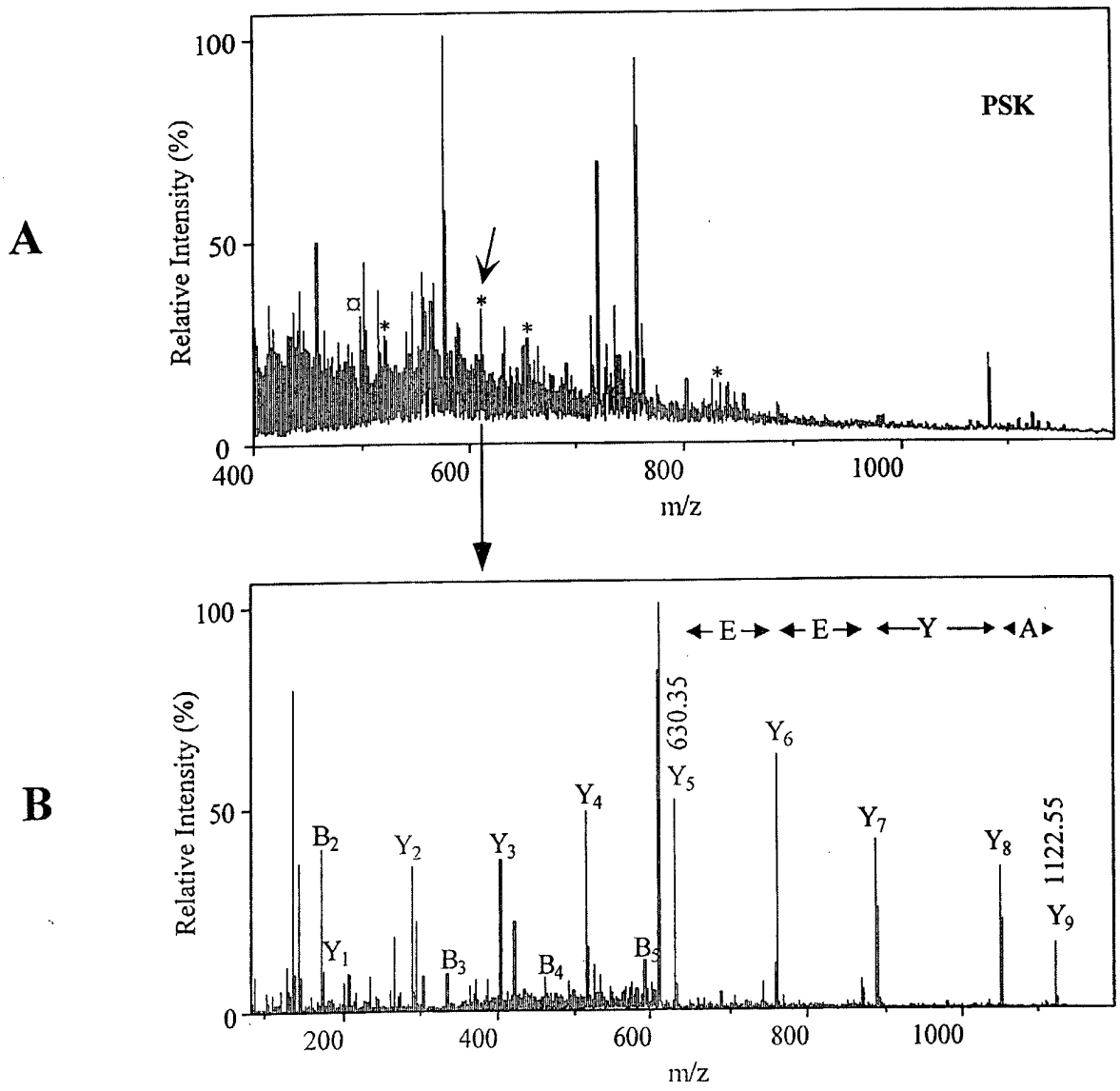

A
B
PSK
Relative Intensity (%)
m/z
100
50
0
400
600
800
1000
200
¤
*
E
E
Y
A
630.35
1122.55
B2
Y2
Y3
Y4
Y5
Y6
Y7
Y8
Y9
Y1
B3
B4
B5

Fig. 3

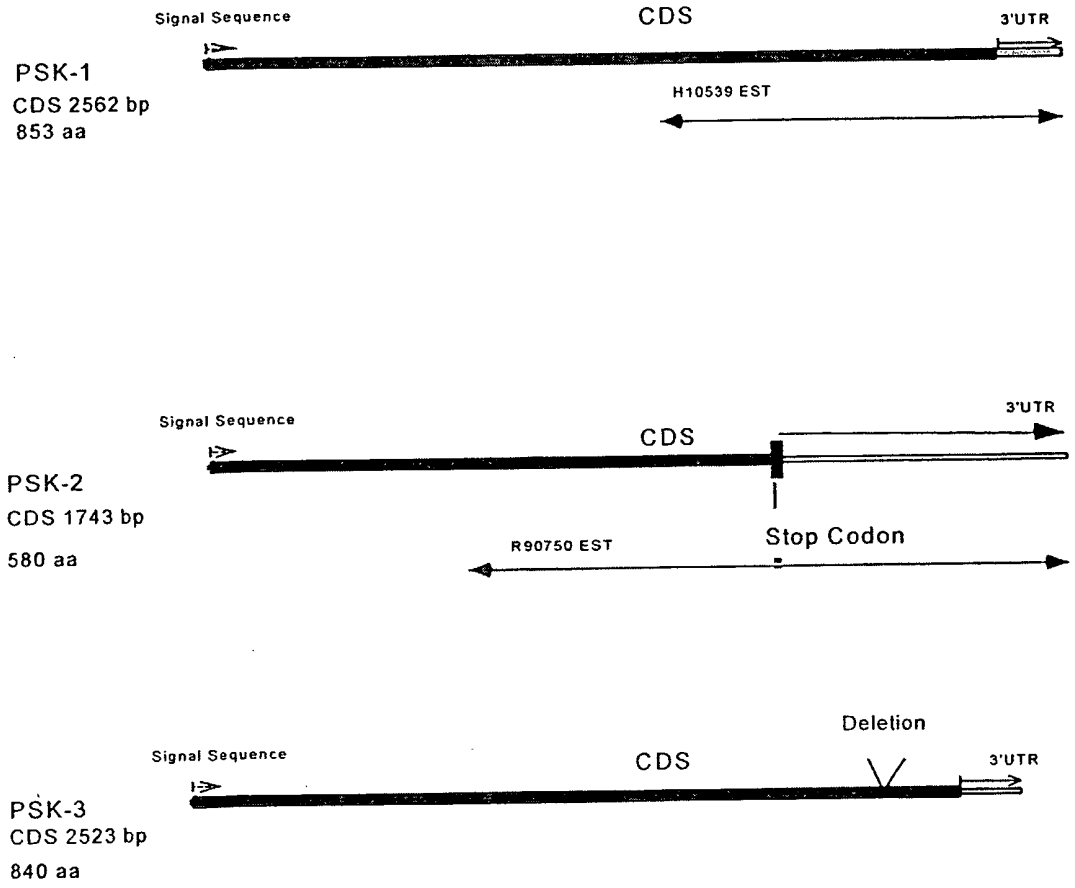

Signal Sequence
CDS
3'UTR
PSK-1
CDS 2562 bp
853 aa
H10539 EST
Signal Sequence
CDS
3'UTR
PSK-2
CDS 1743 bp
580 aa
R90750 EST
Stop Codon
Deletion
Signal Sequence
CDS
3'UTR
PSK-3
CDS 2523 bp
840 aa

Fig. 4

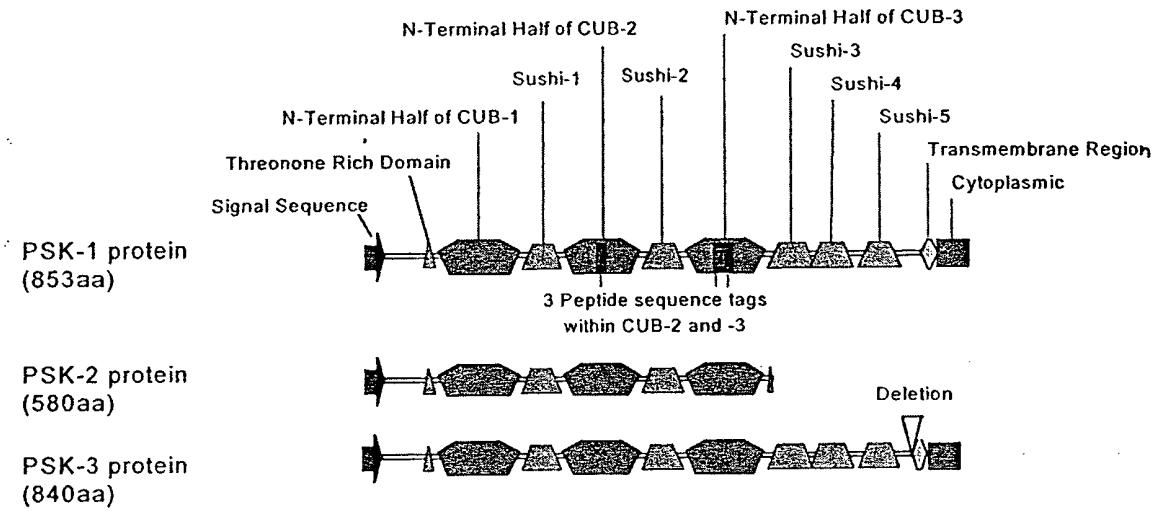

N-Terminal Half of CUB-2
N-Terminal Half of CUB-3
Sushi-3
Sushi-1
Sushi-2
Sushi-4
N-Terminal Half of CUB-1
Sushi-5
Transmembrane Region
Threonone Rich Domain
Cytoplasmic
Signal Sequence
PSK-1 protein
(853aa)
3 Peptide sequence tags
within CUB-2 and -3
PSK-2 protein
(580aa)
Deletion
PSK-3 protein
(840aa)

Fig. 5

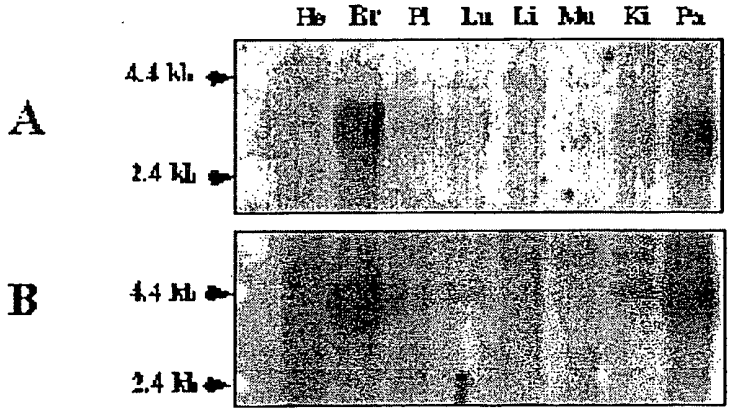


Fig. 6

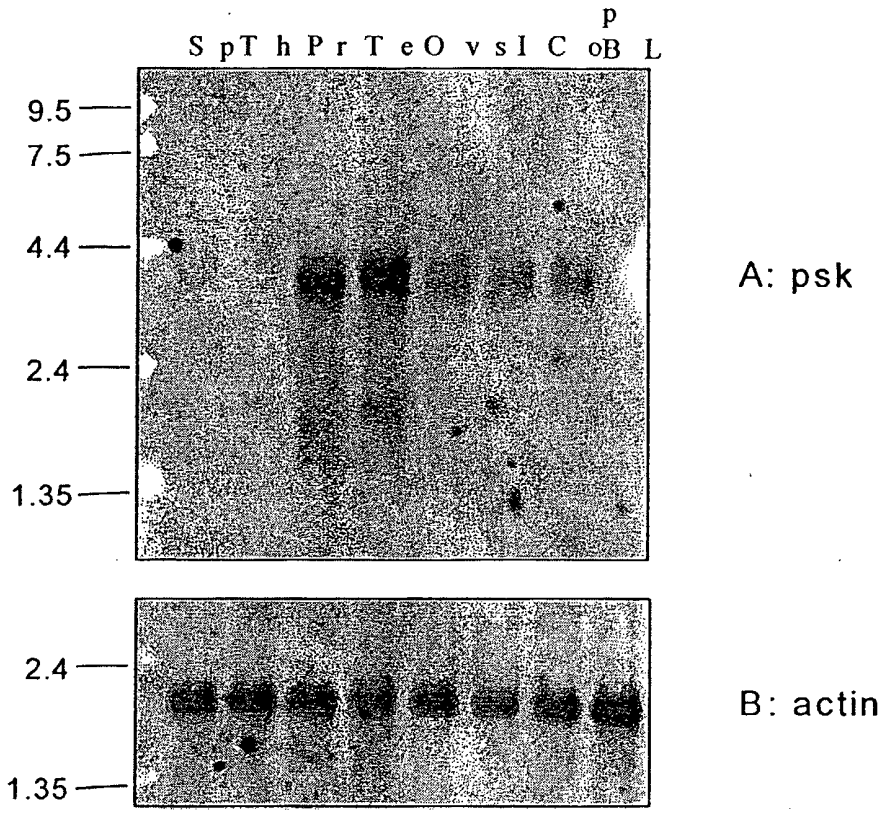


Fig. 7

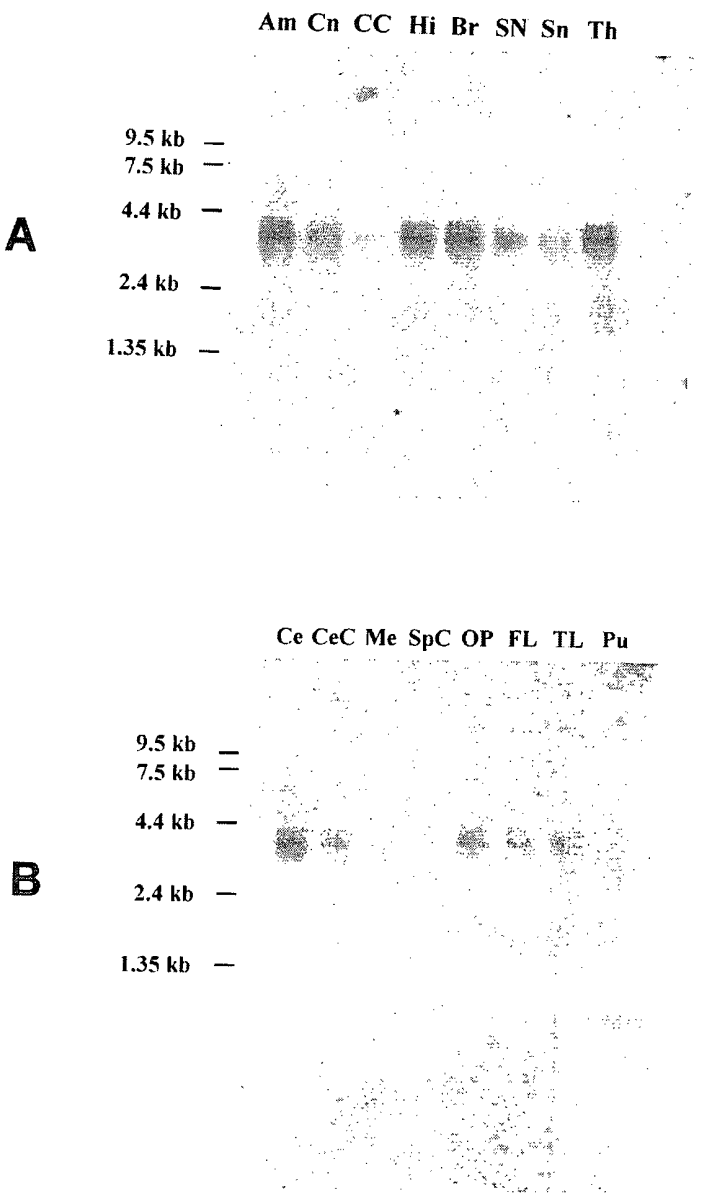

Am Cn CC Hi Br SN Sn Th
9.5 kb
7.5 kb
4.4 kb
2.4 kb
1.35 kb
A
Ce CeC Me SpC OP FL TL Pu
9.5 kb
7.5 kb
4.4 kb
2.4 kb
1.35 kb
B

Fig. 8

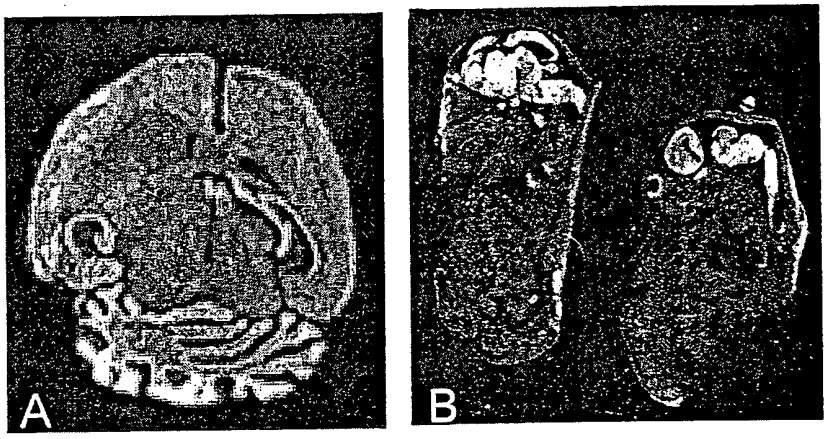


Fig. 9

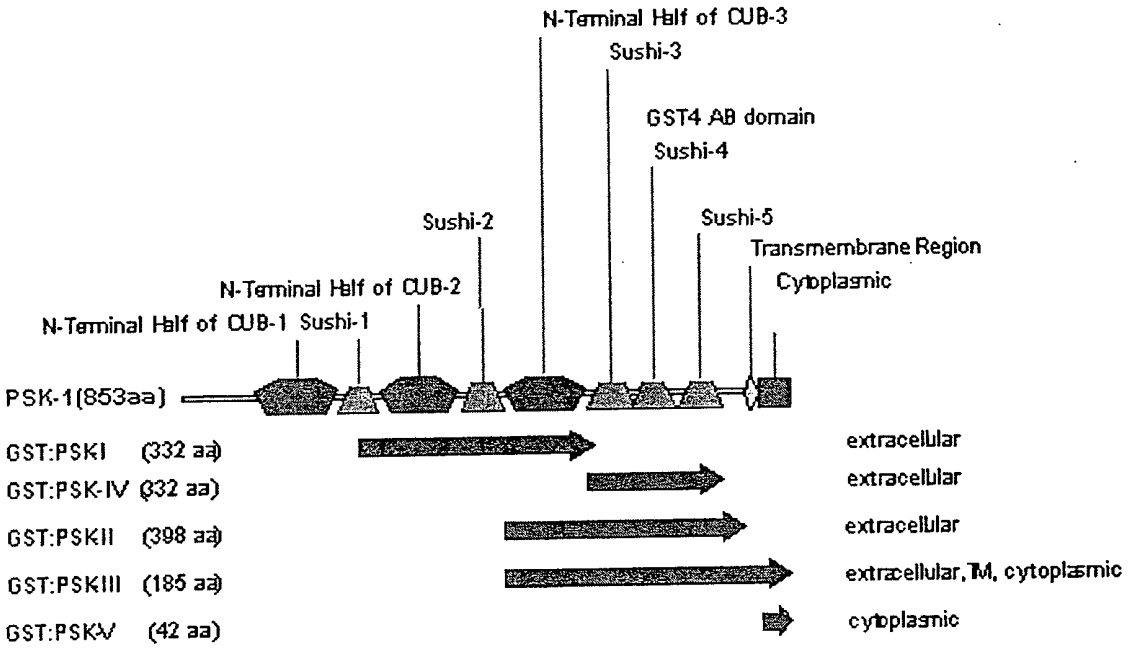
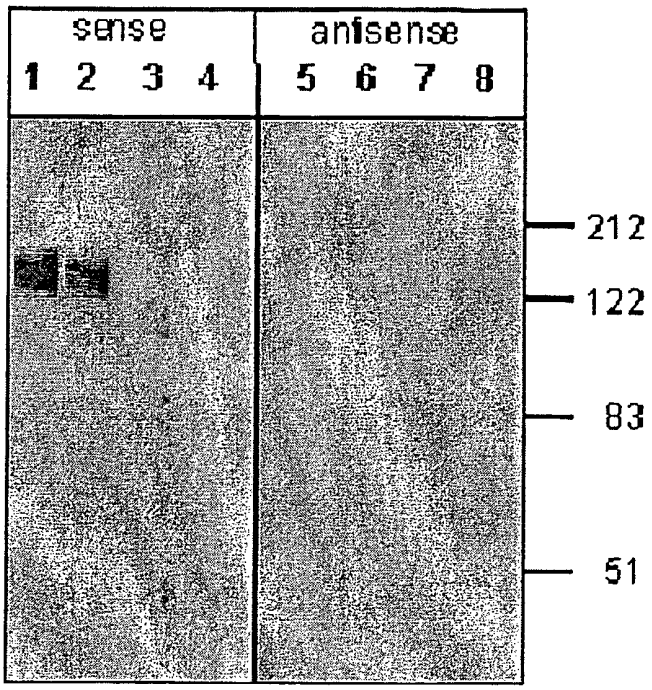

N-Terminal Half of CUB-3
Sushi-3
GST4 AB domain
Sushi-4
Sushi-2
Sushi-5
Transmembrane Region
Cytoplasmic
N-Terminal Half of CUB-2
N-Terminal Half of CUB-1
Sushi-1
PSK-1(853aa)
GST:PSKI (332 aa)
GST:PSK-IV (332 aa)
GST:PSKII (398 aa)
GST:PSKIII (185 aa)
GST:PSKV (42 aa)
extracellular
extracellular
extracellular
extracellular, TM, cytoplasmic
cytoplasmic

Fig. 10

A


sense
antisense
1 2 3 4
5 6 7 8
212
122
83
51

B

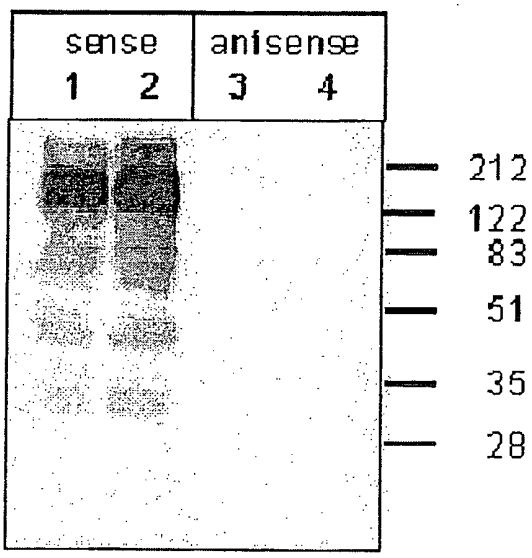

sense
antisense
1 2
3 4
212
122
83
51
35
28

Fig. 11

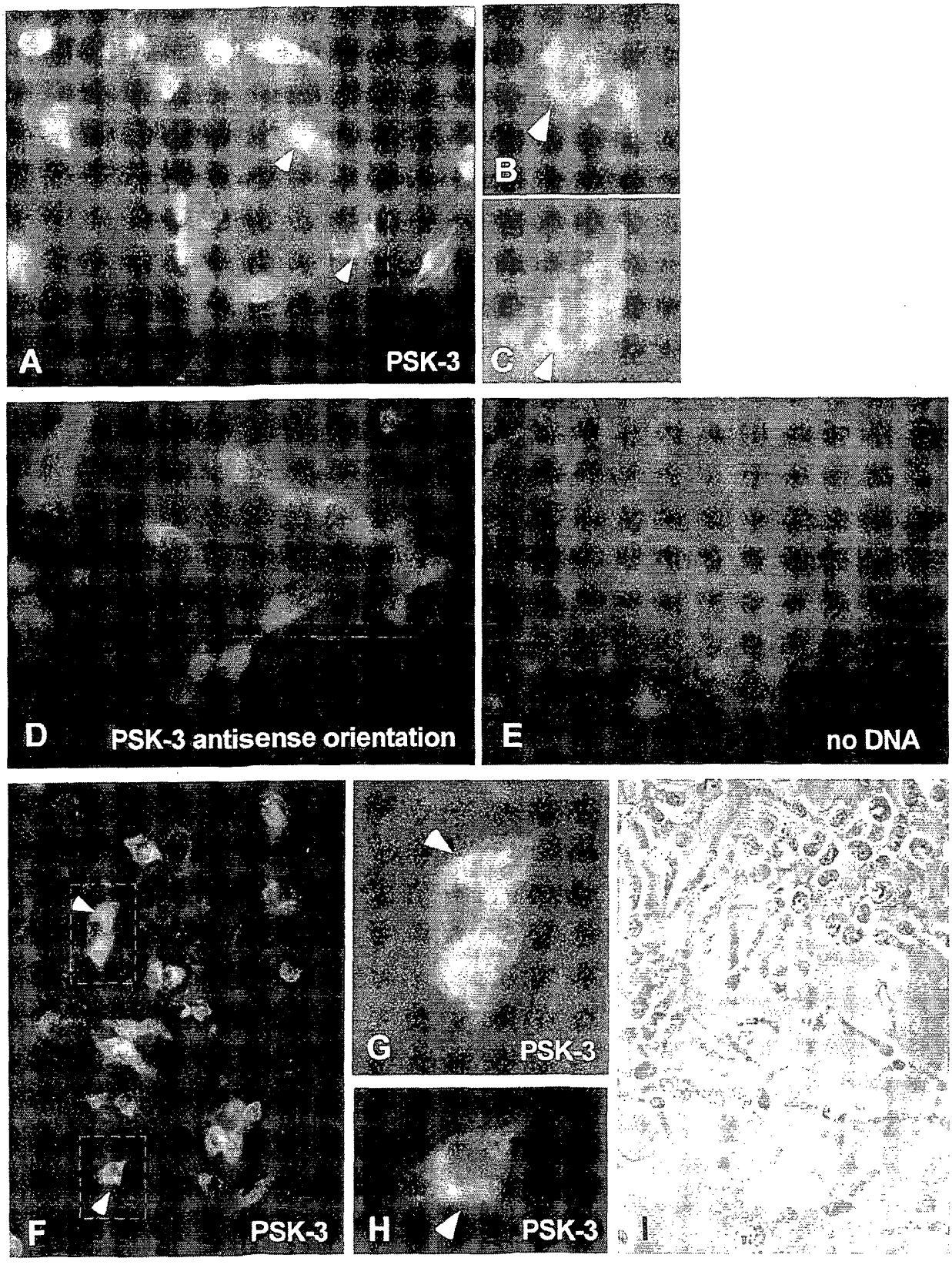

A
PSK-3
B
C
D
PSK-3 antisense orientation
E
no DNA
F
PSK-3
G
PSK-3
H
PSK-3
I

Fig. 12

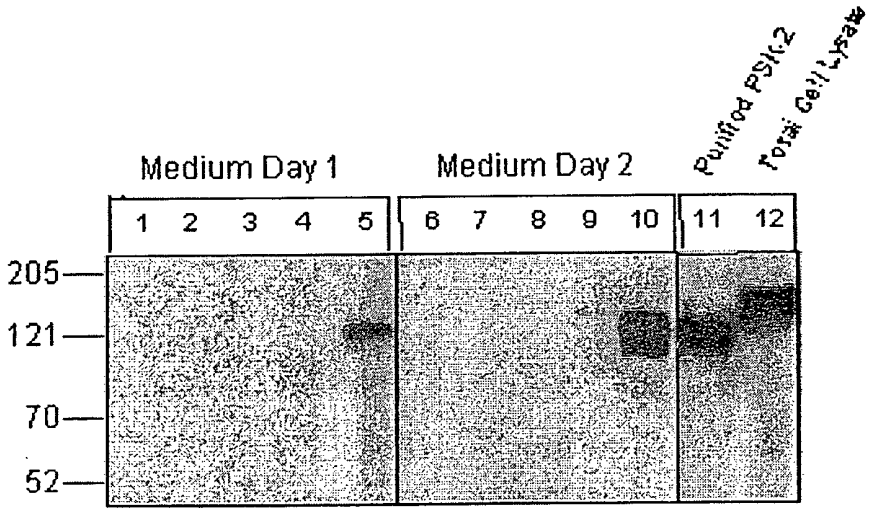

Purified PSP-2
Total Cell Lysate
Medium Day 1
Medium Day 2
1 2 3 4 5 6 7 8 9 10 11 12
205
121
70
52

Fig. 13

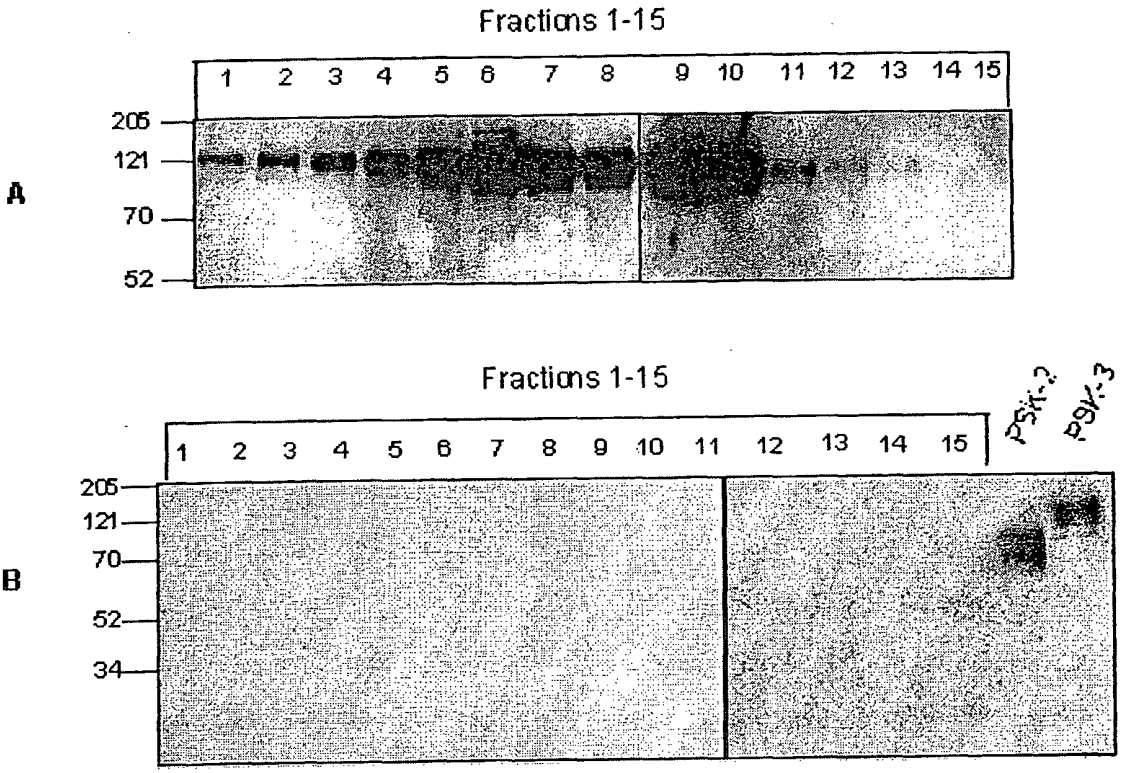

A
Fractions 1-15
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
205
121
70
52
B
Fractions 1-15
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
PSK-2
PSK-3
205
121
70
52
34

Fig. 14

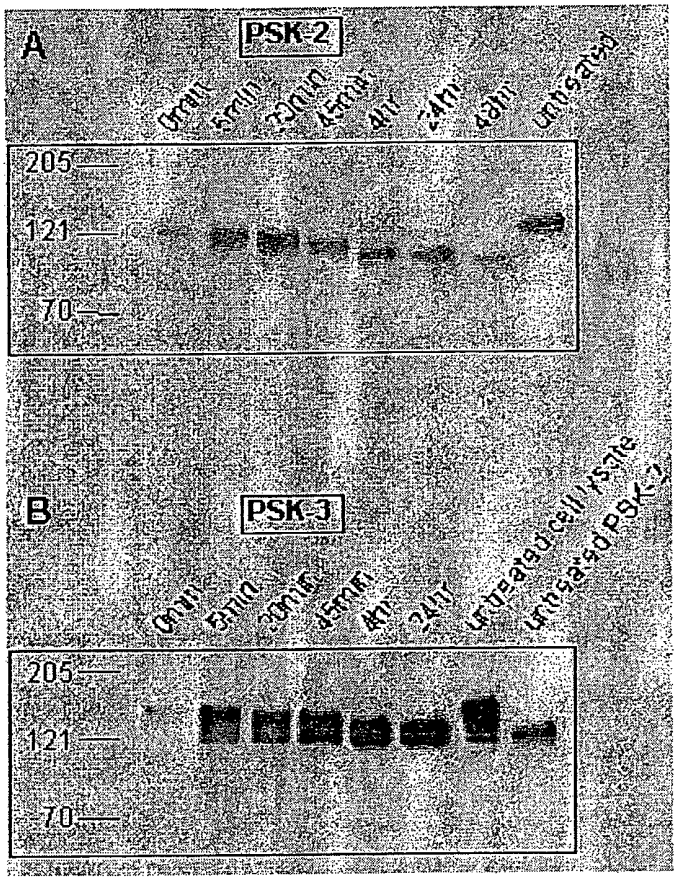


Fig. 15

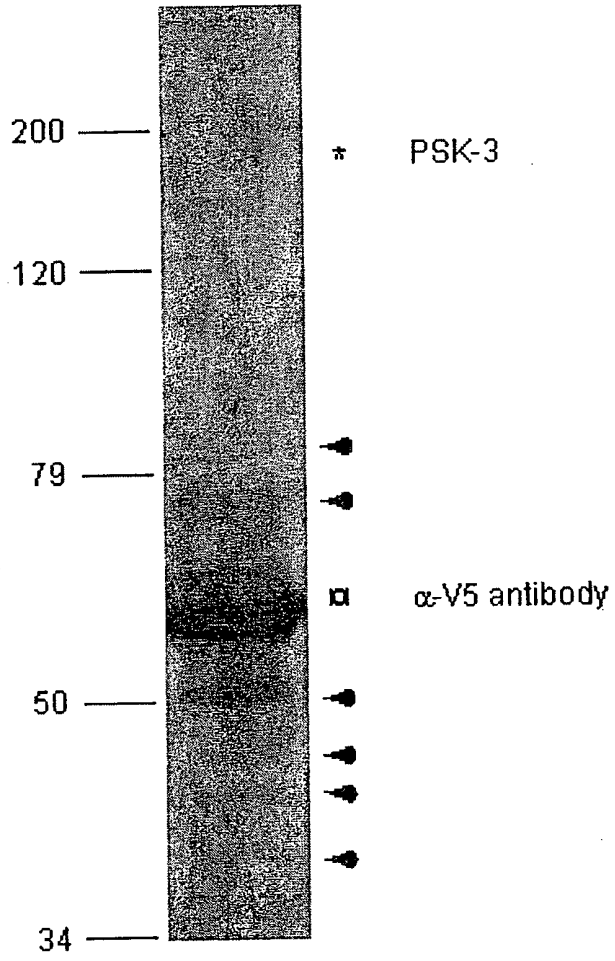


Fig. 16

17/18

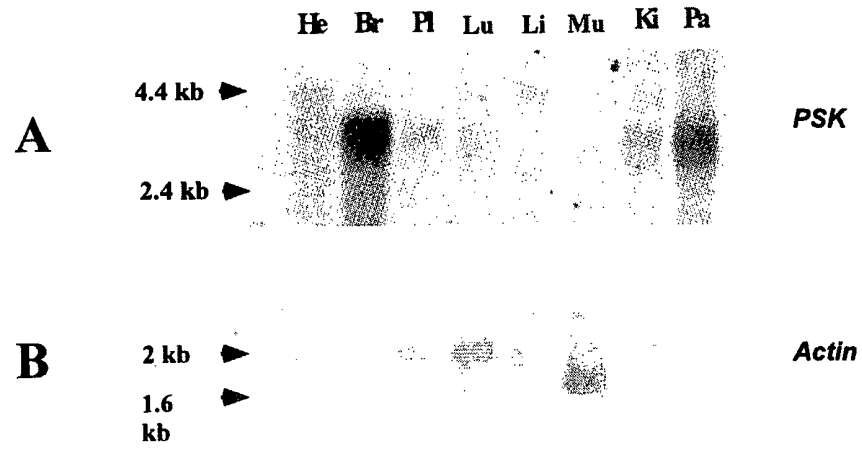


Figure 1. Northern Blot Analysis of Human PSK Gene Expression

(Top Panel) Multiple Tissue Northern Blot, containing 2 μg of poly(A)$^+$ RNA from each of the following tissues, was hybridized to DNA fragments corresponding to the 3 prime PSK2 DNA fragment, including the nucleic acids from base 827-2855 (A) and to the Actin gene probe provided from Clontech with the multiple tissue Northern blots (B):indicated PSK and Actin genes: heart (He), brain (Br), placenta (Pl), Lung (Lu), liver (Li), skeletal muscle (Mu), kidney (Ki), and pancreas (Pa).

(Bottom Panel) The same fragments were used to probe a human multiple tissue Northern Blot, containing 2μg of poly(A)$^+$ RNA from each of the following tissues: Spleen (Sp), thymus (Th), prostate (Pr), testis (Te), ovaries (Ov), small intestine (sI), Colon (Co), and peripheral blood leucocytes (PBL) . Molecular size markers are indicated on the left.

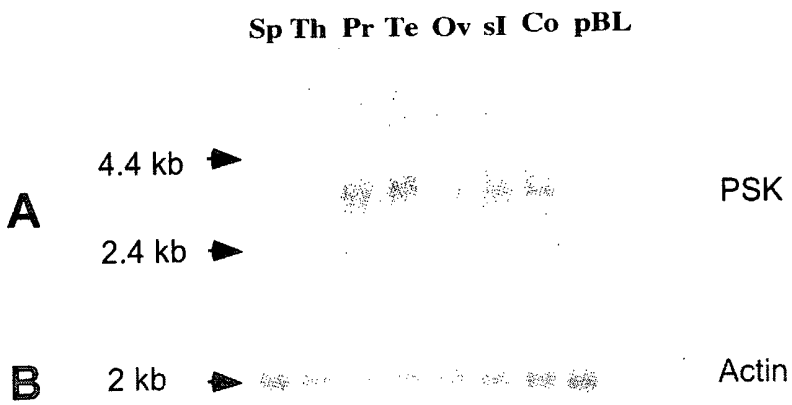


Fig. 17

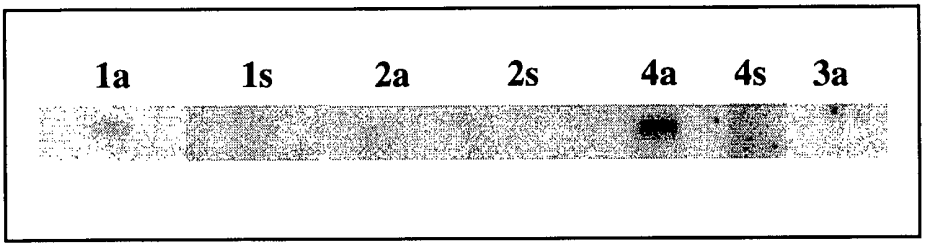


Figure 4. Northern Blot Analysis of mouse PSK Gene Expression

Northern Blot, containing equal amounts of poly(A)$^+$ RNA from adult mouse brain, was hybridized to different oligonucleotides from the human PSK2 cDNA. The olgonucleotides are derived from the following nucleotide sequences of the PSK1 cDNA, the sense oligos read from 5' to 3' and the antisense oligos read as complementatry sequence from 3' to 5': , 1s and 1a from 2362-2406, 2s and 2a from 2251-2295, 3s (here not shown) and 3a from 2687-2731, and 4s and 4a from 469-513. The oligonucleotides 1s and 1a contain the nucleotides which are deleted in PSK3, the oligonucleotides 3s and 3a localize to the 3'-untranslated region of the PSK genes.

Fig. 18

SEQUENCE LISTING

<110> Petra Schrotz-King

<120> Human seizure related proteins

<130> 22704PCl

<160> 6

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 5710
<212> DNA
<213> human

<400> 1
cgcccgggca ggtcggagac ggaggaaagg tggcagccag attacttaga gaggcacaga 60
gcgggcccgt ccagcctctg cctcctttcc accgtcggtc taatgaatct ctccgtgtct 120
ggagagagat cggggtgagt cgccatgggg actcccaggg cccagcaccc gccgcctccc 180
cctctctcta gccccactca gcggtacccc tgagggtccc gggtcgtggg cggcggaggg 240
cagctgctgt tcctaattct gctgagctgt ccctggatcc agggtctgcc cctgaaggag 300
gtcgacgaca aggattaaga cgactcgaca gggacctagg tcccagacgg ggacttcctc 360
gaggagatat tgccagagcc tggaagtgag acccccacgg tggcctctga ggccctggct 420
ctcctctata acggtctcgg accttcactc tgggggtgcc accggagact ccgggaccga 480
gaactgcttc atggggccct gctgaggagg ggcccagaga tgggctacct gccagggcct 540
cttgacgaag taccccggga cgactcctcc ccgggtctct acccgatgga cggtcccgga 600
ccccttgggc ctgagggagg agaggaggag acgacgacca ccatcatcac cacgacaact 660
ggggaacccg gactccctcc tctcctcctc tgctgctggt ggtagtagtg gtgctgttga 720
gttaccacta cggtgaccag cccagttctg tgtaataaca acatctccga gggcgaaggg 780
caatggtgat gccactggtc gggtcaagac acattattgt tgtagaggct cccgcttccc 840
tatgtggagt ctccagatct ggggagcccc gtcagccgca ccctggggct cctggactgc 900
atacacctca gaggtctaga cccctcgggg cagtcggcgt gggaccccga ggacctgacg 960

acttacagca tccatgtcta ccctggctac ggcattgaga tccaggtgca gacgctgaac
1020
tgaatgtcgt aggtacagat gggaccgatg ccgtaactct aggtccacgt ctgcgacttg
1080
ctgtcacagg aagaggagct cctggtgctg gctggtgggg gatccccagg cctggccccc
1140
gacagtgtcc ttctcctcga ggaccacgac cgaccacccc ctaggggtcc ggaccggggg
1200
cgactcctgg ccaactcatc catgcttgga gaaggacaag tccttcggag cccaaccaac
1260
gctgaggacc ggttgagtag gtacgaacct cttcctgttc aggaagcctc gggttggttg
1320
cggctgcttc tgcacttcca gagcccacgg gtcccaaggg gcggtggctt caggatccac
1380
gccgacgaag acgtgaaggt ctcgggtgcc cagggttccc cgccaccgaa gtcctaggtg
1440
tatcaggcct acctcctgag ctgtggcttc cctccccggc cggcccatgg ggacgtgagt
1500
atagtccgga tggaggactc gacaccgaag ggaggggccg gccgggtacc cctgcactca
1560
gtgacggacc tgcaccctgg gggcactgcc acctttcact gtgattcggg ctaccagctg
1620
cactgcctgg acgtgggacc cccgtgacgg tggaaagtga cactaagccc gatggtcgac
1680
cagggagagg agaccctcat ctgcctcaat ggcacccggc catcctggaa cggtgaaacc
1740
gtccctctcc tctgggagta gacggagtta ccgtgggccg gtaggacctt gccactttgg
1800
cccagctgca tggcatcctg tggtggcacc atccacaatg ccaccctggg ccgcatcgtg
1860
gggtcgacgt accgtaggac accaccgtgg taggtgttac ggtgggaccc ggcgtagcac
1920
tccccagagc ctgggggagc cgtagggccc aacctcacct gccgttgggt cattgaagca
1980
aggggtctcg gacccccctcg gcatcccggg ttggagtgga cggcaaccca gtaacttcgt
2040
gctgaggggc gccggctgca cctgcacttt gaaagggtct cgctggatga ggacaatgac
2100
cgactccccg cggccgacgt ggacgtgaaa ctttcccaga gcgacctact cctgttactg
2160
cggctgatgg tgcgctcagg gggcagcccc ctatcccccg tgatctatga ttcggacatg
2220
gccgactacc acgcgagtcc cccgtcgggg gataggggggc actagatact aagcctgtac
2280
gacgatgtcc ccgagcgggg tctcatcagt gacgcccagt ccctctacgt ggagctgctg
2340
ctgctacagg ggctcgcccc agagtagtca ctgcgggtca gggagatgca cctcgacgac
2400
tcagagacac ctgccaatcc cctgctgtta agccttcgat ttgaagcctt tgaggaggat
2460
agtctctgtg gacggttagg ggacgacaat tcggaagcta aacttcggaa actcctccta
2520

cgctgcttcg ccccttcct ggcacatgga aatgtcacta ccacggaccc tgagtatcgc
2580
gcgacgaagc gggggaagga ccgtgtacct ttacagtgat ggtgcctggg actcatagcg
2640
ccaggggcac tggcaacctt ctcgtgcctc ccaggatatg ccctggagcc ccctgggccc
2700
ggtccccgtg accgttggaa gagcacggag ggtcctatac gggacctcgg gggacccggg
2760
cccaatgcca tcgaatgtgt ggatcccaca gaaccccact ggaacgacac agagccggcc
2820
gggttacggt agcttacaca cctagggtgt cttggggtga ccttgctgtg tctcggccgg
2880
tgcaaagcca tgtgtggagg ggagctgtcg gaaccagctg gcgtggtcct ctctcccgac
2940
acgtttcggt acacacctcc cctcgacagc cttggtcgac cgcaccagga gagagggctg
3000
tggccccaga gctatagccc gggccaagac tgcgtgtggg gcgtgcacgt ccaggaagag
3060
accggggtct cgatatcggg cccggttctg acgcacaccc cgcacgtgca ggtccttctc
3120
aagcgcatct tgctccaagt tgagatattg aatgtgcggg aaggggacat gctgacgctg
3180
ttcgcgtaga acgaggttca actctataac ttacacgccc ttcccctgta cgactgcgac
3240
ttcgacgggg acggtcccag cgcccgagtc ttggcccagc tgcggggacc tcagccgcgc
3300
aagctgcccc tgccagggtc gcgggctcag aaccgggtcg acgcccctgg agtcggcgcg
3360
cgccgccttc tctcctctgg gcccgacctc acactgcagt ttcaggcacc gcccgggccc
3420
gcggcggaag agaggagacc cgggctggag tgtgacgtca aagtccgtgg cgggcccggg
3480
ccaaatccag gcctgggcca gggcttcgta ttgcacttca aagaggtccc gaggaacgac
3540
ggtttaggtc cggacccggt cccgaagcat aacgtgaagt ttctccaggg ctccttgctg
3600
acgtgccccg agctgccacc tccggagtgg ggctggagaa cggcatccca cggggacctg
3660
tgcacggggc tcgacggtgg aggcctcacc ccgacctctt gccgtagggt gcccctggac
3720
atccggggca cggtgctcac ctaccagtgc gagcctggct acgagctgct aggctccgac
3780
taggccccgt gccacgagtg gatggtcacg ctcggaccga tgctcgacga tccgaggctg
3840
attctcactt gccagtggga cctgtcttgg agcgccgcgc cgcccgcctg ccaaaagatc
3900
taagagtgaa cggtcaccct ggacagaacc tcgcggcgcg gcgggcggac ggttttctag
3960
atgacttgtg ctgaccctgg cgagattgcc aacgggcacc gcaccgcctc ggacgccggc
4020
tactgaacac gactgggacc gctctaacgg ttgcccgtgg cgtggcggag cctgcggccg
4080

ttccccgttg gctcccacgt ccagtaccgc tgcctgccag ggtacagcct cgaggggca
4140
aaggggcaac cgagggtgca ggtcatggcg acggacggtc ccatgtcgga gctcccccgt
4200
gccatgctca cctgctacag ccgggacaca ggcacaccca agtggagcga tagggtcccc
4260
cggtacgagt ggacgatgtc ggccctgtgt ccgtgtgggt tcacctcgct atcccagggg
4320
aaatgcgcct tgaagtacga gccgtgcctg aacccggggg ttcccgagaa tggctaccag
4380
tttacgcgga acttcatgct cggcacggac ttgggccccc aagggctctt accgatggtc
4440
acgctgtaca agcaccacta ccaggcgggc gagtctctgc gcttcttctg ctatgagggc
4500
tgcgacatgt tcgtggtgat ggtccgcccg ctcagagacg cgaagaagac gatactcccg
4560
tttgagctta tcggcgaggt caccatcacc tgtgtgcccg gccacccctc ccagtggacc
4620
aaactcgaat agccgctcca gtggtagtgg acacacgggc cggtggggag ggtcacctgg
4680
agccagcccc cactctgcaa agttgcctat gaggagctcc tggacaaccg aaaactggaa
4740
tcggtcgggg gtgagacgtt tcaacggata ctcctcgagg acctgttggc ttttgacctt
4800
gtgacccaga ccacagatcc atcacggcag ctggaagggg ggaacctggc cctggccatc
4860
cactgggtct ggtgtctagg tagtgccgtc gaccttcccc ccttggaccg ggaccggtag
4920
ctgctgcctc taggcttggt cattgtcctc ggcagtggcg tttacatcta ctacaccaag
4980
gacgacggag atccgaacca gtaacaggag ccgtcaccgc aaatgtagat gatgtggttc
5040
cttcagggaa agtccctttt cggcttctcg ggctcccact cctacagccc catcaccgtg
5100
gaagtccctt tcagggaaaa gccgaagagc ccgagggtga ggatgtcggg gtagtggcac
5160
gagtcggact tcagcaaccc gctgtatgaa gctggggata cgcgggagta tgaagtttcc
5220
ctcagcctga agtcgttggg cgacatactt cgacccctat gcgccctcat acttcaaagg
5280
atctgaaccc caagactaca gctgcaggac ccaggacgcc cctcccctcc tcattcgggc
5340
tagacttggg gttctgatgt cgacgtcctg ggtcctgcgg ggaggggagg agtaagcccg
5400
agagggaaat acgggacccg gtctctgcct cctggctgcc ctcctccctg gctgtgtaaa
5460
tctcccttta tgccctgggc cagagacgga ggaccgacgg gaggagggac cgacacattt
5520
tagtctccct atcccacgag ggggctttga tggccctgga gatcctacag taaataaacc
5580
atcagaggga tagggtgctc ccccgaaact accgggacct ctaggatgtc atttatttgg
5640

agcatcctgc cgcccaaaaa aaaaaaaaaa aaaaatcgta ggacggcggg ttttttttt
5700
ttttttttt
5710

<210> 2
<211> 853
<212> PRT
<213> human

<400> 2

```
Met Gly Thr Pro Arg Ala Gln His Pro Pro Pro Pro Gln Leu Leu Phe
 1               5                  10                  15
Leu Ile Leu Leu Ser Cys Pro Trp Ile Gln Gly Leu Pro Leu Lys Glu
            20                  25                  30
Glu Glu Ile Leu Pro Glu Pro Gly Ser Glu Thr Pro Thr Val Ala Ser
        35                  40                  45
Glu Ala Leu Ala Glu Leu Leu His Gly Ala Leu Leu Arg Arg Gly Pro
    50                  55                  60
Glu Met Gly Tyr Leu Pro Gly Pro Pro Leu Gly Pro Glu Gly Gly Glu
65                  70                  75                  80
Glu Glu Thr Thr Thr Thr Ile Ile Thr Thr Thr Thr Val Thr Thr Thr
                85                  90                  95
Val Thr Ser Pro Val Leu Cys Asn Asn Asn Ile Ser Glu Gly Glu Gly
            100                 105                 110
Tyr Val Glu Ser Pro Asp Leu Gly Ser Pro Val Ser Arg Thr Leu Gly
        115                 120                 125
Leu Leu Asp Cys Thr Tyr Ser Ile His Val Tyr Pro Gly Tyr Gly Ile
    130                 135                 140
Glu Ile Gln Val Gln Thr Leu Asn Leu Ser Gln Glu Glu Glu Leu Leu
145                 150                 155                 160
Val Leu Ala Gly Gly Gly Ser Pro Gly Leu Ala Pro Arg Leu Leu Ala
                165                 170                 175
Asn Ser Ser Met Leu Gly Glu Gly Gln Val Leu Arg Ser Pro Thr Asn
            180                 185                 190
Arg Leu Leu Leu His Phe Gln Ser Pro Arg Val Pro Arg Gly Gly Gly
        195                 200                 205
Phe Arg Ile His Tyr Gln Ala Tyr Leu Leu Ser Cys Gly Phe Pro Pro
    210                 215                 220
Arg Pro Ala His Gly Asp Val Ser Val Thr Asp Leu His Pro Gly Gly
225                 230                 235                 240
Thr Ala Thr Phe His Cys Asp Ser Gly Tyr Gln Leu Gln Gly Glu Glu
                245                 250                 255
Thr Leu Ile Cys Leu Asn Gly Thr Arg Pro Ser Trp Asn Gly Glu Thr
            260                 265                 270
Pro Ser Cys Met Ala Ser Cys Gly Gly Thr Ile His Asn Ala Thr Leu
        275                 280                 285
Gly Arg Ile Val Ser Pro Glu Pro Gly Gly Ala Val Gly Pro Asn Leu
    290                 295                 300
Thr Cys Arg Trp Val Ile Glu Ala Ala Glu Gly Arg Arg Leu His Leu
305                 310                 315                 320
His Phe Glu Arg Val Ser Leu Asp Glu Asp Asn Asp Arg Leu Met Val
                325                 330                 335
```

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

```
Ser Gln Pro Pro Leu Cys Lys Val Ala Tyr Glu Glu Leu Leu Asp Asn
        755                 760                 765
Arg Lys Leu Glu Val Thr Gln Thr Thr Asp Pro Ser Arg Gln Leu Glu
    770                 775                 780
Gly Gly Asn Leu Ala Leu Ala Ile Leu Leu Pro Leu Gly Leu Val Ile
785                 790                 795                 800
Val Leu Gly Ser Gly Val Tyr Ile Tyr Tyr Thr Lys Leu Gln Gly Lys
                805                 810                 815
Ser Leu Phe Gly Phe Ser Gly Ser His Ser Tyr Ser Pro Ile Thr Val
            820                 825                 830
Glu Ser Asp Phe Ser Asn Pro Leu Tyr Glu Ala Gly Asp Thr Arg Glu
        835                 840                 845
Tyr Glu Val Ser Ile
    850
```

<210> 3
<211> 5710
<212> DNA
<213> human

<400> 3

```
cgcccgggca ggtcggagac ggaggaaagg tggcagccag attacttaga gaggcacaga
    60
gcgggcccgt ccagcctctg cctcctttcc accgtcggtc taatgaatct ctccgtgtct
   120
ggagagagat cggggtgagt cgccatgggg actcccaggg cccagcaccc gccgcctccc
   180
cctctctcta gccccactca gcggtacccc tgagggtccc gggtcgtggg cggcggaggg
   240
cagctgctgt tcctaattct gctgagctgt ccctggatcc agggtctgcc cctgaaggag
   300
gtcgacgaca aggattaaga cgactcgaca gggacctagg tcccagacgg ggacttcctc
   360
gaggagatat tgccagagcc tggaagtgag acccccacgg tggcctctga ggccctggct
   420
ctcctctata acggtctcgg accttcactc tgggggtgcc accggagact ccgggaccga
   480
gaactgcttc atggggccct gctgaggagg ggcccagaga tgggctacct gccagggcct
   540
cttgacgaag taccccggga cgactcctcc ccgggtctct acccgatgga cggtcccgga
   600
ccccttgggc ctgagggagg agaggaggag acgacgacca ccatcatcac cacgacaact
   660
ggggaacccg gactccctcc tctcctcctc tgctgctggt ggtagtagtg gtgctgttga
   720
gttaccacta cggtgaccag cccagttctg tgtaataaca acatctccga gggcgaaggg
   780
caatggtgat gccactggtc gggtcaagac acattattgt tgtagaggct cccgcttccc
   840
tatgtggagt ctccagatct ggggagcccc gtcagccgca ccctggggct cctggactgc
   900
atacacctca gaggtctaga cccctcgggg cagtcggcgt gggaccccga ggacctgacg
```

960
acttacagca tccatgtcta ccctggctac ggcattgaga tccaggtgca gacgctgaac
1020
tgaatgtcgt aggtacagat gggaccgatg ccgtaactct aggtccacgt ctgcgacttg
1080
ctgtcacagg aagaggagct cctggtgctg gctggtgggg gatccccagg cctggccccc
1140
gacagtgtcc ttctcctcga ggaccacgac cgaccacccc ctaggggtcc ggaccggggg
1200
cgactcctgg ccaactcatc catgcttgga gaaggacaag tccttcggag cccaaccaac
1260
gctgaggacc ggttgagtag gtacgaacct cttcctgttc aggaagcctc gggttggttg
1320
cggctgcttc tgcacttcca gagcccacgg gtcccaaggg gcggtggctt caggatccac
1380
gccgacgaag acgtgaaggt ctcgggtgcc cagggttccc cgccaccgaa gtcctaggtg
1440
tatcaggcct acctcctgag ctgtggcttc cctccccggc cggcccatgg ggacgtgagt
1500
atagtccgga tggaggactc gacaccgaag ggaggggccg gccgggtacc cctgcactca
1560
gtgacggacc tgcaccctgg gggcactgcc acctttcact gtgattcggg ctaccagctg
1620
cactgcctgg acgtgggacc cccgtgacgg tggaaagtga cactaagccc gatggtcgac
1680
cagggagagg agaccctcat ctgcctcaat ggcacccggc catcctggaa cggtgaaacc
1740
gtccctctcc tctgggagta gacggagtta ccgtgggccg gtaggacctt gccactttgg
1800
cccagctgca tggcatcctg tggtggcacc atccacaatg ccaccctggg ccgcatcgtg
1860
gggtcgacgt accgtaggac accaccgtgg taggtgttac ggtgggaccc ggcgtagcac
1920
tccccagagc ctgggggagc cgtagggccc aacctcacct gccgttgggt cattgaagca
1980
aggggtctcg gacccctcg gcatcccggg ttggagtgga cggcaaccca gtaacttcgt
2040
gctgaggggc gccggctgca cctgcacttt gaaagggtct cgctggatga ggacaatgac
2100
cgactccccg cggccgacgt ggacgtgaaa ctttcccaga gcgacctact cctgttactg
2160
cggctgatgg tgcgctcagg gggcagcccc ctatcccccg tgatctatga ttcggacatg
2220
gccgactacc acgcgagtcc cccgtcgggg gatagggggc actagatact aagcctgtac
2280
gacgatgtcc ccgagcgggg tctcatcagt gacgcccagt ccctctacgt ggagctgctg
2340
ctgctacagg ggctcgcccc agagtagtca ctgcgggtca gggagatgca cctcgacgac
2400
tcagagacac ctgccaatcc cctgctgtta agccttcgat ttgaagcctt tgaggaggat
2460
agtctctgtg gacggttagg ggacgacaat tcggaagcta aacttcggaa actcctccta

2520
cgctgcttcg ccccttcct ggcacatgga aatgtcacta ccacggaccc tgagtatcgc
2580
gcgacgaagc gggggaagga ccgtgtacct ttacagtgat ggtgcctggg actcatagcg
2640
ccaggggcac tggcaacctt ctcgtgcctc ccaggatatg ccctggagcc ccctgggccc
2700
ggtccccgtg accgttggaa gagcacggag ggtcctatac gggacctcgg gggacccggg
2760
cccaatgcca tcgaatgtgt ggatcccaca gaaccccact ggaacgacac agagccggcc
2820
gggttacggt agcttacaca cctagggtgt cttggggtga ccttgctgtg tctcggccgg
2880
tgcaaagcca tgtgtggagg ggagctgtcg gaaccagctg gcgtggtcct ctctcccgac
2940
acgtttcggt acacacctcc cctcgacagc cttggtcgac cgcaccagga gagagggctg
3000
tggccccaga gctatagccc gggccaagac tgcgtgtggg gcgtgcacgt ccaggaagag
3060
accggggtct cgatatcggg cccggttctg acgcacaccc cgcacgtgca ggtccttctc
3120
aagcgcatct tgctccaagt tgagatattg aatgtgcggg aaggggacat gctgacgctg
3180
ttcgcgtaga acgaggttca actctataac ttacacgccc ttcccctgta cgactgcgac
3240
ttcgacgggg acggtcccag cgcccgagtc ttggcccagc tgcggggacc tcagccgcgc
3300
aagctgcccc tgccagggtc gcgggctcag aaccgggtcg acgcccctgg agtcggcgcg
3360
cgccgccttc tctcctctgg gcccgacctc acactgcagt ttcaggcacc gcccgggccc
3420
gcggcggaag agaggagacc cgggctggag tgtgacgtca aagtccgtgg cgggcccggg
3480
ccaaatccag gcctgggcca gggcttcgta ttgcacttca aagaggtccc gaggaacgac
3540
ggtttaggtc cggacccggt cccgaagcat aacgtgaagt ttctccaggg ctccttgctg
3600
acgtgccccg agctgccacc tccgtagtgg ggctggagaa cggcatccca cggggacctg
3660
tgcacggggc tcgacggtgg aggcatcacc ccgacctctt gccgtagggt gcccctggac
3720
atccggggca cggtgctcac ctaccagtgc gagcctggct acgagctgct aggctccgac
3780
taggccccgt gccacgagtg gatggtcacg ctcggaccga tgctcgacga tccgaggctg
3840
attctcactt gccagtggga cctgtcttgg agcgccgcgc cgcccgcctg ccaaaagatc
3900
taagagtgaa cggtcaccct ggacagaacc tcgcggcgcg gcgggcggac ggttttctag
3960
atgacttgtg ctgaccctgg cgagattgcc aacgggcacc gcaccgcctc ggacgccggc
4020
tactgaacac gactgggacc gctctaacgg ttgcccgtgg cgtggcggag cctgcggccg

4080
ttccccgttg gctcccacgt ccagtaccgc tgcctgccag ggtacagcct cgaggggca
4140
aaggggcaac cgagggtgca ggtcatggcg acggacggtc ccatgtcgga gctcccccgt
4200
gccatgctca cctgctacag ccgggacaca ggcacaccca agtggagcga tagggtcccc
4260
cggtacgagt ggacgatgtc ggccctgtgt ccgtgtgggt tcacctcgct atcccagggg
4320
aaatgcgcct tgaagtacga gccgtgcctg aacccggggg ttcccgagaa tggctaccag
4380
tttacgcgga acttcatgct cggcacggac ttgggccccc aagggctctt accgatggtc
4440
acgctgtaca agcaccacta ccaggcgggc gagtctctgc gcttcttctg ctatgagggc
4500
tgcgacatgt tcgtggtgat ggtccgcccg ctcagagacg cgaagaagac gatactcccg
4560
tttgagctta tcggcgaggt caccatcacc tgtgtgcccg gccacccctc ccagtggacc
4620
aaactcgaat agccgctcca gtggtagtgg acacacgggc cggtggggag ggtcacctgg
4680
agccagcccc cactctgcaa agttgcctat gaggagctcc tggacaaccg aaaactggaa
4740
tcggtcgggg gtgagacgtt tcaacggata ctcctcgagg acctgttggc ttttgacctt
4800
gtgacccaga ccacagatcc atcacggcag ctggaagggg ggaacctggc cctggccatc
4860
cactgggtct ggtgtctagg tagtgccgtc gaccttcccc ccttggaccg ggaccggtag
4920
ctgctgcctc taggcttggt cattgtcctc ggcagtggcg tttacatcta ctacaccaag
4980
gacgacggag atccgaacca gtaacaggag ccgtcaccgc aaatgtagat gatgtggttc
5040
cttcagggaa agtccctttt cggcttctcg ggctcccact cctacagccc catcaccgtg
5100
gaagtccctt tcagggaaaa gccgaagagc ccgagggtga ggatgtcggg gtagtggcac
5160
gagtcggact tcagcaaccc gctgtatgaa gctggggata cgcgggagta tgaagtttcc
5220
ctcagcctga agtcgttggg cgacatactt cgacccctat gcgccctcat acttcaaagg
5280
atctgaaccc caagactaca gctgcaggac ccaggacgcc cctcccctcc tcattcgggc
5340
tagacttggg gttctgatgt cgacgtcctg ggtcctgcgg ggaggggagg agtaagcccg
5400
agagggaaat acgggacccg gtctctgcct cctggctgcc ctcctccctg gctgtgtaaa
5460
tctcccttta tgccctgggc cagagacgga ggaccgacgg gaggagggac cgacacattt
5520
tagtctccct atcccacgag ggggctttga tggccctgga gatcctacag taaataaacc
5580
atcagaggga tagggtgctc ccccgaaact accgggacct ctaggatgtc atttatttgg

5640
agcatcctgc cgcccaaaaa aaaaaaaaaa aaaaatcgta ggacggcggg ttttttttt
5700
ttttttttt
5710

<210> 4
<211> 580
<212> PRT
<213> human

<400> 4

```
Met Gly Thr Pro Arg Ala Gln His Pro Pro Pro Pro Gln Leu Leu Phe
 1               5                  10                  15
Leu Ile Leu Leu Ser Cys Pro Trp Ile Gln Gly Leu Pro Leu Lys Glu
            20                  25                  30
Glu Glu Ile Leu Pro Glu Pro Gly Ser Glu Thr Pro Thr Val Ala Ser
        35                  40                  45
Glu Ala Leu Ala Glu Leu Leu His Gly Ala Leu Leu Arg Arg Gly Pro
    50                  55                  60
Glu Met Gly Tyr Leu Pro Gly Pro Pro Leu Gly Pro Glu Gly Gly Glu
65                  70                  75                  80
Glu Glu Thr Thr Thr Thr Ile Ile Thr Thr Thr Thr Val Thr Thr Thr
                85                  90                  95
Val Thr Ser Pro Val Leu Cys Asn Asn Asn Ile Ser Glu Gly Glu Gly
            100                 105                 110
Tyr Val Glu Ser Pro Asp Leu Gly Ser Pro Val Ser Arg Thr Leu Gly
        115                 120                 125
Leu Leu Asp Cys Thr Tyr Ser Ile His Val Tyr Pro Gly Tyr Gly Ile
    130                 135                 140
Glu Ile Gln Val Gln Thr Leu Asn Leu Ser Gln Glu Glu Glu Leu Leu
145                 150                 155                 160
Val Leu Ala Gly Gly Gly Ser Pro Gly Leu Ala Pro Arg Leu Leu Ala
                165                 170                 175
Asn Ser Ser Met Leu Gly Glu Gly Gln Val Leu Arg Ser Pro Thr Asn
            180                 185                 190
Arg Leu Leu Leu His Phe Gln Ser Pro Arg Val Pro Arg Gly Gly Gly
        195                 200                 205
Phe Arg Ile His Tyr Gln Ala Tyr Leu Leu Ser Cys Gly Phe Pro Pro
    210                 215                 220
Arg Pro Ala His Gly Asp Val Ser Val Thr Asp Leu His Pro Gly Gly
225                 230                 235                 240
Thr Ala Thr Phe His Cys Asp Ser Gly Tyr Gln Leu Gln Gly Glu Glu
                245                 250                 255
Thr Leu Ile Cys Leu Asn Gly Thr Arg Pro Ser Trp Asn Gly Glu Thr
            260                 265                 270
Pro Ser Cys Met Ala Ser Cys Gly Gly Thr Ile His Asn Ala Thr Leu
        275                 280                 285
Gly Arg Ile Val Ser Pro Glu Pro Gly Gly Ala Val Gly Pro Asn Leu
    290                 295                 300
Thr Cys Arg Trp Val Ile Glu Ala Ala Glu Gly Arg Arg Leu His Leu
305                 310                 315                 320
His Phe Glu Arg Val Ser Leu Asp Glu Asp Asn Asp Arg Leu Met Val
```

```
                325                 330                 335
Arg Ser Gly Gly Ser Pro Leu Ser Pro Val Ile Tyr Asp Ser Asp Met
            340                 345                 350
Asp Asp Val Pro Glu Arg Gly Leu Ile Ser Asp Ala Gln Ser Leu Tyr
        355                 360                 365
Val Glu Leu Leu Ser Glu Thr Pro Ala Asn Pro Leu Leu Leu Ser Leu
    370                 375                 380
Arg Phe Glu Ala Phe Glu Glu Asp Arg Cys Phe Ala Pro Phe Leu Ala
385                 390                 395                 400
His Gly Asn Val Thr Thr Thr Asp Pro Glu Tyr Arg Pro Gly Ala Leu
                405                 410                 415
Ala Thr Phe Ser Cys Leu Pro Gly Tyr Ala Leu Glu Pro Pro Gly Pro
            420                 425                 430
Pro Asn Ala Ile Glu Cys Val Asp Pro Thr Glu Pro His Trp Asn Asp
        435                 440                 445
Thr Glu Pro Ala Cys Lys Ala Met Cys Gly Gly Glu Leu Ser Glu Pro
    450                 455                 460
Ala Gly Val Val Leu Ser Pro Asp Trp Pro Gln Ser Tyr Ser Pro Gly
465                 470                 475                 480
Gln Asp Cys Val Trp Gly Val His Val Gln Glu Glu Lys Arg Ile Leu
                485                 490                 495
Leu Gln Val Glu Ile Leu Asn Val Arg Glu Gly Asp Met Leu Thr Leu
            500                 505                 510
Phe Asp Gly Asp Gly Pro Ser Ala Arg Val Leu Ala Gln Leu Arg Gly
        515                 520                 525
Pro Gln Pro Arg Arg Arg Leu Leu Ser Ser Gly Pro Asp Leu Thr Leu
    530                 535                 540
Gln Phe Gln Ala Pro Pro Gly Pro Pro Asn Pro Gly Leu Gly Gln Gly
545                 550                 555                 560
Phe Val Leu His Phe Lys Glu Val Pro Arg Asn Asp Thr Cys Pro Glu
                565                 570                 575
Leu Pro Pro Pro
            580
```

<210> 5
<211> 5598
<212> DNA
<213> human

<400> 5

```
cgcccgggca ggtcggagac ggaggaaagg tggcagccag attacttaga gaggcacaga
    60
gcgggcccgt ccagcctctg cctcctttcc accgtcggtc taatgaatct ctccgtgtct
   120
ggagagagat cggggtgagt cgccatgggg actcccaggg cccagcaccc gccgcctccc
   180
cctctctcta gccccactca gcggtacccc tgagggtccc gggtcgtggg cggcggaggg
   240
cagctgctgt tcctaattct gctgagctgt ccctggatcc agggtctgcc cctgaaggag
   300
gtcgacgaca aggattaaga cgactcgaca gggacctagg tcccagacgg ggacttcctc
   360
```

gaggagatat tgccagagcc tggaagtgag acccccacgg tggcctctga ggccctggct
420
ctcctctata acggtctcgg accttcactc tgggggtgcc accggagact ccgggaccga
480
gaactgcttc atggggccct gctgaggagg ggcccagaga tgggctacct gccagggcct
540
cttgacgaag taccccggga cgactcctcc ccgggtctct acccgatgga cggtcccgga
600
ccccttgggc ctgagggagg agaggaggag acgacgacca ccatcatcac cacgacaact
660
ggggaacccg gactccctcc tctcctcctc tgctgctggt ggtagtagtg gtgctgttga
720
gttaccacta cggtgaccag cccagttctg tgtaataaca acatctccga gggcgaaggg
780
caatggtgat gccactggtc gggtcaagac acattattgt tgtagaggct cccgcttccc
840
tatgtggagt ctccagatct ggggagcccc gtcagccgca ccctggggct cctggactgc
900
atacacctca gaggtctaga cccctcgggg cagtcggcgt gggaccccga ggacctgacg
960
acttacagca tccatgtcta ccctggctac ggcattgaga tccaggtgca gacgctgaac
1020
tgaatgtcgt aggtacagat gggaccgatg ccgtaactct aggtccacgt ctgcgacttg
1080
ctgtcacagg aagaggagct cctggtgctg gctggtgggg gatccccagg cctggccccc
1140
gacagtgtcc ttctcctcga ggaccacgac cgaccacccc ctaggggtcc ggaccggggg
1200
cgactcctgg ccaactcatc catgcttgga gaaggacaag tccttcggag cccaaccaac
1260
gctgaggacc ggttgagtag gtacgaacct cttcctgttc aggaagcctc gggttggttg
1320
cggctgcttc tgcacttcca gagcccacgg gtcccaaggg gcggtggctt caggatccac
1380
gccgacgaag acgtgaaggt ctcgggtgcc cagggttccc cgccaccgaa gtcctaggtg
1440
tatcaggcct acctcctgag ctgtggcttc cctccccggc cggcccatgg ggacgtgagt
1500
atagtccgga tggaggactc gacaccgaag ggaggggccg gccgggtacc cctgcactca
1560
gtgacggacc tgcaccctgg gggcactgcc acctttcact gtgattcggg ctaccagctg
1620
cactgcctgg acgtgggacc cccgtgacgg tggaaagtga cactaagccc gatggtcgac
1680
cagggagagg agaccctcat ctgcctcaat ggcacccggc catcctggaa cggtgaaacc
1740
gtccctctcc tctgggagta gacggagtta ccgtgggccg gtaggacctt gccactttgg
1800
cccagctgca tggcatcctg tggtggcacc atccacaatg ccaccctggg ccgcatcgtg
1860
gggtcgacgt accgtaggac accaccgtgg taggtgttac ggtgggaccc ggcgtagcac
1920

tccccagagc ctgggggagc cgtagggccc aacctcacct gccgttgggt cattgaagca
1980
aggggtctcg gacccccctcg gcatcccggg ttggagtgga cggcaaccca gtaacttcgt
2040
gctgaggggc gccggctgca cctgcacttt gaaagggtct cgctggatga ggacaatgac
2100
cgactccccg cggccgacgt ggacgtgaaa ctttcccaga gcgacctact cctgttactg
2160
cggctgatgg tgcgctcagg gggcagcccc ctatcccccg tgatctatga ttcggacatg
2220
gccgactacc acgcgagtcc cccgtcgggg gatagggggc actagatact aagcctgtac
2280
gacgatgtyc ccgagcgggg tctcatcagt gacgcccagt ccctctacgt ggagctgctg
2340
ctgctacarg ggctcgcccc agagtagtca ctgcgggtca gggagatgca cctcgacgac
2400
tcagagacac ctgccaatcc cctgctgtta agccttcgat ttgaagcctt tgaggaggat
2460
agtctctgtg gacggttagg ggacgacaat tcggaagcta aacttcggaa actcctccta
2520
cgctgcttcg ccccttcct ggcacatgga aatgtcacta ccacggaccc tgagtatcgc
2580
gcgacgaagc gggggaagga ccgtgtacct ttacagtgat ggtgcctggg actcatagcg
2640
ccaggggcac tggcaacctt ctcgtgcctc ccaggatatg ccctggagcc ccctgggccc
2700
ggtccccgtg accgttggaa gagcacggag ggtcctatac gggacctcgg gggacccggg
2760
cccaatgcca tcgaatgtgt ggatcccaca gaaccccact ggaacgacac agagccggcc
2820
gggttacggt agcttacaca cctagggtgt cttggggtga ccttgctgtg tctcggccgg
2880
tgcaaagcca tgtgtggagg ggagctgtcg gaaccagctg gcgtggtcct ctctcccgac
2940
acgtttcggt acacacctcc cctcgacagc cttggtcgac cgcaccagga gagagggctg
3000
tggccccaga gctatagccc gggccaagac tgcgtgtggg gcgtgcacgt ccaggaagag
3060
accggggtct cgatatcggg cccggttctg acgcacaccc cgcacgtgca ggtccttctc
3120
aagcgcatct tgctccaagt tgagatattg aatgtgcggg aaggggacat gctgacgctg
3180
ttcgcgtaga acgaggttca actctataac ttacacgccc ttcccctgta cgactgcgac
3240
ttcgacgggg acggtcccag cgcccgagtc ttggcccagc tgcggggacc tcagccgcgc
3300
aagctgcccc tgccagggtc gcgggctcag aaccgggtcg acgcccctgg agtcggcgcg
3360
cgccgccttc tctcctctgg gcccgacctc acactgcagt ttcaggcacc gcccgggccc
3420
gcggcggaag agaggagacc cgggctggag tgtgacgtca aagtccgtgg cgggcccggg
3480

ccaaatccag gcctgggcca gggcttcgta ttgcacttca aagaggtccc gaggaacgac
3540
ggtttaggtc cggacccggt cccgaagcat aacgtgaagt ttctccaggg ctccttgctg
3600
acgtgccccg agctgccacc tccggagtgg ggctggagaa cggcatccca cggggacctg
3660
tgcacggggc tcgacggtgg aggcctcacc ccgacctctt gccgtagggt gcccctggac
3720
atccggggca cggtgctcac ctaccagtgc gagcctggct acgagctgct aggctccgac
3780
taggccccgt gccacgagtg gatggtcacg ctcggaccga tgctcgacga tccgaggctg
3840
attctcactt gccagtggga cctgtcttgg agcgccgcgc cgcccgcctg ccaaaagatc
3900
taagagtgaa cggtcaccct ggacagaacc tcgcggcgcg gcgggcggac ggttttctag
3960
atgacttgtg ctgaccctgg cgagattgcc aacgggcacc gcaccgcctc ggacgccggc
4020
tactgaacac gactgggacc gctctaacgg ttgcccgtgg cgtggcggag cctgcggccg
4080
ttccccgttg gctcccacgt ccagtaccgc tgcctgccag ggtacagcct cgaggggggca
4140
aaggggcaac cgagggtgca ggtcatggcg acggacggtc ccatgtcgga gctcccccgt
4200
gccatgctca cctgctacag ccgggacaca ggcacaccca agtggagcga tagggtcccc
4260
cggtacgagt ggacgatgtc ggccctgtgt ccgtgtgggt tcacctcgct atcccagggg
4320
aaatgcgcct tgaagtacga gccgtgcctg aacccggggg ttcccgagaa tggctaccag
4380
tttacgcgga acttcatgct cggcacggac ttgggccccc aagggctctt accgatggtc
4440
acgctgtaca agcaccacta ccaggcgggc gagtctctgc gcttcttctg ctatgagggc
4500
tgcgacatgt tcgtggtgat ggtccgcccg ctcagagacg cgaagaagac gatactcccg
4560
tttgagctta tcggcgaggt caccatcacc tgtgtgcccg gccacccctc ccagtggacc
4620
aaactcgaat agccgctcca gtggtagtgg acacacgggc cggtggggag ggtcacctgg
4680
agccagcccc cactctgcaa agtgacccag accacagatc catcacggca gctggaaggg
4740
tcggtcgggg gtgagacgtt tcactgggtc tggtgtctag gtagtgccgt cgaccttccc
4800
gggaacctgg ccctggccat cctgctgcct ctaggcttgg tcattgtcct cggcagtggc
4860
cccttggacc gggaccggta ggacgacgga gatccgaacc agtaacagga gccgtcaccg
4920
gtttacatct actacaccaa gcttcaggga aagtcccttt tcggcttctc gggctcccac
4980
caaatgtaga tgatgtggtt cgaagtccct ttcagggaaa agccgaagag cccgagggtg
5040

tcctacagcc ccatcaccgt ggagtcggac ttcagcaacc cgctgtatga agctggggat
5100
aggatgtcgg ggtagtggca cctcagcctg aagtcgttgg gcgacatact tcgaccccta
5160
acgcgggagt atgaagtttc catctgaacc ccaagactac agctgcagga cccaggacgc
5220
tgcgccctca tacttcaaag gtagacttgg ggttctgatg tcgacgtcct gggtcctgcg
5280
ccctcccctc ctcattcggg cagagggaaa tacgggaccc ggtctctgcc tcctggctgc
5340
gggaggggag gagtaagccc gtctcccttt atgccctggg ccagagacgg aggaccgacg
5400
cctcctccct ggctgtgtaa atagtctccc tatcccacga gggggctttg atggccctgg
5460
ggaggaggga ccgacacatt tatcagaggg atagggtgct cccccgaaac taccgggacc
5520
agatcctaca gtaaataaac cagcatcctg ccgcccaaat ctaggatgtc atttatttgg
5580
tcgtaggacg gcgggttt
5598

<210> 6
<211> 840
<212> PRT
<213> human

<400> 6

Met Gly Thr Pro Arg Ala Gln His Pro Pro Pro Pro Gln Leu Leu Phe
1 5 10 15
Leu Ile Leu Leu Ser Cys Pro Trp Ile Gln Gly Leu Pro Leu Lys Glu
20 25 30
Glu Glu Ile Leu Pro Glu Pro Gly Ser Glu Thr Pro Thr Val Ala Ser
35 40 45
Glu Ala Leu Ala Glu Leu Leu His Gly Ala Leu Leu Arg Arg Gly Pro
50 55 60
Glu Met Gly Tyr Leu Pro Gly Pro Pro Leu Gly Pro Glu Gly Gly Glu
65 70 75 80
Glu Glu Thr Thr Thr Thr Ile Ile Thr Thr Thr Thr Val Thr Thr Thr
85 90 95
Val Thr Ser Pro Val Leu Cys Asn Asn Asn Ile Ser Glu Gly Glu Gly
100 105 110
Tyr Val Glu Ser Pro Asp Leu Gly Ser Pro Val Ser Arg Thr Leu Gly
115 120 125
Leu Leu Asp Cys Thr Tyr Ser Ile His Val Tyr Pro Gly Tyr Gly Ile
130 135 140
Glu Ile Gln Val Gln Thr Leu Asn Leu Ser Gln Glu Glu Glu Leu Leu
145 150 155 160
Val Leu Ala Gly Gly Gly Ser Pro Gly Leu Ala Pro Arg Leu Leu Ala
165 170 175
Asn Ser Ser Met Leu Gly Glu Gly Gln Val Leu Arg Ser Pro Thr Asn
180 185 190
Arg Leu Leu Leu His Phe Gln Ser Pro Arg Val Pro Arg Gly Gly Gly
195 200 205

Phe Arg Ile His Tyr Gln Ala Tyr Leu Leu Ser Cys Gly Phe Pro Pro
210 215 220
Arg Pro Ala His Gly Asp Val Ser Val Thr Asp Leu His Pro Gly Gly
225 230 235 240
Thr Ala Thr Phe His Cys Asp Ser Gly Tyr Gln Leu Gln Gly Glu Glu
245 250 255
Thr Leu Ile Cys Leu Asn Gly Thr Arg Pro Ser Trp Asn Gly Glu Thr
260 265 270
Pro Ser Cys Met Ala Ser Cys Gly Gly Thr Ile His Asn Ala Thr Leu
275 280 285
Gly Arg Ile Val Ser Pro Glu Pro Gly Gly Ala Val Gly Pro Asn Leu
290 295 300
Thr Cys Arg Trp Val Ile Glu Ala Ala Glu Gly Arg Arg Leu His Leu
305 310 315 320
His Phe Glu Arg Val Ser Leu Asp Glu Asp Asn Asp Arg Leu Met Val
325 330 335
Arg Ser Gly Gly Ser Pro Leu Ser Pro Val Ile Tyr Asp Ser Asp Met
340 345 350
Asp Asp Val Pro Glu Arg Gly Leu Ile Ser Asp Ala Gln Ser Leu Tyr
355 360 365
Val Glu Leu Leu Ser Glu Thr Pro Ala Asn Pro Leu Leu Leu Ser Leu
370 375 380
Arg Phe Glu Ala Phe Glu Glu Asp Arg Cys Phe Ala Pro Phe Leu Ala
385 390 395 400
His Gly Asn Val Thr Thr Thr Asp Pro Glu Tyr Arg Pro Gly Ala Leu
405 410 415
Ala Thr Phe Ser Cys Leu Pro Gly Tyr Ala Leu Glu Pro Pro Gly Pro
420 425 430
Pro Asn Ala Ile Glu Cys Val Asp Pro Thr Glu Pro His Trp Asn Asp
435 440 445
Thr Glu Pro Ala Cys Lys Ala Met Cys Gly Gly Glu Leu Ser Glu Pro
450 455 460
Ala Gly Val Val Leu Ser Pro Asp Trp Pro Gln Ser Tyr Ser Pro Gly
465 470 475 480
Gln Asp Cys Val Trp Gly Val His Val Gln Glu Glu Lys Arg Ile Leu
485 490 495
Leu Gln Val Glu Ile Leu Asn Val Arg Glu Gly Asp Met Leu Thr Leu
500 505 510
Phe Asp Gly Asp Gly Pro Ser Ala Arg Val Leu Ala Gln Leu Arg Gly
515 520 525
Pro Gln Pro Arg Arg Arg Leu Leu Ser Ser Gly Pro Asp Leu Thr Leu
530 535 540
Gln Phe Gln Ala Pro Pro Gly Pro Pro Asn Pro Gly Leu Gly Gln Gly
545 550 555 560
Phe Val Leu His Phe Lys Glu Val Pro Arg Asn Asp Thr Cys Pro Glu
565 570 575
Leu Pro Pro Pro Glu Trp Gly Trp Arg Thr Ala Ser His Gly Asp Leu
580 585 590
Ile Arg Gly Thr Val Leu Thr Tyr Gln Cys Glu Pro Gly Tyr Glu Leu
595 600 605
Leu Gly Ser Asp Ile Leu Thr Cys Gln Trp Asp Leu Ser Trp Ser Ala
610 615 620

```
Ala Pro Pro Ala Cys Gln Lys Ile Met Thr Cys Ala Asp Pro Gly Glu
625                 630                 635                 640
Ile Ala Asn Gly His Arg Thr Ala Ser Asp Ala Gly Phe Pro Val Gly
                645                 650                 655
Ser His Val Gln Tyr Arg Cys Leu Pro Gly Tyr Ser Leu Glu Gly Ala
            660                 665                 670
Ala Met Leu Thr Cys Tyr Ser Arg Asp Thr Gly Thr Pro Lys Trp Ser
        675                 680                 685
Asp Arg Val Pro Lys Cys Ala Leu Lys Tyr Glu Pro Cys Leu Asn Pro
    690                 695                 700
Gly Val Pro Glu Asn Gly Tyr Gln Thr Leu Tyr Lys His His Tyr Gln
705                 710                 715                 720
Ala Gly Glu Ser Leu Arg Phe Phe Cys Tyr Glu Gly Phe Glu Leu Ile
                725                 730                 735
Gly Glu Val Thr Ile Thr Cys Val Pro Gly His Pro Ser Gln Trp Thr
            740                 745                 750
Ser Gln Pro Pro Leu Cys Lys Val Thr Gln Thr Thr Asp Pro Ser Arg
        755                 760                 765
Gln Leu Glu Gly Gly Asn Leu Ala Leu Ala Ile Leu Leu Pro Leu Gly
    770                 775                 780
Leu Val Ile Val Leu Gly Ser Gly Val Tyr Ile Tyr Tyr Thr Lys Leu
785                 790                 795                 800
Gln Gly Lys Ser Leu Phe Gly Phe Ser Gly Ser His Ser Tyr Ser Pro
                805                 810                 815
Ile Thr Val Glu Ser Asp Phe Ser Asn Pro Leu Tyr Glu Ala Gly Asp
            820                 825                 830
Thr Arg Glu Tyr Glu Val Ser Ile
        835                 840
```

INTERNATIONAL SEARCH REPORT

Inte[illegible]onal Application No
PCT/DK 00/00556

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/00 //C12N15/12,C07K14/705

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)
IPC 7 C07K C12N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YU W ET AL: "Large-Scale Concatenation cDNA Sequencing" GENOME RESEARCH, vol. 7, no. 4, April 1997 (1997-04), pages 353-358, XP002901489 Appendix ---	15-33, 35-40
A	KEIKO SHIMIZU-NISHIKAWA ET AL: "Cloning and Characterization of Seizure-Related Gene, Sez-6" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 216, no. 1, 2 November 1995 (1995-11-02), pages 382-389, XP002901490 Appendix -----	1-72

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

☐ Patent family members are listed in annex.

° Special categories of cited documents :

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

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 January 2001	08.03.01

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

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national application No.
PCT/DK 00/00556

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

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

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

2. ☒ Claims Nos.: 1, 18, 41, 48, 49, 69-71
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

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

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

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

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

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

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

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

Remark on Protest

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

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 1, 18, 41, 48, 49, 69-71

Present claims 1, 18 and 41 relate to an extremely large number of possible molecules. Support within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the molecules claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible.

Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts exemplified on pages 52-97 of the description.

Furthermore, claims 48 and 49 relate to methods defined by reference to desirable characteristics, namely the detection of antibodies or ligands. The claims cover all methods having these characteristics, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such methods. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the method by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible.

Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to the methods exemplified on pages 52-97 of the description.

Claims 69-71 relate to a method of treatment of the human or animal body. Thus, the International Search Authority is not required to carry out an international search for these claims (Rule 39.1(iv)). Nevertheless, a search has been executed based on the alleged effects of the compounds in claims 69-71.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.