REPUBLIC OF SOUTH AFRICA
PATENTS ACT, 1978

FORM P.8
(to be lodged in duplicate)

A&A Ref: **146379**

PUBLICATION PARTICULARS AND ABSTRACT
(Section 32(3)(a) - Regulations 22(1)(g) and 31)

21	01	PATENT APPLICATION NO	22	LODGING DATE	43	ACCEPTANCE DATE
		2002/6550		15 August 2002		19-12-2003

51	INTERNATIONAL CLASSIFICATION	NOT FOR PUBLICATION

C07K C12N A61K G01N A61P CLASSIFIED BY: **ISA**

71	FULL NAME(S) OF APPLICANT(S)

Aventis Pharma S.A.
Institut National de la Sante et de la Recherche Medicale (INSERM)

72	FULL NAME(S) OF INVENTOR(S)

KOUTNIKOVA, Hana
FOURNIER, Alain
PRADES, Catherine
ROSIER-MONTUS, Marie-Françoise
BRICE, Alexis
PRADIER, Laurent
ARNOULD-REGUIGNE, Isabelle
CORTI, Olga

EARLIEST PRIORITY CLAIMED	COUNTRY		NUMBER		DATE	
	33	**FR**	31	**00/01,980**	32	**17 February 2000**

NOTE: The country must be indicated by its International Abbreviation - see schedule 4 of the Regulations

54	TITLE OF INVENTION

Compositions useful for regulating parkin gene activity

57	ABSTRACT (NOT MORE THAN 150 WORDS)

NUMBER OF SHEETS	111

The sheet(s) containing the abstract is/are attached.

If no classification is furnished, Form P.9 should accompany this form.
The figure of the drawing to which the abstract refers is attached.

WO 01/608

(57) **Abstract:** The invention concerns novel compounds and their uses, in particular pharmaceutical, diagnostic or as pharmacological targets. More particularly, the invention concerns a novel protein, called PAP1, and novel peptides and compounds capable of modulating at least partially parkin gene activity.

COMPOSITIONS WHICH CAN BE USED FOR REGULATING THE ACTIVITY OF PARKIN

The present invention relates to compositions and methods which can be used for regulating the activity of parkin. It relates in particular to a novel protein, referred to as PAP1, which is a partner of parkin, as well as to the peptides or polypeptides which are derived from or are homologous to this protein. It also relates to compounds which are capable of modulating, at least partially, the activity of parkin, in particular of interfering with the interaction between parkin and PAP1. The present invention can be used in the therapeutic or diagnostic domains, or for forming pharmacological targets which enable the development of novel medicaments.

The parkin gene is mutated in certain familial forms (autosomal recessive juvenile) of Parkinson's disease (Kitada *et al.*, 1998). Parkinson's disease (Lewy, 1912) is one of the most common neurodegenerative diseases, affecting more than 1% of the population over 55 years old. Patients suffering from this disease have neurological disorders which are grouped together under the term "Parkinsonian syndrome", which is characterized by rigidity, bradykinesia, and shaking while resting. These symptoms are the consequence of a degeneration of the

dopaminergic neurons of the substantia nigra of the brain.

Most cases with a Parkinson's disease do not have a familial history. However, familial cases do exist, of which certain correspond to a monogenic form of the disease. At the current time, only three different genes have been identified in certain rare hereditary forms. The first form corresponds to an autosomal dominant form, in which the gene responsible encodes alpha Synuclein (Polymeropoulos *et al.*, 1997). This protein is an abundant constituent of the intracytoplasmic inclusions, termed Lewy bodies, which are used as a marker for Parkinson's disease (Lewy, 1912). The second form, also autosomal dominant, is linked to a mutation in a gene which encodes a hydrolase termed ubiquitin carboxy-terminal hydrolase L1 (Leroy *et al.*, 1998). This enzyme is assumed to hydrolyze ubiquitin polymers or conjugates into ubiquitin monomers. The third form differs from the previous forms in that it has an autosomal recessive transmission and a commencement often before 40 years of age, as well as an absence of Lewy bodies. These diseases respond more favorably to levodopa, a dopamine precursor which is used as treatment for Parkinson's disease. The gene involved in this form encodes a novel protein which is termed parkin (Kitada *et al.*, 1998).

The parkin gene consists of 12 exons which cover a genomic region of more than 500,000 base pairs

on chromosome 6 (6q25.2-q27). At the current time, two major types of mutation of this gene, which are at the origin of the disease, are known; either deletions of variable size in the region which covers exons 2 to 9, or point mutations which produce the premature appearance of a stop codon or the change of an amino acid (Kitada *et al.*, 1998; Abbas *et al.*, 1999; Lucking *et al.*, 1998; Hattori *et al.*, 1998). The nature of these mutations and the autosomal recessive method of transmission suggest a loss of function of the parkin, which leads to Parkinson's disease.

This gene is expressed in a large number of tissues and in particular in the substantia nigra. Several transcripts corresponding to this gene exist, which originate from different alternative splicing (Kitada *et al.*, 1998; Sunada *et al.*, 1998). In the brain, two types of messenger RNA are found, of which one lacks the portion corresponding to exon 5. In leukocytes, parkin messenger RNAs which do not contain the region encoding exons 3, 4 and 5 have been identified. The longest of the parkin messenger RNAs, which is present in the brain, contains 2960 bases and encodes a protein of 465 amino acids.

This protein has a weak homology, in its N-termimal portion, with ubiquitin. Its C-terminal half contains two ring finger motifs, separated by an IBR (In Between Ring) domain, which correspond to a cysteine-rich region and which are able to bind metals,

like the zing finger domains (Morett, 1999). It has been shown, by immunocytochemistry, that parkin is located in the cytoplasm and the Golgi apparatus of neurons of the substantia nigra which contain melanin (Shimura *et al.*, 1999). In addition, this protein is present in certain Lewy bodies of Parkinsonians. The cellular function of parkin has not yet been demonstrated, but it might play a transporter role in synaptic vesicles, in the maturation or degradation of proteins, and in the control of cellular growth, differentiation or development. In the autosomal recessive juvenile forms, parkin is absent, which thus confirms that the loss of this function is responsible for the disease.

The elucidation of the exact role of the parkin protein in the process of degeneration of the dopaminergic neurons thus constitutes a major asset for the understanding of and the therapeutical approach to Parkinson's disease, and more generally diseases of the central nervous system.

The present invention lies in the identification of a partner of parkin, which interacts with this protein under physiological conditions. This partner represents a novel pharmacological target for manufacturing or investigating compounds which are capable of modulating the activity of parkin, in particular its activity on the degeneration of dopaminergic neurons and/or the development of nervous

pathologies. This protein, the antibodies, the corresponding nucleic acids, as well as the specific probes or primers, can also be used for detecting or assaying the proteins in biological samples, in particular nervous tissue samples. These proteins or nucleic acids can also be used in therapeutic approaches, for modulating the activity of parkin and any compound according to the invention which is capable of modulating the interaction between parkin and the polypeptides of the invention.

The present invention results more particularly from the demonstration by the applicant of a novel human protein, referred to as PAP1 (Parkin Associated Protein 1) or LY111, which interacts with parkin. The PAP1 protein (sequence SEQ ID NO. 1 or 2) shows a certain homology with synaptotagmins and is capable of interacting more particularly with the central region of parkin (represented on the sequence SEQ ID NO. 3 or 4). The PAP1 protein has also been cloned, sequenced and characterized using various tissues of human origin, in particular lung (SEQ ID NOS. 12 and 13) and brain (SEQ ID NOS. 42 and 43) as have short forms, corresponding to splicing variants (SEQ ID NOS. 14, 15, 44 and 45).

The present invention also results from the identification and from the characterization of specific regions of the PAP1 protein, which are implicated in the modulation of the function of parkin.

The demonstration of the existence of this protein and of regions which are implicated in its function makes it possible in particular to prepare novel compounds and/or compositions which can be used as pharmaceutical agents, and to develop industrial methods of screening such compounds.

A first subject of the invention thus relates to compounds which are capable of modulating, at least partially, the interaction between the PAP1 protein (or homologs thereof) and parkin (in particular human parkin), or of interfering with the interaction between these proteins.

Another subject of the invention lies in the PAP1 protein and fragments, derivatives and homologs thereof.

Another aspect of the invention lies in a nucleic acid which encodes the PAP1 protein or fragments, derivatives or homologs thereof, as well as any vector which comprises such a nucleic acid, any recombinant cell which contains such a nucleic acid or vector and any nonhuman mammal comprising in its cells such a nucleic acid.

The invention also relates to antibodies which are capable of binding the PAP1 protein and fragments, derivatives and homologs thereof, in particular polyclonal or monoclonal antibodies, more preferably antibodies which are capable of binding the

PAP1 protein and of inhibiting, at least partially, its interaction with parkin.

Another aspect of the invention relates to nucleotide probes or primers, which are specific for PAP1 and which can be used for detecting or amplifying the pap1 gene, or a region of this gene, in any biological sample.

The invention also relates to pharmaceutical compositions, methods for detecting genetic abnormalities, methods for producing polypeptides as defined above and methods for screening or for characterizing active compounds.

As indicated above, a first aspect of the invention lies in a compound which is capable of interfering, at least partially, with the interaction between the PAP1 protein (or homologs thereof) and parkin.

For the purposes of the present invention, the name PAP1 protein refers to the protein per se, as well as to all homologous forms thereof. "Homologous form" is intended to refer to any protein which is equivalent to the protein under consideration, of varied cellular origin and in particular derived from cells of human origin, or from other organisms, and which possesses an activity of the same type. Such homologs also comprise natural variants of the PAP1 protein of sequence SEQ ID NO 2, in particular polymorphic or splicing variants. Such homologs can be

obtained by the experiments of hybridization between the coding nucleic acids (in particular the nucleic acid of sequence SEQ ID NO. 1). For the purposes of the invention, a sequence of this type only has to have a significant percentage of identity to lead to a physiological behavior which is comparable to that of the PAP1 protein as claimed. The expression "significant percentage of identity" is intended to mean a percentage of at least 60%, preferentially of 80%, more preferentially of 90%, and even more preferentially of 95%. In this regard, variants and/or homologues of the sequence SEQ ID NO. 2 are described in the sequences SEQ ID NOS. 13, 15, 43 and 45, identified using tissues of human origin. The name PAP1 therefore also encompasses these polypeptides.

For the purpose of the present invention, the "percentage of identity" between two nucleotide or amino acid sequences can be determined by comparing two optimally aligned sequences through a window of comparison.

The part of the nucleotide or polypeptide sequence within the window of comparison may thus comprise additions or deletions (for example gaps) compared with the reference sequence (which does not comprise these additions or these deletions) so as to obtain an optimal alignment of the two sequences.

The percentage is calculated by determining the number of positions at which an identical nucleic

acid base or amino acid residue is observed for the two sequences (nucleic acid or peptide sequences) compared, then dividing the number of positions at which there is identity between the two bases or amino acid residues by the total number of positions in the window of comparison, and then multiplying the result by 100 in order to obtain the percentage of sequence identity.

The optimal alignment of the sequences for the comparison may be produced on a computer, using known algorithms contained in the WISCONSIN GENETICS SOFTARE PACKAGE, GENETICS COMPUTER GROUP (GCG), 575 Science Doctor, Madison, WISCONSIN.

By way of illustration, the percentage of sequence identity may be produced using the BLAST program (versions BLAST 1.4.9 of March 1996, BLAST 2.0.4 of February 1998 and BLAST 2.0.6 of September 1998), using exclusively the default parameters (Altschul et al., *J. Mol. Biol.*, (1990) **215**: 403-410; Altschul et al, *Nucleic Acids Res.* (1997) **25**: 3389-3402). Blast searches for sequences which are similar/homologous to a "request" reference sequence, using the algorithm of Altschul et al. (above). The request sequence and the databases used may be peptide or nucleic acid based, any combination being possible.

The interference of a compound according to the invention can reveal itself in various ways. Thus, the compound can slow, inhibit or stimulate, at least partially, the interaction between the PAP1 protein, or

a homologous form thereof, and parkin. Preferably, they are compounds which are capable of modulating this interaction in vitro, for example in a double-hybrid type system or in any acellular system for detecting an interaction between two polypeptides. The compounds according to the invention are preferably compounds which are capable of modulating, at least partially, this interaction, preferably of increasing or inhibiting this reaction by at least 20%, more preferably by at least 50%, with respect to a control in the absence of the compound.

In a particular embodiment, they are compounds which are capable of interfering with the interaction between the region of parkin which is represented on the sequence SEQ ID NO. 4 and the region of the PAP1 protein which is represented on the sequence SEQ ID NOS. 2, 13, 15, 43 or 45.

According to a particular embodiment of the invention, the compounds are capable of binding at the level of the domain of interaction between the PAP1 protein, or a homologous form thereof, and parkin.

The compounds according to the present invention can be varied in nature and in origin. In particular, they can be compounds of peptide, nucleic acid (i.e. comprising a string of bases, in particular a DNA or an RNA molecule), lipid or saccharide type, an antibody, etc. and, more generally, any organic or inorganic molecule.

According to a first variant, the compounds of the invention are peptide in nature. The term "peptide" refers to any molecule comprising a string of amino acids, such as for example a peptide, a polypeptide, a protein or an antibody (or antibody fragment or derivative), which if necessary is modified or combined with other compounds or chemical groups. In this respect, the term "peptide" refers more specifically to a molecule comprising a string of at most 50 amino acids, more preferably of at most 40 amino acids. A polypeptide (or a protein) preferably comprises from 50 to 500 amino acids, or more.

According to a first preferred embodiment, the compounds of the invention are peptide compounds comprising all or part of the peptide sequence SEQ ID NO. 2 or a derivative thereof, in particular all or part of the peptide sequence SEQ ID NOS. 13, 15, 43 or 45 or derivatives of these sequences, more particularly of the PAP1 protein, which comprises the sequence SEQ ID NOS. 2, 13, 15, 43 or 45.

For the purposes of the present invention, the term "derivative" refers to any sequence which differs from the sequence under consideration because of a degeneracy of the genetic code, which is obtained by one or more modifications of genetic and/or chemical nature, as well as any peptide which is encoded by a sequence which hybridizes with the nucleic acid sequence SEQ ID NO. 1 or a fragment of this sequence,

for example with the nucleic acid sequence SEQ ID NOS. 12, 14, 42 or 44 or a fragment of these sequences, and which is capable of interfering with the interaction between the PAP1 protein, or a homolog thereof, and parkin. "Modification of genetic and/or chemical nature" can be intended to mean any mutation, substitution, deletion, addition and/or modification of one or more residues. The term "derivative" also comprises the sequences which are homologous to the sequence under consideration, which are derived from other cellular sources and in particular cells of human origin, or from other organisms, and which possess an activity of the same type. Such homologous sequences can be obtained by hybridization experiments. The hybridizations can be carried out with nucleic acid libraries, using the native sequence or a fragment of this sequence as probe, under varied conditions of hybridization (Maniatis *et al.*, 1989). Moreover, the term "fragment" or "part" refers to any portion of the molecule under consideration, which comprises at least 5 consecutive residues, preferably at least 9 consecutive residues, even more preferably at least 15 consecutive residues. Typical fragments can comprise at least 25 consecutive residues.

Such derivatives or fragments can be generated with different aims, such as in particular that of increasing their therapeutic effectiveness or of reducing their side effects, or that of conferring

novel pharmacokinetic and/or biological properties thereon.

As a peptide which is derived from the PAP1 protein and from the homologous forms, mention may be made in particular of any peptide which is capable of interacting with parkin, but which bears an effector region which has been made nonfunctional. Such peptides can be obtained by deletion, mutation or disruption of this effector region on the PAP1 protein and homologous forms. Such modifications can be carried out for example by *in vitro* mutagenesis, by introducing additional elements or synthetic sequences, or by deletions or substitutions of the original elements. When such a derivative as defined above is prepared, its activity as partial inhibitor of the binding of the PAP1 protein and of the homologous forms, on its binding site on parkin, can be demonstrated. Any technique known to persons skilled in the art can of course be used for this.

They can also be fragments of the sequences indicated above. Such fragments can be generated in various ways. In particular they can be synthesized chemically, on the basis of the sequences given in the present application, using the peptide synthesizers known to persons skilled in the art. They can also be synthesized genetically, by expression in a host cell of a nucleotide sequence which encodes the desired peptide. In this case, the nucleotide sequence can be

prepared chemically using an oligonucleotide synthesizer, on the basis of the peptide sequence given in the present application and of the genetic code. The nucleotide sequence can also be prepared from sequences given in the present application, by enzymatic cleavage, ligation, cloning, etc., according to the techniques known to persons skilled in the art, or by screening DNA libraries with probes which are developed from these sequences.

Moreover, the peptides of the invention, i.e. which are capable of modulating, at least partially, the interaction between the PAP1 protein, and homologous forms, and parkin, can also be peptides which have a sequence corresponding to the site of interaction of the PAP1 protein and of the homologous forms on parkin.

Other peptides according to the invention are peptides which are capable of competing with the peptides defined above for the interaction with their cellular target. Such peptides can be synthesized in particular on the basis of the sequence of the peptide under consideration, and their capacity for competing with the peptides defined above can be determined.

A specific subject of the present invention relates to the PAP1 protein. It is more particularly the PAP1 protein comprising the sequence SEQ ID NO. 2 or a fragment or derivative of this sequence, for

example the PAP1 protein with the sequences SEQ ID NOS. 13, 15, 43 or 45 or with fragments of these sequences.

Another subject of the invention lies in polyclonal or monoclonal antibodies or antibody fragments or derivatives, which are directed against the polypeptide as defined above. Such antibodies can be generated by methods known to persons skilled in the art. In particular, these antibodies can be prepared by immunizing an animal against a peptide compound of the invention (in particular a polypeptide or a peptide comprising all or part of the sequence SEQ ID NO. 2), sampling the blood and isolating the antibodies. These antibodies can also be generated by preparing hybridomas according to the techniques known to persons skilled in the art.

More preferably, the antibodies or antibody fragments of the invention have the capacity to modulate, at least partially, the interaction of the claimed peptides with parkin.

Moreover, these antibodies can also be used for detecting and/or assaying the expression of PAP1 in biological samples and, consequently, for providing information on its activation state.

The antibody fragments or derivatives are for example Fab or Fab'2 fragments, single-chain antibodies (ScFv), etc. They are in particular any fragment or derivative which conserves the antigenic specificity of the antibodies from which they are derived.

The antibodies according to the invention are more preferably capable of binding the PAP1 proteins which comprise the sequence SEQ ID NOS. 2, 13, 15, 43 or 45, in particular the region of this protein which is involved in the interaction with parkin. These antibodies (or fragments or derivatives) are more preferably capable of binding an epitope which is present in the sequence between residues 1 and 344 of the sequence SEQ ID NO. 2.

The invention also relates to compounds which are not peptide or not exclusively peptide, which can be used as a pharmaceutical agent. It is in fact possible, from the active protein motifs described in the present application, to prepare molecules which are modulators of the activity of PAP1, which are not exclusively peptide, and which are compatible with pharmaceutical use, in particular by duplicating the active motifs of the peptides with a structure which is not a peptide, or which is not of exclusively peptide nature.

A subject of the present invention is also any nucleic acid which encodes a peptide compound according to the invention. It can be in particular a nucleic acid comprising all or part of the sequence SEQ ID NOS. 1, 12, 14, 42 or 44, or a derivative thereof. "Derived sequence" is intended to mean, for the purposes of the present invention, any sequence which hybridizes with the sequence which is presented in SEQ

ID No. 1, or with a fragment of this sequence, and which encodes a peptide compound according to the invention, as well as the sequences which result from the latter by degeneracy of the genetic code. Nucleic acids according to the invention comprise, for example, all or part of the nucleic acid sequence SEQ ID NOS. 12, 14, 42 or 44.

The present invention also relates to sequences which have a significant percentage of identity with the sequence presented in SEQ ID NO. 1 or with a fragment of this sequence, and which encode a peptide compound having a physiological behavior similar to that of the PAP1 protein. The expression "significant percentage of identity" is intended to mean a percentage of at least 60%, preferentially of 80%, more preferentially of 90%, and even more preferentially of 95%.

The various nucleotide sequences of the invention may or may not be of artificial origin. They can be genomic, cDNA or RNA sequences, hybrid sequences or synthetic or semi-synthetic sequences. These sequences can be obtained either by screening DNA libraries (cDNA library, genomic DNA library), by chemical synthesis, by mixed methods which include the chemical or enzymatic modification of sequences which are obtained by screening of libraries, or by searching for homology in nucleic acid or protein databases. The abovementioned hybridization is preferably carried out

under the conditions described by Sambrook et al. (1989, pages 9.52-9.55).

It is advantageously carried out under high stringency hybridization conditions. For the purpose of the present invention, the expression "high stringency hybridization conditions" will be intended to mean the following conditions:

1 - Membrane competition and PREHYBRIDIZATION:

- Mix: 40 µl of salmon sperm DNA (10 mg/ml)
  + 40 µl of human placental DNA (10 mg/ml)
- Denature for 5 min at 96°C, and then plunge the mixture into ice.
- Remove the 2X SSC buffer and pour 4 ml of formamide mix into the hybridization tube containing the membranes.
- Add the mixture of the two denatured DNAs.
- Incubate at 42°C for 5 to 6 hours, with rotation.

2 - Labeled probe competition:

- Add 10 to 50 µl of Cot 1 DNA, depending on the amount of nonspecific hybridization, to the labeled and purified probe.
- Denature for 7 to 10 min at 95°C.
- Incubate at 65°C for 2 to 5 hours.

3 - Hybridization:

- Remove the prehybridization mix.

- Mix 40 µl of salmon sperm DNA + 40 µl of human placental DNA; denature for 5 min at 96°C and then plunge into ice.

- Add 4 ml of formamide mix, the mixture of the two DNAs and the labeled probe/denatured Cot 1 DNA to the hybridization tube.

- Incubate for 15 to 20 hours at 42°C, with rotation.

4 - Washing:

- Wash once at ambient temperature in 2X SSC, to rinse.

- Wash twice 5 minutes at ambient temperature in 2X SSC and 0.1% SDS,

- wash twice 15 minutes in 0.1X SSC and 0.1% SDS at 65°C.

Wrap the membranes in Saran and expose.

The hybridization conditions described above are suitable for hybridizing, under high stringency conditions, a molecule of nucleic acid with a variable length of 20 nucleotides to several hundreds of nucleotides.

It goes without saying that the hybridization conditions described above may be adjusted as a function of the length of the nucleic acid for which the hybridization is sought or of the type of labeling chosen, according to techniques known to those skilled in the art.

The suitable hybridization conditions may, for example, be adjusted according to the teaching contained in the work by HAMES and HIGGINS (1985)

(Nucleic acid Hybridization: a practical approach, Hames and Higgins Ed., IRL Press, Oxford) or in the work by F. AUSUBEL et al (1999) (Currents Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.).

For the purposes of the invention, a particular nucleic acid encodes a polypeptide comprising the sequence SEQ ID NO. 2 or a fragment or derivative of this sequence, in particular the human PAP1 protein. It is advantageously a nucleic acid comprising the nucleic acid sequence SEQ ID NOS. 1, 12, 14, 42 or 44.

Such nucleic acids can be used for producing the peptide compounds of the invention. The present application thus relates to a method for preparing such peptide compounds, according to which a cell which contains a nucleic acid according to the invention is cultured under conditions for expressing said nucleic acid, and the peptide compound produced is recovered. In this case, the portion which encodes said peptide compound is generally placed under the control of signals which allow its expression in a host cell. The choice of these signals (promoters, terminators, secretion leader sequence, etc.) can vary as a function of the host cell used. Moreover, the nucleic acids of the invention can form part of a vector which can replicate autonomously or can integrate. More particularly, autonomously-replicating vectors can be

prepared using sequences which replicate autonomously in the chosen host. As regards the integrating vectors, they can be prepared for example using sequences which are homologous to certain regions of the genome of the host, which allow, by homologous recombination, the integration of the vector. It can be a vector of plasmid, episomal, chromosomal, viral etc., type.

The host cells which can be used for producing the peptide compounds of the invention via the recombinant pathway are both eukaryotic and prokaryotic hosts. Among the eukaryotic hosts which are suitable, mention may be made of animal cells, yeasts or fungi. In particular, as regards yeasts mention may be made of the yeasts of the genus *Saccharomyces*, *Kluyveromyces*, *Pichia*, *Schwanniomyces*, or *Hansenula*. As regards animal cells, mention may be made of COS, CHO, C127, PC12 etc., cells. Among the fungi, mention may be made more particularly of *Aspergillus* ssp. or *Trichoderma* ssp. As prokaryotic hosts, use of the following bacteria is preferred: *E. coli*, *Bacillus* or *Streptomyces*.

A subject of the present invention is also nonhuman mammals comprising in their cells a nucleic acid or a vector according to the invention.

Such mammals (rodents, canines, rabbits, etc.) can in particular be used for studying the properties of PAP1 and identifying compounds for therapeutic purposes. The modification of the genome of

such a transgenic animal may result from an alteration or a modification of one or more genes by "knock-in" or by "knock-out". This modification may be produced using conventional altering agents or mutagenic agents, or by site-directed mutagenesis. The modification of the genome may also result from an insertion of (a) gene(s) or from replacement of (a) gene(s) in its (their) wild-type or mutated form. The modifications of the genome are advantageously performed on reproductive stem cells and advantageously on the pronuclei. The transgenesis may be performed by microinjection of an expression cassette comprising the modified genes into the two fertilized pronuclei. Thus, an animal according to the present invention may be obtained by injection of an expression cassette comprising a nucleic acid. Preferentially, this nucleic acid is a DNA which may be a genomic DNA (gDNA) or a complementary DNA (cDNA).

The construction of transgenic animals according to the invention may be carried out according to conventional techniques known to those skilled in the art. Those skilled in the art may in particular refer to the production of transgenic animals, and particularly to the production of transgenic mice, as described in US patents 4,873,191, 5,464,764 and 5,789,215, the content of these documents being incorporated herein by way of reference.

Briefly, a polynucleotide construct comprising a nucleic acid according to the invention is

inserted into a line of stem cells of the ES type. The polynucleotide construct is preferably inserted by electroporation, as described by Thomas et al. (1987, Cell, Vol. 51: 503-512).

The cells which have undergone the electroporation step are then screened for the presence of the polynucleotide construct (for example by selection using markers, or alternatively by PCR or DNA analysis on an electrophoresis gel such as Southern blotting) in order to select the positive cells which have integrated the exogenous polynucleotide construct into their genome, where appropriate, following a homologous recombination event. Such a technique is, for example, described by MANSOUR et al. (*Nature* (1988) **336**: 348-352).

Next, the positively selected cells are isolated, cloned and injected into 3.5-day mouse blastocysts, as described by BRADLEY (1987, *Production and Analysis of Chimaeric mice*. In: E.J. ROBERTSON (Ed., teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, page 113). Blastocysts are then introduced into a female host animal and the embryo is allowed to develop to term.

According to an alternative, positively selected cells of the ES type are brought into contact with 2.5-day embryos at an 8-16 cell stage (*morulae*) as described by WOOD et al. (1993, Proc. Natl. Acad. Sci. USA, vol. 90: 4582-4585) or by NAGY et al. (1993, Proc.

Natl. Acad. Sci. USA, vol. 90: 8424-8428), the ES cells being internalized so as to extensively colonize the blastocyst, including the cells which give rise to the germ line.

The descendants are then tested in order to determine those which have integrated the polynucleotide construct (the transgene).

The nucleic acids according to the invention can also be used to prepare genetic antisense or antisense oligonucleotides which can be used as pharmaceutical agents. Antisense sequences are oligonucleotides of short length, which are complementary to the coding strand of a given gene, and consequently are capable of specifically hybridizing with the mRNA transcript, which inhibits its translation into a protein. A subject of the invention is thus antisense sequences which are capable of inhibiting, at least partially, the interaction of the PAP1 proteins on parkin. Such sequences can consist of all or part of the nucleic acid sequences defined above. They are generally sequences, or fragments of sequences, which are complementary to sequences encoding peptides which interact with parkin. Such oligonucleotides can be obtained by fragmentation, etc., or by chemical synthesis.

The claimed sequences can be used in the context of gene therapies, for transferring and expressing, *in vivo*, antisense sequences or peptides

which are capable of modulating the interaction of the PAP1 protein with parkin. In this respect, the sequences can be incorporated in viral or nonviral vectors, which allows their administration *in vivo* (Kahn *et al.*, 1991). As viral vectors in accordance with the invention, mention may be made most particularly of adenovirus, retrovirus, adeno-associated virus (AAV) or herpesvirus type vectors. A subject of the present application is also defective recombinant viruses comprising a nucleic acid which encodes a polypeptide according to the invention, in particular a polypeptide or peptide comprising all or part of the sequence SEQ ID NO. 2 or of a derivative of this sequence, for example all or part of the sequence SEQ ID NOS. 12, 14, 42 or 44 or derivatives of these sequences.

The invention also enables the preparation of nucleotide probes, which may or may not be synthetic, and which are capable of hybridizing with the nucleotide sequences defined above or their complementary strand. Such probes can be used *in vitro* as a diagnostic tool for detecting the expression or overexpression of PAP1, or alternatively for revealing genetic abnormalities (incorrect splicing, polymorphism, point mutations, etc.). These probes can also be used for detecting and isolating homologous nucleic acid sequences which encode peptides as defined above, from other cellular sources and preferably from

cells of human origin. The probes of the invention generally comprise at least 10 bases, and they can for example comprise up to the whole of one of the abovementioned sequences or of their complementary strand. Preferably, these probes are labeled prior to their use. For this, various techniques known to persons skilled in the art can be employed (radioactive, fluorescent, enzymatic, chemical labeling, etc.).

The invention also relates to primers or pairs of primers for amplifying all or part of a nucleic acid encoding a PAP1, for example a primer of sequence chosen from SEQ ID NOS. 16-41.

A subject of the invention is also any pharmaceutical composition which comprises, as an active principle, at least one compound as defined above, in particular a peptide compound.

A subject of the invention is in particular any pharmaceutical composition which comprises, as an active principle, at least one antibody and/or one antibody fragment as defined above, as well as any pharmaceutical composition which comprises, as an active principle at least one nucleic acid or one vector as defined above.

A subject of the invention is also any pharmaceutical composition which comprises, as an active principle, a chemical molecule which is capable

of increasing or of decreasing the interaction between the PAP1 protein and parkin.

Moreover, a subject of the invention is also pharmaceutical compositions in which the peptides, antibodies, chemical molecules and nucleotide sequences defined above are combined mutually or with other active principles.

The pharmaceutical compositions according to the invention can be used for modulating the activity of the parkin protein, and consequently for maintaining the survival of the dopaminergic neurons. More particularly, these pharmaceutical compositions are intended for modulating the interaction between the PAP1 protein and parkin. They are, more preferably, pharmaceutical compositions which are intended for treating diseases of the central nervous system, such as for example Parkinson's disease.

A subject of the invention is also the use of the molecules described above for modulating the activity of parkin or for typing diseases of the central nervous system. In particular, the invention relates to the use of these molecules for modulating, at least partially, the activity of parkin.

The invention also relates to a method for screening or characterizing molecules which are active on the function of parkin, which comprises selecting molecules which are capable of binding the sequence SEQ ID NO. 2 or the sequence SEQ ID NO. 4, or a fragment

(or derivative) of these sequences. The method comprises, advantageously, bringing the molecule(s) to be tested into contact, in vitro, with a polypeptide which comprises the sequence SEQ ID NO. 2 or the sequence SEQ ID NO. 4, or a fragment (or derivative) of these sequences, and selecting molecules which are capable of binding the sequence SEQ ID NO. 2 (in particular the region between residues 1 and 344) or the sequence SEQ ID NO. 4. The molecules tested can be varied in nature (peptide, nucleic acid, lipid, sugar, etc., or mixtures of such molecules, for example combinatory libraries, etc.). As indicated above, the molecules thus identified can be used for modulating the activity of the parkin protein, and represent potential therapeutic agents for treating neurodegenerative pathologies.

Other advantages of the present invention will appear upon reading the following examples and figure, which should be considered as illustrative and nonlimiting.

LEGENDS TO THE FIGURE:

Figure 1: Representation of the vector pLex9-parkin (135-290)

Figure 2: Results of the first 5'-RACE experiment. 8 clones were obtained. The initial electronic sequence is indicated at the bottom of the figure.

Figure 3: Results of the second 5'-RACE experiment. Only two of the 8 clones obtained in the first experiment were validated (clones A12 and D5). The initial electronic sequence is indicated at the bottom of the figure. The complete sequence of the DNAs and proteins is given in the sequences 12-15.

Figure 4: Detail of the organization of clones C5 and D4 of the second 5'-RACE experiment. The resulting consensus sequence is indicated at the top of the figure.

Figure 5: Structure of the transcripts isolated from human brain.

Figure 6: Nucleic acid and protein sequence of LY111 (full length) from human brain.

Double underlined: conserved cysteines of the zinc finger domain. Bold: $C_2$1 domain, italics: $C_2$2 domain.

Figure 7: Nucleic acid and protein sequence of LY111 (short version) from human brain. Double underlined: conserved cysteines of the zinc finger domain. Bold: $C_2$1 domain, italics: $C_2$2 domain.

Figure 8: Location of the short (8b) or long (8a) LY111 protein after expression in Cos-7 cells.

Figure 9: Nucleic acid and protein sequence of LY111 (long version) from human lung.

Figure 9: Nucleic acid and protein sequence of LY111 (long version) from human lung.

Figure 10: Nucleic acid and protein sequence of LY111 (short version) from human brain.

MATERIALS AND TECHNIQUES USED

1) Yeast strains:

Strain L40 of the genus *S. cerevisiae (Mata, his3D200, trp1-901, leu2-3,112, ade2, LYS2:: (lexAop)$_4$-HIS3, URA3::(lexAop)$_8$-LacZ, GAL4, GAL80)* was used to verify the protein-protein interactions when one of the protein partners is fused to the LexA protein. The latter is capable of recognizing the LexA response element which controls the expression of the reporter genes *LacZ* and His3.

It was cultured on the following culture media:

Complete YPD medium: - Yeast extract (10 g/l) (Difco)
- Bactopeptone (20 g/l) (Difco)
- Glucose (20 g/l) (Merck)

This medium was solidified by addition of 20 g/l of agar (Difco).

Minimum YNB medium: - Yeast Nitrogen Base (without amino acids) (6.7 g/l) (Difco)
- Glucose (20 g/l) (Merck)

This medium can be solidified by addition of 20 g/l of agar (Difco). It can also be supplemented with amino acids and/or with 3-amino-1,2,3-triazole by addition of CSM media [CSM-Leu, -Trp, -His (620 mg/l), CSM-Trp (740 mg/l) or CSM-Leu, -Trp (640 mg/l) (Bio101)] and/or of 2.5 mM 3-amino-1,2,4-triazole.

2) Bacterial strains:

Strain TG1 of *Escherichia coli*, of genotype supE, hsdΔ5, thi, Δ(lac-proAB), F'[tra D36 pro A^+B^+ $lacI^q$ lacZΔM15], was used for constructing plasmids, as a means of amplifying and of isolating recombinant plasmids used. It was cultured on the following medium:

Medium LB:
- NaCl (5g/l) (Prolabo)
- Bactotryptone (10 g/l) (Difco)
- Yeast extract (5 g/l) (Difco)

This medium is solidified by addition of 15 g/l of agar (Difco).

Ampicillin was used at 100 µg/ml; this antibiotic is used to select the bacteria which have received the plasmids bearing, as a marker, the gene for resistance to this antibiotic.

Strain HB101 of *Escherichia coli* of genotype supE44, ara14, galK2, lacY1, Δ(gpt-proA)62, rpsL20(Str^r), xyl-5, mtl-1, recA13, Δ(mcrC-mrr), $HsdS^-(r^-m^-)$ was used as means for amplifying and for isolating plasmids which originate from the human lymphocyte cDNA library.

It was cultured on

Medium M9:
- $-Na_2HPO_4$ (7 g/l) (Prolabo)
- $-KH_2PO_4$ (3 g/l) (Prolabo)
- $-NH_4Cl$ (1 g/l) (Prolabo)
- -NaCl (0.5 g/l) (Prolabo)
- -Glucose (20 g/l) (Sigma)
- $-MgSO_4$ (1 mM) (Prolabo)
- -Thiamine (0.001%) (Sigma)

This medium is solidified by addition of 15 g/l of agar (Difco).

Leucine (50 mg/l) (Sigma) and proline (50 mg/l) (Sigma) should be added to the M9 medium to enable the growth of strain HB101.

During the selection of plasmids which originate from the lymphocyte cDNA two-hybrid library, leucine was not added to the medium because the plasmids bear a Leu2 selection marker.

3) Plasmids:

The 5-kb vector pLex9 (pBTM116) (Bartel *et al.*, 1993), which is homologous to pGBT10 and which contains a multiple cloning site located downstream of the sequence which encodes the LexA bacterial repressor, and upstream of a terminator, for forming a fusion protein.

pLex-HaRasVal12; plasmid pLex9, as described in application WO 98/21327, which contains the sequence encoding the HaRas protein mutated at position Val12, which is known to interact with the mammalian Raf protein (Vojtek *et al.*, 1993). This plasmid was used to test the specificity of interaction of the PAP1 protein in strain L40.

pLex9-cAPP; plasmid pLex9 which contains the sequence encoding the cytoplasmic domain of the APP protein, known to interact with the PTB2 domain of FE65. This plasmid was used to test the specificity of interaction of the PAP1 protein in strain L40.

4) Synthetic oligonucleotides:

TTAAGAATTC GGAAGTCCAG CAGGTAG (SEQ ID N°5)

ATTAGGATCC CTACACACAA GGCAGGGAG (SEQ ID N°6)

Oligonucleotides which made it possible to obtain the PCR fragment which corresponds to the central region of parkin, bordered with the sites *EcoRI* and *BamHI*.

GCGTTTGGAA TCACTACAG (SEQ ID N°7)

GGTCTCGGTG TGGCATC (SEQ ID N°8)

CCGCTTGCTT GGAGGAAC (SEQ ID N°9)

CGTATTTCTC CGCCTTGG (SEQ ID N°10)

AATAGCTCGA GTCAGTGCAG GACAAGAG (SEQ ID N°11)

Oligonucleotides which were used to sequence the insert corresponding to the PAP1 gene.

The oligonucleotides are synthesized using an Applied System ABI 394-08 machine. They are removed from the synthesis matrix with ammonia and precipitated twice with 10 volumes of n-butanol, and then taken up in water. The quantification is carried out by measuring the optical density ($1OD_{260}$ corresponds to 30 µg/ml).

5) Preparation of plasmid DNAs

The preparations, in small amount and in large amount, of plasmid DNA were carried out according to the protocols recommended by Quiagen, the manufacturer of the DNA purification kits:

- Quiaprep Spin Miniprep kit, reference: 27106
- Quiaprep Plasmid Maxiprep kit, reference: 12613.

6) Enzymatic amplification of DNA by PCR (Polymerase Chain Reaction):

The PCR reactions are carried out in a final volume of 100 µl in the presence of the DNA matrix, of dNTP (0.2 mM), of PCR buffer (10 mM Tris-HCl pH 8.5, 1 mM $MgCl_2$, 5 mM KCl, 0.01% gelatin), of 10 to 20 pmol of each one of the oligonucleotides and of 2.5 IU of Ampli Taq DNA polymerase (Perkin Elmer). The mixture is covered with 2 drops of liquid petroleum jelly to limit the evaporation of the sample. The machine used is the "Crocodile II" from Appligene.

We used a matrix denaturation temperature of 94°C, a hybridization temperature of 52°C and a temperature for elongation by the enzyme at 72°C.

7) Ligations:

All the ligation reactions are carried out at 37°C for one hour in a final volume of 20 µl, in the presence of 100 to 200 ng of vector, 0.1 to 0.5 µg of insert, 40 IU of T4 DNA ligase enzyme (Biolabs) and a ligation buffer (50 mM Tris-HCl pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP). The negative control consists of ligating the vector in the absence of insert.

8) Transformation of bacteria:

The transformation of bacteria with a plasmid is carried out according to the following protocol: 10 µl of the ligation volume are used to transform the TG1 bacteria, according to the method of Chung (Chung *et al.*, 1989). After transformation, the bacteria are

plated on an LB medium + ampicillin and incubated for 16 h at 37°C.

9) Separation and extraction of DNAs:

The separation of DNAs is carried out as a function of their size, on agarose gel electrophoresis according to Maniatis (Maniatis *et al.*, 1989): 1% agarose gel (Gibco BRL) in a TBE buffer (90 mM Tris base; 80 mM borate; 2 mM EDTA).

10) Fluorescent sequencing of plasmid DNAs:

The sequencing technique used is derived from the method of Sanger (Sanger *et al.*, 1977) and adapted for sequencing by fluorescence, which is developed by Applied Biosystems. The protocol used is that described by the designers of the system (Perkin Elmer, 1997).

11) Transformation of yeast:

The plasmids are introduced into the yeast using a conventional technique for transforming yeast, which was developed by Gietz (Gietz *et al.*, 1992) and which was modified in the following way:

In the specific case of the transformation of yeast with the lymphocyte cDNA library, the yeast used contains the plasmid pLex9-parkin (135-290), which encodes the central portion of parkin fused to the LexA protein. It is cultured in 200 ml of YNB minimum medium, supplemented with amino acids CSM-Trp, at 30°C with shaking until a density of 10^7 cells/ml is attained. To carry out the transformation of the yeasts, according to the above protocol, the cell

suspension was separated into 10 50-µl tubes, into which 5 µg of the library were added. Heat shock was carried out for 20 minutes, and the cells were collected by centrifugation and resuspended in 100 ml of YPD medium for 1 h at 30°C, and in 100 ml of YNB medium, supplemented with CSM-Leu, -Trp, for 3 h 30 at 30°C. The efficiency of the transformation is determined by placing various dilutions of transformed cells on solid YNB medium which is supplemented with CSM-Trp, -Leu. After 3 days of culture at 30°C, the colonies obtained were counted, and the degree of transformation per µg of lymphocyte library DNA was determined.

12) Isolation of plasmids extracted from yeast:

5 ml of a yeast culture, which is incubated for 16 h at 30°C, are centrifuged, and taken up in 200 µl of a lysis buffer (1M Sorbitol, 0.1 M KH_2PO_4/K_2HPO_4 pH 7.4, 12.5 mg/ml zymolyase) and incubated for 1 h at 37°C. The lysate is then treated according to the protocol recommended by Quiagen, the manufacturer of the DNA purification kit, Quiaprep Spin Miniprep kit, ref 27106.

13) β-galactosidase activity assay:

A sheet of nitrocellulose is preplaced on the Petri dish containing the yeast clones, which are separated from each other. This sheet is then immersed in liquid nitrogen for 30 seconds, in order to rupture the yeasts and thus to release the β-galactosidase

activity. After thawing, the sheet of nitrocellulose is placed, colonies facing upwards, in another Petri dish containing a Whatman paper which has been presoaked in 1.5 ml of PBS solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7) containing 15 µl of X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactoside) at 40 mg/ml of N,N-dimethylformamide. The dish is then placed in an incubator at 37°C. The assay is termed positive when the colonies on the membrane turn blue after 12 hours.

EXAMPLE 1: CONSTRUCTION OF A VECTOR WHICH ALLOWS THE EXPRESSION OF A FUSION PROTEIN IN WHICH FUSION IS BETWEEN THE CENTRAL PORTION OF PARKIN AND THE LEXA BACTERIAL REPRESSOR.

Screening a library using the double-hybrid system requires the central region of parkin to be fused to a DNA binding protein, such as the LexA bacterial repressor. The expression of this fusion protein is carried out using the vector pLex9 (cf. materials and methods), into which the sequence encoding the central region of parkin, which is in the sequence presented in sequence SEQ ID NO. 3 or 4, is introduced, in the same reading frame as the sequence corresponding to the LexA protein.

The 468 bp-fragment of DNA corresponding to the 156 amino acids of the central region of parkin which begins at amino acid 135, was obtained by PCR using the oligonucleotides (sequence SEQ ID NO. 5 and No. 6), which also made it possible to introduce the site *EcoRI* at the 5′ end and a stop codon and a site *BamHI* at the 3′ end. The PCR fragment was introduced between the *EcoRI* and *BamHI* sites of the multiple cloning site of the plasmid pLex9, downstream of the sequence encoding the protein LexA, to give the vector pLex9-parkin (135-290) (Fig. 1).

The construct was verified by sequencing the DNA. This verification made it possible to show that this fragment does not have mutations generated during

the PCR reaction, and that it was fused in the same open reading frame as that of the fragment corresponding to LexA.

EXAMPLE 2: SCREENING A LYMPHOCYTE FUSION LIBRARY

We used the double-hybrid method (Fields and Song, 1989).

Screening a fusion library makes it possible to identify clones producing proteins which are fused to the transactivating domain of GAL4, and which are able to interact with the protein of interest described in Example 1 (central region of parkin). This interaction makes it possible to reconstitute a transactivator which will then be capable of inducing the expression of the reporter genes His3 and *LacZ* in strain L40.

To carry out this screening we chose a fusion library which is prepared from cDNA originating from peripheral human lymphocytes, supplied by Richard Benarous (Peytavi *et al.*, 1999). Yeasts were transformed with the lymphocyte library and positive clones were selected as described above.

During screening, it is necessary to maintain the probability that each separate plasmid from the fusion library is present in at least one yeast at the same time as the plasmid pLex9-parkin (135-290). To maintain this probability, it is important to have a good efficiency of transformation of the yeast. For

this, we chose a protocol for transforming yeast which gives an efficiency of 2.6×10^5 transformed cells per µg of DNA. In addition, as cotransforming yeast with two different plasmids reduces this efficiency, we preferred to use a yeast which is pretransformed with the plasmid pLex9-parkin (135-290). This strain L40 pLex9-parkin (135-290), of phenotype His-, Lys-, Leu-, Ade-, was transformed with 50 µg of plasmid DNA from the fusion library. This amount of DNA enabled us to obtain, after estimation, 1.3×10^7 transform cells, which corresponds to a number which is slightly higher than the number of separate plasmids which constitute the library. According to this result, it may be considered that virtually all of the plasmids of the library were used to transform the yeasts. The selection of the transform cells, which are capable of reconstituting a functional transactivator, was done on a YNB medium which was supplemented with 2.5 mM 3-amino-1,2,4-triazole and 620 mg/l of CSM (Bio101), and which contains no histidine, no leucine and no tryptophan.

Upon termination of this selection, many clones with a His+ phenotype were obtained. A β-galactosidase activity assay was carried out on these transformants, to validate, by the expression of the other reporter gene, *LacZ*, this number of clones obtained. 115 clones had the His+, β-Gal+ double

phenotype, which can correspond to a protein-protein interaction.

EXAMPLE 3: ISOLATION OF THE LIBRARY PLASMIDS IN THE CLONES SELECTED.

To identify the proteins which are able to interact with the central region of parkin, the fusion library plasmids contained in the yeasts which were selected during the double-hybrid screening were extracted. To be able to obtain a large amount thereof, this isolation requires a prior transformation of *E. coli* with an extract of DNA from the positive yeast strains. As the library plasmid which is contained in this extract is a yeast/*E. coli* shuttle plasmid, it can easily replicate in the bacterium. The library plasmid was selected by complementing the auxotrophic HB101 bacterium for leucine, on leucine-lacking medium.

The plasmid DNAs from the bacterial colonies which are obtained after transformation with extracts of DNA from yeasts were analyzed by digestion with restriction enzymes and separation of the DNA fragments on an agarose gel. Among the 115 clones analyzed, one clone containing a library plasmid, which showed a different profile from the others, was obtained. This plasmid, termed pGAD-Ly111b, was studied more precisely.

EXAMPLE 4: DETERMINATION OF THE SEQUENCE OF THE INSERT CONTAINED IN THE PLASMID IDENTIFIED.

Sequencing of the insert contained in the plasmid identified was carried out, firstly, using the oligonucleotide SEQ ID No. 7, which is complementary to the sequence GAL4TA, close to the *EcoRI* site of insertion of the lymphocyte cDNA library; then, secondly, using the oligonucleotides SEQ ID No. 8 to SEQ ID No. 11, which correspond to the sequence of the insert which is obtained during the course of the sequencing. The sequence obtained is presented on the sequence SEQ ID NO. 1. The protein thus identified was referred to as PAP1 (Parkin-Associated Protein 1).

Comparison of the sequence of this insert with the sequences which are contained in the GENBank and EMBL (European Molecular Biology Lab) databases showed a homology of 25%, at the protein level, with various members of the synaptotagmin family. The synaptotagmins are part of a family of membrane proteins which are encoded by at least eleven different genes, which are expressed in the brain and other tissues. They contain a single transmembrane domain and two calcium-regulated domains which are termed C_2. It is in this domain that the homology between the synaptotagmins and the PAP1 protein is found. No other significant homology was observed.

EXAMPLE 5: ANALYSIS OF THE SPECIFICITY OF INTERACTION BETWEEN THE CENTRAL REGION OF PARKIN AND THE PAP1 PROTEIN.

To determine the specificity of interaction between the fragment corresponding to the PAP1 protein and the central region of parkin, a two-hybrid test for specific interaction with other nonrelevant proteins was carried out. To carry out this test, we transformed strain L40 with the control plasmids plex9-cAPP or pLex9-HaRasVal12, which respectively encode the cytoplasmic domain of the APP or the HaRasVal12 protein, which are fused to the domain the DNA binding of LexA, in place of the plasmid pLex9-parkin (135-290), and with the plasmid which is isolated in the screening of the two-hybrid library. A β-Gal activity assay was carried out on the cells which were transformed with the various plasmids, to determine a protein-protein interaction. According to the result of the assay, only the yeasts which were transformed with the plasmid which was isolated during the screening of the two-hybrid library, and with the plasmid pLex9-parkin (135-290), had a β-Gal+ activity, which thus shows an interaction between the central region of parkin and the PAP1 protein. This interaction thus turns out to be specific, since this fragment of PAP1 does not seem to interact with the cAPP or HaRasVal12 proteins.

These results thus show the existence of a novel protein, referred to as PAP1, which is capable of interacting specifically with parkin. This protein, which is related to the synaptotagmins, shows no significant homology with known proteins, and can be used in therapeutic or diagnostic applications, for producing antibodies, probes or peptides, or for screening active molecules.

EXAMPLE 6: CLONING OF THE PAP1 GENE FROM A HUMAN LUNG DNA LIBRARY

With the aim of identifying the complete sequence of the human PAP1 gene and of characterizing the existence of variant forms, two electronic extension approaches were carried out using the sequence SEQ ID NO. 1. Two electronic sequences were thus obtained, of 1644 bp and 1646 bp, respectively, comprising a 330 bp extension compared to the sequence SEQ ID NO. 1. However, analysis of these sequences showed differences in the consensus region, which were apparent after translation. Thus, a 420aa ORF is obtained in one case, and a 230aa ORF with the other sequence. The protein sequence obtained was compared with known sequences and revealed a 24% homology, over the 293 overlapping amino acids, with human synaptogamin I (p65)(p21579). The function of synaptogamin I may be a regulatory role in membrane interactions during the trafficking of synaptic

vesicles in the area of the synapse. It binds acid phospholipids with a certain specificity. In addition, a calcium-dependent interaction between synaptogamin and receptors for activated protein kinase C has been reported. The synaptogamin may also bind three other proteins, namely neurexin, syntaxin and ap2. Given the abrupt and early disappearance of any homology between the sequences identified and the synaptogamin family, it is possible that the sequence identified has a deletion compared to the natural sequence. To verify this hypothesis and validate the sequences, an RT-PCR sequencing experiment was carried out using the 1644 bp sequence. The sequence to obtain comprises a 420aa ORF having a homology of the same order with synaptogamins.

To attempt to obtain a larger sequence, and to verify whether the sequence obtained could correspond to a splicing form, a 5'-RACE extension experiment was initiated from the 3' region of the validated sequence, using the oligos L1 and L2 on a preparation of human lung cDNA.

The results obtained are given in figure 2 and show the identification of 8 clones corresponding to 6 different terminal 5' ends. Three of them contain a STOP codon which interrupts the ORF (clones A12, F2, F12) and clone A3 contains no OPR. The presence of various transcripts was confirmed by RT-PCR and nested RT-PCR (table 1).

Table 1

RT-PCR	Primary	Secondary PCR U3-L3	Secondary PCR U1-L4	Secondary PCR C-B
U3-L3	170			
A-L4	153		+	
A-L3	Smear	+		
U1-L4	130			
U1-L3	Smear	+		
U1-B	415			+
U2-B	515			+
Expected size		170	130	120

The pairs of primers U3-L3 and C-B are specific for the common fragment of the sequence, the oligos A and U1 are specific for the initial sequence and for clone C11, the oligo L4 is specific for the initial sequence and the primer U2 is specific for clone A3. A second 5'-RACE was performed with the oligos L3 and L7 located in the region common to the various clones (figure 2). The results obtained are given in figures 3 and 4. The presence of various transcripts was confirmed by RT-PCR and nested RT-PCR (table 2).

Table 2

RT-PCR	Result	Secondary PCR C-B	Secondary PCR U3-B	Secondary PCR U5-L7
U4-F	Smear	+	+	+
U5-F	Smear	+	+	+
U3-F	1550 bp	+	+	
Expected size (bp)		120	385	530

The sequence of the primers and oligonucleotides is given in tables 3 and 4 (SEQ ID NOS. 16-37).

Table 3

					SEQ ID
LY111_U4	CCAGTTCTGCCTGTTCATC	23	to	41	16
LY111_U5	TTCAAAACACAGAGGAGGAG	319	to	338	17
LY111_U3	GAATTTGGTCAGTTTAGAGG	759	to	778	18
LY111_L7	TTCTGGGATTTGGAGAGCTTTTTCAC	851	to	825	19
LY111_L6	TCTGTCTGTCCCACACACTGCC	914	to	892	20
LY111_L3	GACTGGCTCCGTCTCTCTG	928	to	910	21
LY111_C	AAGCAACAGAATCTCCCATCC	1029	to	1049	22
LY111_B	GCATTGTCAAAATTGCCCATC	1147	to	1127	23
LY111_E	AGGCGGAGAAATACGAAGAC	1543	to	1562	24
LY111_D	GCAGAGTGAGACAGCCCTTAAC	1767	to	1746	25
Ly111_L2	CTTCCTCAGGACTGGCGACTTCAG	1811	to	1782	26
Ly111_L1	CAAGCGGTCGTTCATTCCAAAGAG	1934	to	1913	27
LY111_F	AAGAGGAGATAACCCACCAGAG	2288	to	2269	28

Table 4

LY111_A	TCGTAGAGCAGCAGGTCCAAG	14	to	34	46
LY111_U1	AGGGCTGCTGGCTATTTTTC	36	to	55	29
LY111_L4	TAAGAAATGGGTTGTGAAC	148	to	166	30
LY111_C	AAGCAACAGAATCTCCCATCC	1029	to	1049	31
LY111_B	GCATTGTCAAAATTGCCCATC	1147	to	1127	32
LY111_E	AGGCGGAGAAATACGAAGAC	1543	to	1562	33
LY111_D	GCAGAGTGAGACAGCCCTTAAC	1767	to	1746	34
Ly111_L2	CTTCCTCAGGACTGGCGACTTCAG	1811	to	1782	35
Ly111_L1	CAAGCGGTCGTTCATTCCAAAGAG	1934	to	1913	36
LY111_F	AAGAGGAGATAACCCACCAGAG	2288	to	2269	37

This set of results makes it possible to validate the consensus sequence corresponding to the long isoform (figure 9, SEQ ID NOS. 12 and 13) and to the short isoform (figure 10, SEQ ID NOS. 14 and 15) of the PAP1 protein identified from human lung. In the remainder of the examples, this protein is also referred to using the term LY111. The long isoform is encoded by a 1833 bp ORF located at residues 237-2069 of SEQ ID NO. 12, and comprises 610 amino acids. The polyadenylation signal is located starting from

nucleotide 2315. The short isoform is encoded by a 942 bp ORF located at residues 429-1370 of SEQ ID NO. 14, and comprises 313 amino acids. The polyadenylation signal is located starting from nucleotide 1616.

Northern blotting experiments were then carried out on various human tissues with probes (amplimer CD and E-F) and made it possible to reveal a 6 kb transcript in muscle, a transcript in heart (3 kb) and also a 6 kb transcript in fetal liver. Example 7 describes, moreover, the cloning of a transcript in human fetal brain.

Various homology studies were carried out in various protein databases, the results of which are given in table 5 below.

Table 5

Databank	Homology
Genpept116	G5926736 (AB025258) granuphilin-a Identity: 31%(215/679), Homology (POS): 46%(322/679)
	G5926738 (AB025259) granuphilin-b Identity: 31%(150/479), Homology (POS): 47%(230/479)
	G1235722 (D70830) Doc2 beta (homo sapiens) Identity: 25%(74/292), Homology (POS): 43%(127/292)
	G289718 (L15302) Synaptogamin-1 Identity: 26%(77/293), Homology (POS): 45%(133/293)
Swissprot	SP:SYT1_CAEEL Synaptogamin 1 Identity: 26%(77/293), Homology (POS): 45%(133/293)
	SP:SYT2_MOUSE Synaptogamin II Identity: 24%(72/293), Homology (POS): 44% (131/293)

EXAMPLE 7: CLONING OF TWO FULL-LENGTH TRANSCRIPTS OF PAP1 (LY111B) FROM HUMAN FETAL BRAIN COMPLEMENTARY DNA

In order to confirm the presence of a full-length Ly111b transcript in human brain, a PCR was carried out using complementary DNA derived from human fetal brain (Marathon Ready cDNA, Clontech), using, as primers, the oligonucleotides LyF1 (AAT GGA AGG GCG TGA CTC, figure 5, SEQ ID NO. 38) and HA71 (CCT CAC GCC TGC TGC AAC CTG, SEQ ID NO. 39). A weakly represented DNA fragment of approximately two kilobases was amplified. The product of this first PCR was used as matrix for a nested PCR, carried out with the oligonucleotides LyEcoF (GCACGAATTC ATG GCC CAA GAA ATA GAT CTG, SEQ ID NO. 40) and HA72 (CTG TCT TCG TAT TTC TCC GCC TTG, SEQ ID NO. 41). The amplified products were digested with the restriction enzymes EcoRI (integrated into the oligonucleotide LyEcoF)and BstEII (figure 5) and inserted into the expression vector pcDNA3, and their sequence was then determined. Analysis of the sequence of clones obtained revealed the presence of two potential full-length Ly111b transcripts in human fetal brain (figure 5). The first of these transcripts ($Ly111b_{full A}$) corresponds to the mRNA identified in human lung (example 6) and encodes a 609 amino acid protein ($pLy111b_{full A}$; figures 5 and 6, SEQ ID NOS. 42-43). The second ($Ly111b_{full B}$) probably represents a product of alternative splicing of a common primary mRNA. In this transcript, which is identical to $Ly111b_{full A}$, the sequence between nucleotides 752 and 956 of the sequence validated in human lung is absent (SEQ ID NO.

42). $Ly111b_{full B}$ thus encodes a 541 amino acid protein ($pLy111b_{full B}$) identical to $pLy111_{full A}$, in which, however, the domain between amino acids 172 and 240 (figures 5 and 7, SEQ ID NOS. 44-45) has come to be missing. The two proteins $pLy111b_{full A/full B}$ integrate the domain of interaction with the fragment of Parkin comprising amino acids 135 to 290, identified in yeast (initial sequence Ly111b, figure 5), and can therefore theoretically maintain this interaction.

The $pLy111b_{full A/full B}$ proteins belong to the RIM/Rabphilin family

$pLy111b_{full A/full B}$ show homology with the proteins of the RIM/rabphilin family (Wang Y, Sugita S & Südhof TG. The RIM/NIM family of neuronal C2 domain proteins. *J Biol Chem* (2000) 275, 20033-20044) and in particular with the granulophilins (Wang Jie, Takeuchi T, Yokota H & Izumi T. Novel Rabphilin-3-like protein associates with insulin-containing granules in pancreatic beta cells. *J Biol Chem* (1999) 274, 28542-28548). They are characterized by the presence of a zinc finger domain in the N-terminal portion and of two C_2 domains in the C-terminal portion (figures 6 and 7). The zinc finger domain of the proteins of the RIM/rabphilin family has been implicated in the interaction with the Rab proteins. The latter, which are GTP-binding proteins, are essential components of the membrane trafficking machinery in eukaryotic cells.

In addition, it has been described that the C_2 domains of the proteins of the RIM/rabphilin family may bind membranes through the interaction with phospholipids.

Expression of the pLy111b$_{fullA/fullB}$ proteins in cells of the cos-7 line: colocalization with Parkin

The coding sequence of the Ly111b$_{fullA/B}$ transcripts was inserted into the eukaryotic expression vector pcDNA3 in frame with the sequence encoding an N-terminal myc epitope (pcDNA3-mycLy111b$_{fullA/B}$). Cells of the cos-7 line transfected using these vectors produce proteins with an apparent molecular weight of approximately 67 kDa (pcDNA3-mycLy111b$_{fullA}$) and 60 kDa (pcDNA3-mycLy111b$_{fullB}$), corresponding to the expected molecular weight. These proteins, revealed by immunolabeling using an antibody directed against the N-terminal myc epitope, are distributed nonhomogeneously, in a punctate manner, in the cytoplasm, the extensions and sometimes the nucleus of the cells of the cos-7 line (figure 8a, b, column A). When they are overexpressed with Parkin, revealed using the anti-Parkin antibody Asp5, in the cells of the cos-7 line (figure 8a, b, column B), a similar distribution is a colocalization of these proteins may be observed (figure 8a, b, column C).

BIBLIOGRAPHIC REFERENCES

Abbas, N. *et al.* (1999). A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. *Hum Mol Genet.* **8**, 567-574.

Bartel, P.L. *et al.* (1993). D.A. Hartley Ed, Oxford University Press, 153

Chung, C.T. *et al.* (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution, *Proc. Natl. Acad. Sci. USA.* **86**, 2172-2175.

Fields, S. and **Song,** O. (1989). A novel genetic system to detect protein-protein interactions. *Nature.* **340,** 245-246.

Gietz, R.D. *et al.* (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.,* **20**, 1425.

Hattori, N. *et al.* (1998). Molecular genetic analysis of a novel Parkin gene in Japanese families with autosomal recessive juvenile parkinsonism: evidence for variable homozygous deletions in the Parkin gene in affected individuals. *Ann Neurol.* **6**, 935-941.

Kahn, A. *et al.*, (1991) Thérapie génique: espoirs et limites, *Médecine et Sciences.* **7**, 705-714.

Kitada, T. *et al.* (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature.* **392,** 605-608.

Leroy, E. *et al.* (1998). The ubiquitin pathway in Parkinson's disease. *Nature.* **395**, 451-452.

Lewy, F.H. (1912). in *Handbuch der Neurologie* (Lewandowski, M., ed) pp 920-933, Springer, Berlin.

Lucking, C. *et al.* (1998). Homozygous deletions in parkin gene in European and North African families with autosomal recessive juvenile parkinsonism. *Lancet.* **352,** 1355-1356.

Maniatis, T. *et al.* (1989). Molecular cloning, second edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Morett, E. (1999). A novel transactivation domain in parkin. *Trends Biochem Sci.* **24**, 229-231.

Peytavi, R. *et al.* (1999). HEED, the product of the human homolog of the murine eed gene, binds to the matrix protein of HIV-1. *J. Biol. Chem.* **274,** 1635-1645.

Polymeropoulos, M.H. *et al.* (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science.* **276**, 2045-2047.

Sanger, F. *et al.* (1997). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.

Shimura, H. *et al.* (1999). Immunohistochemical and subcellular localization of Parkin protein: absence of protein in autosomal recessive juvenile parkinsonism patients. *Ann Neurol.* **45**, 668-672.

Sunada, Y. *et al.* (1998), Differential expression of the parkin gene in the human brain and peripheral leukocytes. *Neurosci Lett.* **254**, 180-182.

Vojtek, AB. *et al.*, (1993). Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell.* **74**, 205-214.

CLAIMS

1. Compound capable of modulating, at least partially, the interaction between the PAP1 protein, or a homolog of this protein, and parkin.

2. Compound according to claim 1, characterized in that it slows, inhibits or stimulates, at least partially, said interaction.

3. Compound according to either of claims 1 and 2, characterized in that it is capable of binding the domain of interaction between the PAP1 protein, or a homolog of this protein, and parkin.

4. Compound according to one of claims 1 to 3, characterized in that it is a compound of peptide, nucleic acid, lipid or saccharide type, or an antibody.

5. Compound according to claim 4, characterized in that it is a peptide compound comprising all or part of the peptide sequence SEQ ID NO. 2 or a derivative thereof.

6. Compound according to claim 4, characterized in that it is a peptide compound comprising a region of which the sequence corresponds to all or a functional part of the site of interaction of the PAP1 protein with parkin.

7. Compound according to claim 4, characterized in that it is a peptide compound which is

derived from the PAP1 protein (and/or from the homologous forms), and which bears an effector region which has been made nonfunctional.

8. Polypeptide comprising the sequence SEQ ID NO. 2 or a derivative or fragment of this sequence.

9. Polypeptide according to claim 8, comprising at least 5 consecutive residues of the sequence SEQ ID NO. 2, preferably at least 9, more preferably at least 15.

10. Polypeptide according to claim 8, comprising all or part of the sequence SEQ ID NOS. 13, 15, 43 or 45 or a variant of these sequences, in particular at least 5 consecutive residues, preferably at least 9, more preferably at least 15 consecutive residues of the sequence SEQ ID NOS. 13, 15, 43 or 45.

11. Nucleic acid encoding a peptide compound according to one of claims 4 to 10.

12. Nucleic acid according to claim 11, characterized in that it comprises all or part of the sequence SEQ ID NOS. 1, 12, 14, 42 or 44 or a sequence which is derived from this sequence.

13. Nucleic acid encoding a polypeptide according to claim 8 or 11.

14. Nucleic acid, in particular a nucleotide probe, which is capable of hybridizing with a nucleic acid according to one of claims 11 to 13, or with their complementary strand.

15. Vector comprising a nucleic acid according to one of claims 11 to 14.

16. Defective recombinant virus comprising a nucleic acid according to one of claims 11 to 14.

17. Nucleic acid chosen from the nucleic acids of sequence SEQ ID NOS. 16-41 and 46.

18. Antibody or antibody fragment or derivative, characterized in that it is directed against a peptide compound according to one of claims 4 to 10.

19. Antibody according to claim 18, characterized in that it recognizes a polypeptide according to claim 9 or 10.

20. Pharmaceutical composition comprising at least one compound according to one of claims 1 to 10, or an antibody according to claim 18 or 19.

21. Nonpeptide compound or a compound which is not of exclusively peptide nature, which is capable of modulating, at least partially, the interaction of the PAP1 protein, or a homolog of this protein, with parkin.

22. Compound according to claim 21, characterized in that the active motifs of a peptide according to one of claims 5 to 7 have been duplicated with a structure which is not a peptide or which is not of exclusively peptide nature.

23. Pharmaceutical composition comprising at least one nucleic acid according to one of claims 11 to 14, or one vector according to claim 15 or 16.

24. Pharmaceutical composition comprising a peptide compound according to one of claims 4 to 10.

25. Composition according to claim 20, 23 or 24, intended for treating neurodegenerative pathologies.

26. Method for screening or for characterizing active molecules, which comprises a step of selecting molecules which are capable of binding the sequence SEQ ID NO. 2 or the sequence SEQ ID NO. 4, or a fragment of these sequences.

27. Method for screening or for characterizing active molecules, which comprises a step of selecting molecules which are capable of binding a sequence chosen from SEQ ID NOS. 13, 15, 43 and 45, or a fragment of these sequences.

28. Method for producing a peptide compound according to one of claims 4 to 10, comprising the culture of a cell which contains a nucleic acid according to one of claims 11 to 14 or a vector according to claim 15 or 16, under conditions for expressing said nucleic acid, and the recovery of the peptide compound produced.

29. Human PAP1 protein in isolated form.

30. Cell containing a nucleic acid according to one of claims 11 to 14 or a vector according to claim 15 or 16.

31. Nonhuman mammal comprising in its cells a nucleic acid according to one of claims 11 to 14.

32. Use of a compound as claimed in any one of claims 1 to 10, an antibody as claimed in claim 18 or 19, a nucleic acid as claimed in claim 11 or 14, a vector as claimed in claim 15 or 16, or a peptide compound as claimed in any one of claims 4 to 10, in the manufacture of a preparation for treating neurodegenerative pathologies.

33. A substance or composition for use in a method for treating neurodegenerative pathologies, said substance or composition comprising a compound as claimed in any one of claims 1 to 10, an antibody as claimed in claim 18 or 19, a nucleic acid as claimed in claim 11 or 14, a vector as claimed in claim 15 or 16, or a peptide compound as claimed in any one of claims 4 to 10, and said method comprising administering said substance or composition.

34. A compound according to any one of claims 1 to 7, or 21, substantially as herein described and illustrated.

35. A polypeptide according to claim 8, substantially as herein described and illustrated.

36. A nucleic acid according to any one of claims 11 to 14 or 17, substantially as herein described and illustrated.

37. A vector according to claim 15, substantially as herein described and illustrated.

38. A virus according to claim 16, substantially as herein described and illustrated.

39. An antibody according to claim 18,

substantially as herein described and illustrated.

40. A composition according to claim 20, or claim 23, or claim 24, substantially as herein described and illustrated.

41. A method according to claim 26, or claim 27, substantially as herein described and illustrated.

42. A method according to claim 28, substantially as herein described and illustrated.

43. A protein according to claim 29, substantially as herein described and illustrated.

44. A cell according to claim 30, substantially as herein described and illustrated.

45. A mammal according to claim 31, substantially as herein described and illustrated.

46. Use according to claim 32, substantially as herein described and illustrated.

47. A substance or composition for use in a method of treatment according to claim 33, substantially as herein described and illustrated.

48. A new compound; a new polypeptide; a new nucleic acid; a new vector; a new virus; a new antibody; a new composition; a new method for screening or characterizing molecules; a new method for producing a peptide; a new protein; a new cell; a new mammal; a new use of a compound as claimed in any one of claims 1 to 10, an antibody as claimed in claim 18 or 19, a nucleic acid as claimed in claim 11 or 14, a vector as claimed in claim 15 or 16, or a peptide compound as claimed in any one of claims 4 to 10; or a substance or composition for a new use in a method of treatment; substantially as herein described.