



(86) Date de dépôt PCT/PCT Filing Date: 1995/05/30
(87) Date publication PCT/PCT Publication Date: 1995/12/07
(45) Date de délivrance/Issue Date: 2012/09/25
(85) Entrée phase nationale/National Entry: 1996/10/16
(86) N° demande PCT/PCT Application No.: US 1995/006722
(87) N° publication PCT/PCT Publication No.: 1995/033052
(30) Priorités/Priorities: 1994/05/27 (US08/250,795);
1994/12/20 (US08/360,144)

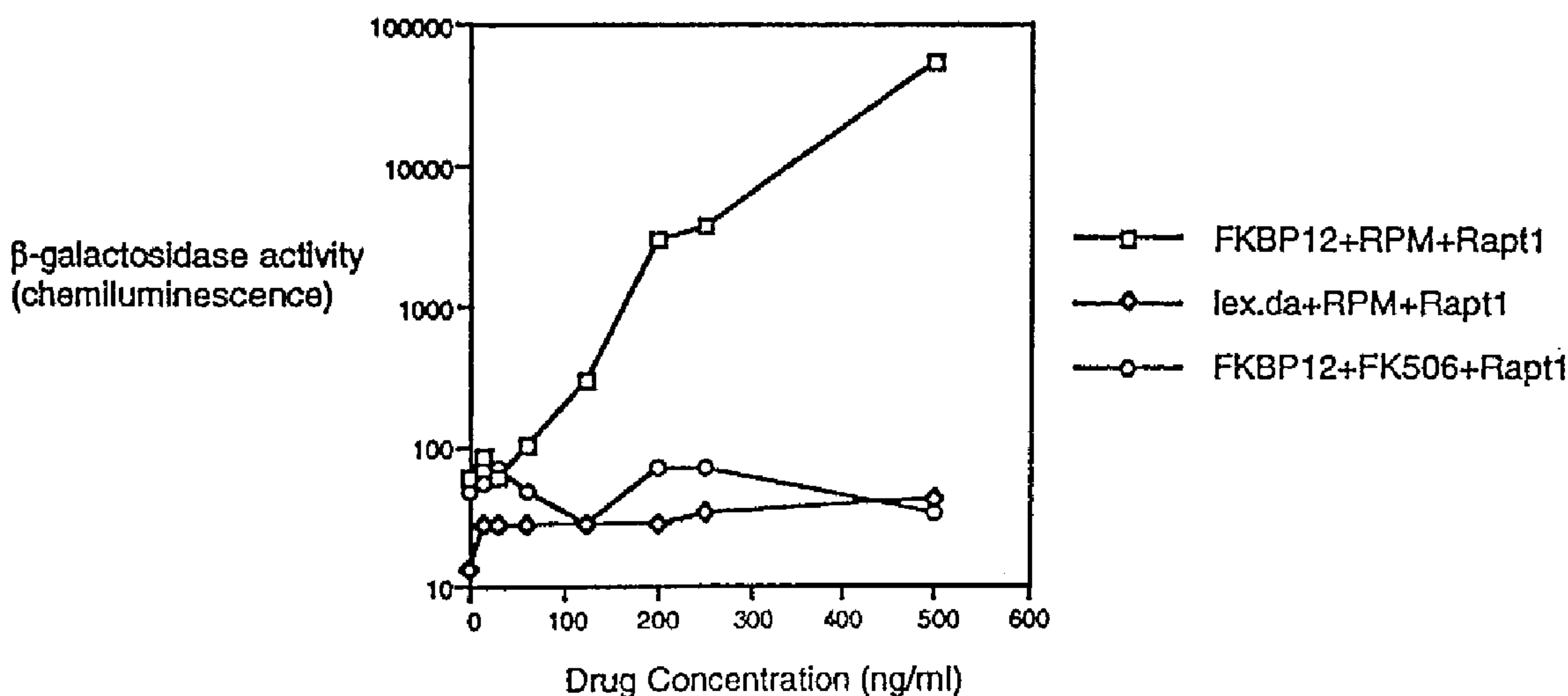
(51) Cl.Int./Int.Cl. *C12N 15/62* (2006.01),
C07K 14/40 (2006.01), *C07K 14/47* (2006.01),
C07K 16/18 (2006.01), *C07K 19/00* (2006.01),
C12N 15/12 (2006.01), *C12N 15/54* (2006.01),
C12N 9/00 (2006.01), *C12N 9/12* (2006.01),
C12Q 1/25 (2006.01), *C12Q 1/68* (2006.01),
G01N 33/566 (2006.01), *G01N 33/68* (2006.01)

(72) Inventeurs/Inventors:
BERLIN, VIVIAN, US;
CHIU, MARIE ISABEL, US;
COTTAREL, GUILLAUME, US;
DAMAGNEZ, VERONIQUE, US

(73) Propriétaire/Owner:

(54) Titre : PROTEINES CIBLES IMMUNODEPRESSIVES

(54) Title: IMMUNOSUPPRESSANT TARGET PROTEINS



(57) Abrégé/Abstract:

The present invention relates to the discovery of novel proteins of mammalian origin which are immediate downstream targets for FKBP/rapamycin complexes.

(73) Propriétaires(suite)/Owners(continued):ARIAD GENE THERAPEUTICS, INC., US

(74) Agent: BORDEN LADNER GERVAIS LLP

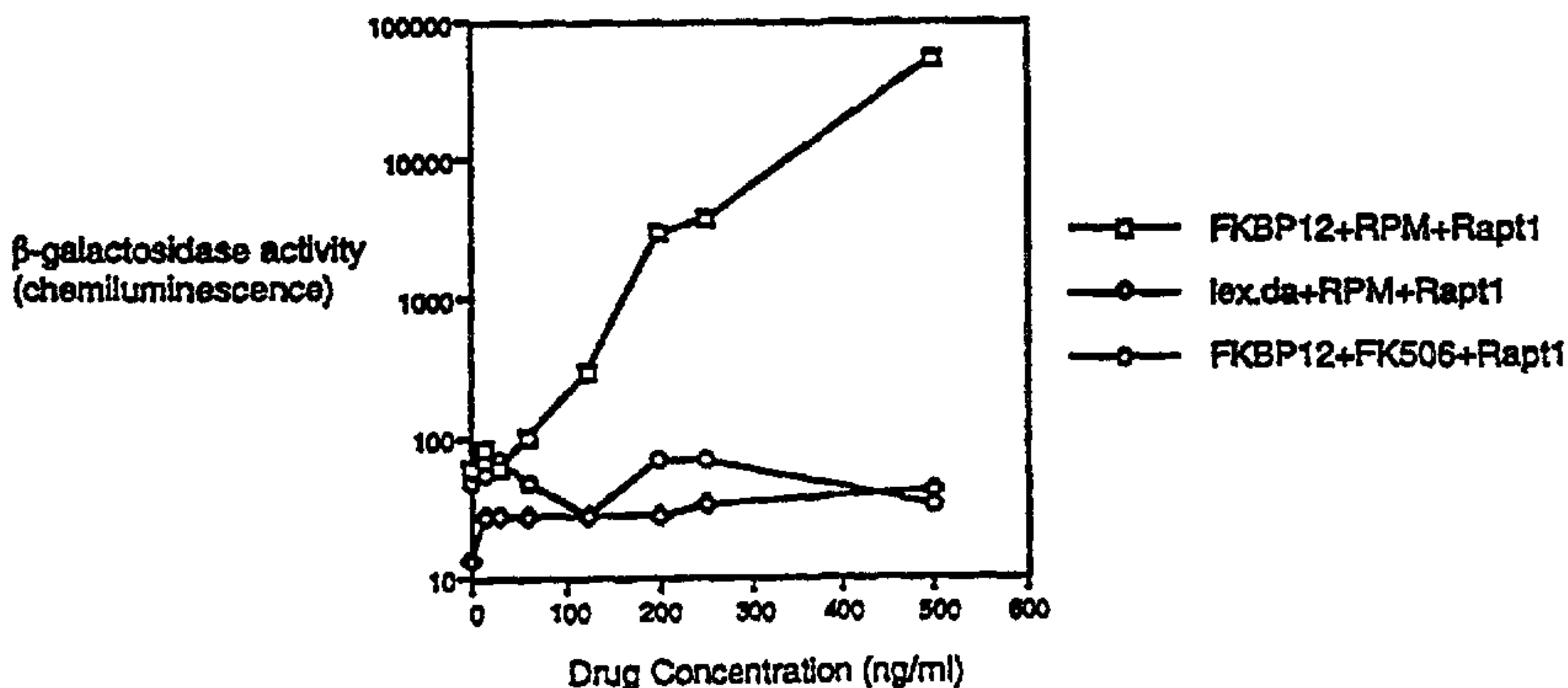


INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 15/12, 15/31, 15/62, 15/81, 1/19, C07K 14/47, 14/40, 16/18, C12Q 1/68, G01N 33/53	A1	(11) International Publication Number: WO 95/33052 (43) International Publication Date: 7 December 1995 (07.12.95)
(21) International Application Number: PCT/US95/06722 (22) International Filing Date: 30 May 1995 (30.05.95) (30) Priority Data: 08/250,795 27 May 1994 (27.05.94) US 08/360,144 20 December 1994 (20.12.94) US (71) Applicant: MITOTIX, INC. [US/US]; Building 600, One Kendall Square, Cambridge, MA 02139 (US). (72) Inventors: BERLIN, Vivian; 634 Main Street, Dunstable, MA 01827 (US). CHIU, Marie, Isabel; Apartment #3, 288 Commonwealth Avenue, Boston, MA 02109 (US). COTTAREL, Guillaume; 15 Rickerhill Road, West Roxbury, MA 02132 (US). DAMAGNEZ, Veronique; 119 Sciarappa Street, #3, Cambridge, MA 02141 (US). (74) Agents: VINCENT, Matthew, P. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).		(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>

2188061

(54) Title: IMMUNOSUPPRESSANT TARGET PROTEINS



(57) Abstract

The present invention relates to the discovery of novel proteins of mammalian origin which are immediate downstream targets for FKBP/rapamycin complexes.

Immunosuppressant Target Proteins

Background of the Invention

Cyclosporin A, FK506, and rapamycin are microbial products with potent
5 immunosuppressive properties that result primarily from a selective inhibition of T
lymphocyte activation. Rapamycin was first described as an antifungal antibiotic extracted
from a streptomycete (*Streptomyces hygroscopicus*) (Vezina et al. (1975) *J. Antibiot.*, 28:721;
Sehgal et al. (1975) *J. Antibiot.* 28:727; and Sehgal et al., U.S. Pat. No. 3,929,992).
Subsequently, the macrolide drug rapamycin was shown to exhibit immunosuppressive as
10 well as antineoplastic and antiproliferative properties (Morris (1992) *Transplant Res* 6:39-
87).

Each of these compounds, cyclosporin A, FK506 and rapamycin, suppress the
immune system by blocking distinctly different biochemical reactions which would
ordinarily initiate the activation of immune cells. Briefly, cyclosporin A and FK506 act soon
15 after Ca^{2+} -dependent T-cell activation to prevent the synthesis of cytokines important for the
perpetuation and amplification of the immune response. Rapamycin acts later to block
multiple affects of cytokines on immune cells including the inhibition of interleukin-2 (IL2)-
triggered T-cell proliferation, but its antiproliferative effects are not restricted solely to T and
B cells. Rapamycin also selectively inhibits the proliferation of growth factor-dependent and
20 growth factor-independent nonimmune cells. Rapamycin is generally believed to inhibit cell
proliferation by blocking specific signaling events necessary for the initiation of S phase in a
number of cell types, including lymphocytes (Bierer et al. (1990) *PNAS* 87:9231-9235; and
Dumont et al. (1990) *J. Immunol* 144:1418-1424), as well as non-immune cells, such as
hepatocytes (Francavilla et al. (1992) *Hepatology* 15:871-877; and Price et al. (1992) *Science*
25 257:973-977). Several lines of evidence suggest that the association of rapamycin with
different members of a family of intracellular FK506/rapamycin binding proteins (FKBPs) is
necessary for the inhibition of G_1 progression as mediated by rapamycin. For instance, the
actions of rapamycin are reversed by an excess of the structurally FKBP-ligands FK506 or
506BD (Bierer et al. *supra.*; Dumont et al. *supra.*; and Bierer et al. (1990) *Science* 250:556-
30 559).

Cyclosporin A binds to a class of proteins called cyclophilins (Walsh et al. (1992) *J.*
Biol. Chem. 267:13115-13118), whereas the primary targets for both FK506 and rapamycin,
as indicated above, are the FKBPs (Harding et al. (1989) *Nature* 341:758-7601; Siekienka et
al. (1989) *Nature* 341:755-757; and Soltoff et al. (1992) *J. Biol. Chem.* 267:17472-17477).
35 Both the cyclophilin/cyclosporin and FKBP12/FK506 complexes bind to a specific protein
phosphatase (calcineurin) which is hypothesized to control the activity of IL-2 gene specific

transcriptional activators (reviewed in Schreiber (1991) *Cell* 70:365-368). In contrast, the downstream cellular targets for the rapamycin-sensitive signaling pathway have not been especially well characterized, particularly with regard to the identity of the direct target of the FKBP-rapamycin complex.

5 The TOR1 and TOR2 genes of *S. cerevisiae* were originally identified by mutations that rendered cells resistant to rapamycin (Heitman et al. (1991) *Science* 253:905-909) and there was early speculation that the FKBP/rapamycin complex might inhibit the cellular function of the TOR gene product by binding directly to a phosphoserine residue of either TOR1 or TOR2. Subsequently, however, new models for rapamycin drug interaction have
10 been proposed which do not involve direct binding of the FKBP/rapamycin complex to the TOR proteins. For example, based on experimental data regarding cyclin-cdk activity in rapamycin treated cells, the Schreiber laboratory wrote in Albers et al. (1993) *J. Biol. Chem.* 268:22825-22829:

15 "Although it is possible the TOR2 gene product is a direct target of the FKBP-rapamycin complex, a more likely explanation is that the TOR2 gene product lies downstream of the direct target of rapamycin and that the TOR2 mutation caused the protein to be constitutively active. If the latter model is correct, then the TOR2 gene product joins p70^{s6k},
20 cyclin-dependent kinases, and cyclin D1 as proteins that lie downstream of the direct target of the FKBP-rapamycin complex and have been shown to play important roles in cell cycle progression. The identification of the direct target of the FKBP-rapamycin complex will likely reveal an upstream
25 component of the signal transduction pathway that leads to G1 progression and will help delineate the signal transduction pathways that link growth factor-mediated signaling events and cyclin-cdk activity required for cell cycle progression."

 Likewise, after studying the role of TOR1 and TOR2 mutations in rapamycin-
30 resistant yeast cells, Livi group wrote in Cafferkey et al. (1993) *Mol. Cell Biol.* 13:6012-6023:

35 "Thus, the amino acid changes that we have identified in the rapamycin-released DRR1 [TOR1] protein may allow it to compensate for the loss of the proliferative signal inhibited by rapamycin by constitutively activating an alternative signal rather than by preventing its association with the FKBP12-
40 rapamycin complex. The positions of the mutations within the kinase domain, but in a region not shared by the PI 3-kinases, support this idea. Therefore, it is entirely possible that DRR1 is not a component of the rapamycin-sensitive pathway in wild-type yeast cells. Instead, missense mutations in DRR1 at Ser-

WO 95/33052

PCT/US95/06722

3

1972 may alter its normal activity and allow it to substitute for the function of an essential protein which is the true target of rapamycin."

5 It is an object of the present invention to identify cellular proteins which are the direct downstream target proteins for the FKBP/rapamycin complex, and isolate the genes encoding those proteins.

Summary of the Invention

10 The present invention relates to the discovery of novel proteins of mammalian origin which are immediate downstream targets for FKBP/rapamycin complexes. As described herein, a drug-dependent interaction trap assay was used to isolate a number of proteins which interact with an FK506-binding protein/rapamycin complex, and which are collectively referred to herein as "RAP-binding proteins" or "RAP-BPs". In particular, several mammalian genes (orthologs) have been cloned for a protein referred to herein as "RAPT1", which protein is apparently related to the yeast TOR1 and TOR2 gene products.
15 Furthermore, a novel ubiquitin-conjugating enzyme, referred to herein as "rap-UBC", has been cloned based on its ability to bind FKBP/rapamycin complexes. In addition, a RAPT1-like protein was cloned from the human pathogen *Candida*. The present invention, therefore, makes available novel proteins (both recombinant and purified forms), recombinant genes, antibodies to RAP-binding proteins, and other novel reagents and assays for diagnostic and
20 therapeutic use.

The present invention relates to the discovery in eukaryotic cells, particularly human cells, of novel protein-protein interactions between the FK506-binding protein/rapamycin complexes and certain cellular proteins, referred to hereinafter as "RAP-binding proteins" or "RAP-BP".

25 The invention provides an isolated nucleic acid comprising a nucleotide sequence which encodes a polypeptide including an FKBP/rapamycin binding domain that specifically binds to an FKBP/rapamycin complex, which FKBP/rapamycin binding domain comprises an amino acid sequence which is (a) at least 70% homologous to a sequence in SEQ ID NO: 2 or 12, or (b)
30 encoded by a nucleotide sequence which hybridizes under stringency conditions of 0.2 x SSC 65°C to SEQ ID NO: 1 or 11.

The invention also provides a nucleic acid construct comprising
35 (i) a coding sequence for a polypeptide including an FKBP/rapamycin binding domain that specifically binds to an FKBP/rapamycin complex, which FKBP/rapamycin binding domain comprises an amino acid sequence which is (a) at least 70%

3a

homologous to a sequence in SEQ ID NO: 2 or 12, or (b) encoded by a nucleotide sequence which hybridizes under stringency conditions of 0.2 x SSC 65°C to SEQ ID NO: 1 or 11; and

- (ii) a transcriptional regulatory sequence operably linked to the coding sequence and which is heterologous to the coding sequence.

The invention further provides a nucleic acid construct comprising a coding sequence for a fusion protein comprising

- (i) a portion of a RAPT1 protein which can be encoded by a nucleic acid which hybridizes to one of SEQ ID NOs. 1, 11, 13, 15 or 17 under stringent conditions including a wash step of 2.0 x SSC at 22°C, and which portion binds an FKBP/rapamycin complex in a rapamycin dependent manner; and
- (ii) one or more additional polypeptide portions which do not occur in a native RAPT1 protein.

The invention additionally provides a kit comprising

- (i) a first coding sequence for a first fusion protein including a rapamycin binding domain of an FK506 binding protein (FKBP); and
- (ii) a second coding sequence for a second fusion protein including an FKBP/rapamycin binding domain that specifically binds to an FKBP/rapamycin complex;

wherein one of the first or second fusion proteins includes a polymerase activation domain and the other includes a DNA binding domain.

The invention further comprises a nucleic acid encoding a fusion polypeptide consisting of: a portion of a RAPT1 protein of 200 amino acid residues or less, which RAPT1 portion (a) is encoded by a nucleic acid which hybridizes to one of SEQ ID NOs. 1 or 11 under stringent conditions including a wash step of 0.2x SSC at 65 °C., and (b) binds an FKBP/rapamycin complex in a rapamycin dependent manner; and one or more additional polypeptide portions heterologous to a native RAPT1 protein.

The invention additionally comprises a nucleic acid encoding a fusion polypeptide consisting essentially of: a portion of a RAPT1 protein of 200 amino acid residues or less, which RAPT1 portion (a) is encoded by a nucleic acid which hybridizes to SEQ ID NO. 1 or 11 under stringent conditions including a wash step of 0.2x. SSC at 65 °C.,

3b

and (b) binds an FKBP/rapamycin complex in a rapamycin dependent manner; wherein the RAPT1 portion does not include an endogenous kinase activity; and one or more additional polypeptide portions heterologous to a native RAPT1 protein.

5 In general, the invention features a mammalian RAPT1 polypeptide, preferably a substantially pure preparation of a RAPT1 polypeptide, or a recombinant RAPT1 polypeptide. In preferred embodiments the polypeptide has a biological activity associated with its binding to rapamycin, e.g., it retains the ability to bind to an FKBP/rapamycin complex, though it may be able to either agonize or antagonize assembly of rapamycin-dependent complexes. The polypeptide
10 can be identical to a polypeptide shown in one of SEQ ID No: 2 or 12, or it can merely be homologous to that sequence. For instance, the polypeptide preferably has an amino acid sequence at least 70% homologous to the amino acid sequence of at least one of either SEQ ID No: 2 or 12, though higher sequence homologies of, for example, 80%, 90% or 95% are also contemplated, and will generally be preferred. The polypeptide can comprise the full length protein, or a portion
15 of a full length protein, such as the RAPT1 polypeptides represented in either SEQ ID No: 2 or 12, or an even

4

smaller fragment of that protein. which fragment may be, for instance, at least 5, 10, 20, 50, 100, or 150 amino acids in length. As described below, the RAPT1 polypeptide can be either an agonist (e.g. mimics), or alternatively, an antagonist of a biological activity of a naturally occurring form of the protein, e.g., the polypeptide is able to modulate assembly of rapamycin complexes, such as complexes involving FK506-binding proteins, or cell cycle regulatory proteins.

In a preferred embodiment, a peptide having at least one biological activity of the subject RAPT1 polypeptides may differ in amino acid sequence from the sequence in SEQ ID No: 2 or 12, but such differences result in a modified protein which functions in the same or similar manner as the native RAPT1 protein or which has the same or similar characteristics of the native RAPT1 protein. However, homologs of the naturally occurring protein are contemplated which are antagonistic of the normal cellular role of the naturally occurring protein.

In yet other preferred embodiments, the RAPT1 protein is a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated to the RAPT1 polypeptide portion, e.g. the second polypeptide portion is glutathione-S-transferase, e.g. the second polypeptide portion is a DNA binding domain of transcriptional regulatory protein, e.g. the second polypeptide portion is an RNA polymerase activating domain, e.g. the fusion protein is functional in a two-hybrid assay.

Yet another aspect of the present invention concerns an immunogen comprising a RAPT1 peptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the RAPT1 polypeptide; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g. a unique determinant, from a protein represented by SEQ ID No: 2 and/or 12.

A still further aspect of the present invention features an antibody preparation specifically reactive with an epitope of the RAPT1 immunogen.

In still another aspect, the invention features a RAPT1-like polypeptide from a *Candida* species (caRAPT1), preferably a substantially pure preparation of a caRAPT1 polypeptide, or a recombinant caRAPT1 polypeptide. As above, in preferred embodiments the caRAPT1 polypeptide has a biological activity associated with its binding to rapamycin, e.g., it retains the ability to bind to a rapamycin complex, such as an FKBP/rapamycin complex. The polypeptide can be identical to the polypeptide shown in SEQ ID No: 14, or it can merely be homologous to that sequence. For instance, the caRAPT1 polypeptide preferably has an amino acid sequence at least 60% homologous to the amino acid sequence

in SEQ ID No: 14, though higher sequence homologies of, for example, 80%, 90% or 95% are also contemplated. The caRAP1 polypeptide can comprise the entire polypeptide represented in SEQ ID No: 14, or it can comprise a fragment of that protein, which fragment may be, for instance, at least 5, 10, 20, 50 or 100 amino acids in length. The caRAP1 polypeptide can be either an agonist (e.g. mimics), or alternatively, an antagonist of a biological activity of a naturally occurring form of the protein.

In a preferred embodiment, a peptide having at least one biological activity of the subject caRAP1 polypeptide may differ in amino acid sequence from the sequence in SEQ ID No: 14, but such differences result in a modified protein which functions in the same or similar manner as the native caRAP1 or which has the same or similar characteristics of the native protein. However, homologs of the naturally occurring caRAP1 protein are contemplated which are antagonistic of the normal cellular role of the naturally occurring protein.

In yet other preferred embodiments, the caRAP1 protein is a recombinant fusion protein which includes a second polypeptide portion, e.g., a second polypeptide having an amino acid sequence unrelated to the caRAP1 sequence, e.g. the second polypeptide portion is glutathione-S-transferase, e.g. the second polypeptide portion is a DNA binding domain of transcriptional regulatory protein, e.g. the second polypeptide portion is an RNA polymerase activating domain, e.g. the fusion protein is functional in a two-hybrid assay.

Yet another aspect of the present invention concerns an immunogen comprising a caRAP1 peptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the caRAP1 polypeptide; e.g. a humoral response, e.g. an antibody response; e.g. a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant, e.g. a unique determinant, from a protein represented by SEQ ID No: 14.

A still further aspect of the present invention features an antibody preparation specifically reactive with an epitope of the caRAP1 immunogen.

Still another embodiment of the present invention features fragments of a RAP1, e.g., hRAP1 or mRAP1, or other RAP1-like polypeptide, e.g., caRAP1, TOR1 or TOR2, which fragments retaining the ability to bind to an FK-binding protein in a rapamycin dependent manner. Accordingly, the present invention facilitates the generation of drug screening assays, particularly the high-throughout assays described below, for the identification immunosuppressants, anti-mycotic agents, and the like which act through the binding of the rapamycin-binding domain of the RAP1-like proteins. For instance, the present invention provides portions of the RAP1-like proteins which are easier to manipulate than the full length protein. The full length protein is, because of its size, more

6

difficult to express as a recombinant protein or a fusion protein which would retain rapamycin-binding activity, and may very well be insoluble. Accordingly, the present invention provides soluble polypeptides which include a soluble portion of a RAPT1-like polypeptide that binds to said FKBP/rapamycin complex, such as the rapamycin-binding domain represented by an amino acid sequence selected from the group consisting Val26-Tyr160 of SEQ ID No. 2 (mRATP1), Val2012-Tyr2144 of SEQ ID No. 12 (hRAPT1), Val41-Tyr173 of SEQ ID No. 14 (caRAPT1), Val1-Tyr133 of SEQ ID No. 16 (TOR1), and Val1-Arg133 of SEQ ID No. 18 (TOR2).

Another aspect of the present invention provides a substantially isolated nucleic acid having a nucleotide sequence which encodes a RAPT1 polypeptide. In preferred embodiments: the encoded polypeptide specifically binds a rapamycin complexes and/or is able to either agnoize or antagonize assembly of rapamycin-containing protein complexes. The coding sequence of the nucleic acid can comprise a RAPT1-encoding sequence which can be identical to the cDNA shown in SEQ ID No: 1 or 11, or it can merely be homologous to that sequence. For instance, the RAPT1-encoding sequence preferably has a sequence at least 70% homologous to one or both of the nucleotide sequences in SEQ ID No: 1 or 11, though higher sequence homologies of, for example, 80%, 90% or 95% are also contemplated. The nucleic acid can comprise the nucleotide sequence represented in SEQ ID No: 1, or it can comprise a fragment of that nucleic acid, which fragment may be, for instance, encode a fragment of which is, for example, at least 5, 10, 20, 50, 100 or 133 amino acids in length. The polypeptide encoded by the nucleic acid can be either an agonist (e.g. mimics), or alternatively, an antagonist of a biological activity of a naturally occurring form of the RAPT1 protein, e.g., the polypeptide is able to modulate rapamycin-mediated protein complexes.

Furthermore, in certain preferred embodiments, the subject RAPT1 nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, which regulatory sequence is operably linked to the RAPT1 gene sequence. Such regulatory sequences can be used in to render the RAPT1 gene sequence suitable for use as an expression vector.

In yet a further preferred embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No: 1 and/or 11; preferably to at least 20 consecutive nucleotides, and more preferably to at least 40 consecutive nucleotides. In yet another embodiment, the nucleic acid hybridizes to region of the human or mouse RAPT1 genes corresponding to the binding domain for rapamycin.

Another aspect of the present invention provides a substantially isolated nucleic acid having a nucleotide sequence which encodes a caRAPT1 polypeptide. In preferred embodiments: the encoded polypeptide specifically binds a rapamycin complexes and/or is able to either agnoize or antagonize assembly of rapamycin-containing protein complexes.

5 The coding sequence of the nucleic acid can comprise a caRAPT1-encoding sequence which can be identical to the cDNA shown in SEQ ID No: 13, or it can merely be homologous to that sequence. For instance, the caRAPT1-encoding sequence preferably has a sequence at least 60% homologous to the nucleotide sequences in SEQ ID No: 13, though higher sequence homologies of, for example, 80%, 90% or 95% are also contemplated. The nucleic
10 acid can comprise the nucleotide sequence represented in SEQ ID No: 13, or it can comprise a fragment of that nucleic acid, which fragment may be, for instance, encode a fragment of which is, for example, at least 5, 10, 20, 50, 100 or 140 amino acids in length. The polypeptide encoded by the nucleic acid can be either an agonist (e.g. mimics), or alternatively, an antagonist of a biological activity of a naturally occurring form of the
15 caRAPT1 protein, e.g., the polypeptide is able to modulate rapamycin-mediated protein complexes.

Furthermore, in certain preferred embodiments, the subject caRAPT1 nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, which regulatory sequence is operably linked
20 to the caRAPT1 gene sequence. Such regulatory sequences can be used in to render the caRAPT1 gene sequence suitable for use as an expression vector.

In yet a further preferred embodiment, the nucleic acid hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID No: 13; preferably to at least 20 consecutive nucleotides, and more preferably to at
25 least 40 consecutive nucleotides.

The invention also features transgenic non-human animals, e.g. mice, rats, rabbits or pigs, having a transgene, e.g., animals which include (and preferably express) a heterologous form of one of the RAP-BP genes described herein, e.g. a gene derived from humans, or which misexpress an endogenous RAP-BP gene, e.g., an animal in which expression of one
30 or more of the subject RAP-binding proteins is disrupted. Such a transgenic animal can serve as an animal model for studying cellular disorders comprising mutated or mis-expressed RAP-BP alleles or for use in drug screening.

The invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence
35 which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of one of SEQ ID Nos: 1, 11 or 13, or naturally occurring mutants

8

thereof. In preferred embodiments, the probe/primer further includes a label group attached thereto and able to be detected. The label group can be selected, e.g., from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Probes of the invention can be used as a part of a diagnostic test kit for identifying transformed cells, such as for detecting in a sample of cells isolated from a patient, a level of a nucleic acid encoding one of the subject RAP-binding proteins; e.g. measuring the RAP-BP mRNA level in a cell, or determining whether the genomic RAP-BP gene has been mutated or deleted. Preferably, the oligonucleotide is at least 10 nucleotides in length, though primers of 20, 30, 50, 100, or 150 nucleotides in length are also contemplated.

10 In yet another aspect, the invention provides assay systems for screening test compounds for an molecules which induce an interaction between a RAP-binding protein and a rapamycin/protein complexes. An exemplary method includes the steps of (i) combining a RAP-binding protein of the invention, an FK506-binding protein, and a test compound, e.g., under conditions wherein, but for the test compound, the FK506-binding protein and the RAP-binding protein are unable to interact; and (ii) detecting the formation of a drug-dependent complex which includes the FK506-binding protein and the RAP-binding protein. A statistically significant change, such as an increase, in the formation of the complex in the presence of a test compound (relative to what is seen in the absence of the test compound) is indicative of a modulation, e.g., induction, of the interaction between the FK506-binding protein and the RAP-binding protein. Moreover, primary screens are provided in which the FK506-binding protein and the RAP-binding protein are combined in a cell-free system and contacted with the test compound; i.e. the cell-free system is selected from a group consisting of a cell lysate and a reconstituted protein mixture. Alternatively, FK506-binding protein and the RAP-binding protein are simultaneously expressed, e.g., recombinantly, in a cell, and the cell is contacted with the test compound, e.g. as an interaction trap assay (two hybrid assay).

The present invention also provides a method for treating an animal having unwanted cell growth characterized by a loss of wild-type function of one or more of the subject RAP-binding proteins, comprising administering a therapeutically effective amount of an agent able to inhibit the interaction of the RAP-binding protein with other cellular or viral proteins. In one embodiment, the method comprises administering a nucleic acid construct encoding a polypeptides represented in one of SEQ ID Nos: 2 or 12, under conditions wherein the construct is incorporated by cells deficient in that RAP-binding protein, and under conditions wherein the recombinant gene is expressed, e.g. by gene therapy techniques. In other embodiments, the action of a naturally-occurring RAP-binding protein is antagonized by therapeutic expression of a RAP-BP homolog which is an antagonist of, for example, assembly of rapamycin-mediated complexes, or by delivery of an antisense nucleic acid molecule which inhibits transcription and/or translation of the targeted RAP-BP gene.

Another aspect of the present invention provides a method of determining if a subject, e.g. a human patient, is at risk for a disorder characterized by unwanted cell proliferation. The method includes detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a protein represented by one of SEQ ID Nos: 1 or 11, or a homolog thereof; (ii) the mis-expression of a gene encoding a protein represented by one of SEQ ID Nos: 1 or 11; or (iii) the mis-incorporation of a RAP-binding protein in a regulatory protein complex, e.g. a rapamycin-containing complex. In preferred embodiments: detecting the genetic lesion includes ascertaining the existence of at least one of: a deletion of one or more nucleotides from the RAP-BP gene; an addition of one or more nucleotides to the gene, an substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of the protein.

For example, detecting the genetic lesion can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of one of SEQ ID Nos: 1 or 11, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the RAP-BP gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the RAP-BP gene and, optionally, of the flanking nucleic acid sequences. For instance, the probe/primer can be employed in a polymerase chain reaction (PCR) or in a ligation chain reaction (LCR). In alternate embodiments, the level of the RAP-binding protein is detected in an immunoassay using an antibody which is specifically immunoreactive with a protein represented by one of SEQ ID Nos: 1 or 11.

In similar fashion, *Candida* infection can be detected by use of probes/primers which hybridize to a *Candida* gene encoding a RAPT1-like protein. For instance, the method can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of one of SEQ ID No: 13, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the caRAPT1 gene; (ii) exposing the probe/primer to nucleic acid of a biological sample, e.g., tissue biopsy, fluid sample, stool, etc.; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of a *Candida* organism.

Another aspect of the present invention concerns a novel *in vivo* method for the isolation of genes encoding proteins which physically interact with a "bait" protein/drug complex. The method relies on detecting the reconstitution of a transcriptional activator in the presence of the drug, particularly wherein the drug is a non-peptidyl small organic

10 molecule (e.g. <2500K), e.g. a macrolide, e.g. rapamycin, FK506 or cyclosporin. In particular, the method makes use of chimeric genes which express hybrid proteins. The first hybrid comprises the DNA-binding domain of a transcriptional activator fused to the bait protein. The second hybrid protein contains a transcriptional activation domain fused to a "fish" protein, e.g. a test protein derived from a cDNA library. If the fish and bait proteins are able to interact in a drug-dependent manner, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the marker gene can be detected and used to score for the interaction of the bait protein/drug complex with another protein.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Description of the Figures

Figure 1 illustrates the map of the pACT vector used to clone the human RAPT1 clone. The RAPT1-containing version of pACT, termed "pIC524" has been deposited with the ATCC.

Figure 2 illustrates the interaction of FKBP12 and hRAPT1 (rapamycin-binding domain) as a function of rapamycin concentration. Interaction is detected as β -galactosidase

11

activity. No interaction is detected if FK506 is used in place of rapamycin, or if lex.da (a control plasmid) replaces FKBP12.

Figure 3 illustrates the relative strengths of interaction between pairs of FK506-binding proteins and rapamycin-binding domain (BD) fusions in the presence of varying concentrations of rapamycin, measured by β -galactosidase expression (see Example 8). The yeast reporter strain VBY567 was transformed with the indicated pairs of plasmids. LexA DNA-binding domain fusions to human FKBP12, yeast FKBP12 and an unrelated sequence serving as negative control were used as "baits". The VP16 acidic activation domain fusions to human RAPT1 BD, human RAPT1 BD containing the serine to arginine substitution, yeast Tor1 BD, yeast Tor2 BD (not shown) and *Candida albicans* RAPT1 BD were tested for interaction against the bait fusions. Transformants containing each pair of plasmids were tested for β -galactosidase expression on media containing the chromogenic substrate X-gal. Colonies were scored as either white (open bars) or blue (solid bars) after growth at 30°C for 2 days. The levels of β -galactosidase expression were qualitatively scored by the intensity of the blue color, ranging from 1 (light blue) to 4 (deep blue).

Detailed Description of the Invention

Recent studies have provided some remarkable insights into the molecular basis of eukaryotic cell cycle regulation. Passage of a mammalian cell through the cell cycle is regulated at a number of key control points. Among these are the points of entry into and exit from quiescence (G_0), the restriction point, the G_1/S transition, and the G_2/M transition (for review, see Draetta (1990) *Trends Biol Sci* 15:378-383; and Sherr (1993) *Cell* 73:1059-1065). Ultimately, information from these check-point controls is integrated through the regulated activity of a group of related kinases, the cyclin-dependent kinases (CDKs). For example, the G_1 -to-S phase transition is now understood to be timed precisely by the transient assembly of multiprotein complexes involving the periodic interaction of a multiplicity of cyclins and cyclin-dependent kinases.

To illustrate, stimulation of quiescent T lymphocytes by cell-bound antigens triggers a complex activation program resulting in cell cycle entry (G_0 -to- G_1 transition) and the expression of high affinity interleukin-2 (IL-2) receptors. The subsequent binding of IL-2 to its high affinity receptor drives the progression of activated T cells through a late G_1 -phase "restriction point" (Pardee (1989) *Science* 246:603-608), after which the cells are committed to complete a relatively autonomous program of DNA replication and, ultimately, mitosis.

One important outcome of the information concerning eukaryotic cell cycle regulation is the delineation of a novel class of molecular targets for potential growth-modulatory drugs. The macrolide ester, rapamycin, is a potent immunosuppressant whose mechanism of action is related to the inhibition of cytokine-dependent T cell proliferation (Bierer et al. (1990) *PNAS* 87:9231-9235; Dumont et al. (1990) *J. Immunol* 144:1418-1424; Sigal et al. (1991) *Transplant Proc* 23:1-5; and Sigal et al. (1992) *Annu Rev Immunol* 10:519-560). Rapamycin specifically interferes with a late G₁-phase event required for the progression of IL-2 stimulated cells into S-phase (Morice et al. (1993) *J Biol Chem* 268:3734-3738). The location of the cell cycle arrest point induced by rapamycin hints that this drug interferes with the regulatory proteins that govern the G₁-to-S phase transition, particularly in lymphocytes.

As described herein, the present invention relates to the discovery of novel proteins of mammalian origin which are immediate downstream targets for FKBP/rapamycin complexes. As described below, a drug-dependent interaction trap assay was used to isolate a number of proteins which bind the FKBP12/rapamycin complex, and which are collectively referred to herein as "RAP-binding proteins" or "RAP-BPs". In particular, mouse and human genes have been cloned for a protein (referred to herein as "RAPT1") which is apparently related to the yeast TOR1 and TOR2 gene products. Furthermore, a novel ubiquitin-conjugating enzyme (referred to herein as "rap-UBC") has been cloned based on its ability to bind FKBP/rapamycin complexes. The present invention, therefore, makes available novel proteins (both recombinant and purified forms), recombinant genes, antibodies to RAP-binding proteins, and other novel reagents and assays for diagnostic and therapeutic use. Moreover, drug discovery assays are provided for identifying agents which can modulate the binding of one or more of the subject RAP-binding proteins with FK506-binding proteins. Such agents can be useful therapeutically to alter the growth and/or differentiation of a cell, but can also be used *in vitro* as cell-culture additives for controlling proliferation and/or differentiation of cultured cells and tissue. Other aspects of the invention are described below or will be apparent to those skilled in the art in light of the present disclosure.

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

The term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a RAP-binding protein of the present invention, including both exon

and (optionally) intron sequences. A "recombinant gene" refers to nucleic acid encoding a RAP-binding protein and comprising RAP-BP encoding exon sequences, though it may optionally include intron sequences which are either derived from a chromosomal RAP-BP gene or from an unrelated chromosomal gene. Exemplary recombinant genes encoding
5 illustrative RAP-binding proteins include a nucleic acid sequence represented by one of SEQ ID Nos: 1, 11, 13 or 23. The term "intron" refers to a DNA sequence present in a given RAP-BP gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" refers to the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer.
10 "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of the RAP-binding protein of the present invention or where anti-sense expression occurs from the transferred gene, the expression for a naturally-occurring form of the RAP-binding protein is disrupted.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively
20 linked are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such
25 other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are
30 operably linked. In preferred embodiments, transcription of a recombinant RAP-BP gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those
35 sequences which control transcription of the naturally-occurring form of the RAP-binding protein.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of a lymphoid lineage, e.g. B or T lymphocytes, or alternatively, e.g. hepatic cells. In an illustrative embodiment, gene constructs utilizing lymphoid-specific promoters can be used as a part of gene therapy to provide dominant negative mutant forms of a RAP-binding protein to render lymphatic cells resistant to rapamycin by directing expression of the mutant form of RAP-BP in only lymphatic tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, a bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of a subject RAP-binding protein, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant RAP-BP gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the *Xenopus* genus, and transgenic chickens can also provide important tools for understanding, for example, embryogenesis and tissue patterning. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that the recombinant RAP-BP gene is present and/or expressed in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a RAP-binding protein), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into

which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

5 As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a RAP-binding protein" may thus refer to one or more genes within a
10 particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" refers to sequence similarity between two peptides or between two
15 nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

20 "Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the
25 term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding one of the subject RAP-binding proteins with a second amino acid sequence defining a domain foreign to and not substantially homologous with any domain of the subject RAP-BP. A chimeric protein may present a foreign domain which is found (albeit in a
30 different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergeneric", etc. fusion of protein structures expressed by different kinds of organisms. For example, a fusion protein of the present invention can be represented by the general formula Z_1 - Z_2 - Z_3 , wherein Z_2 represents all or a portion of a polypeptide sequence of a RAP-binding protein, and Z_1 and Z_3 each represent polypeptide sequences which are
35 heterologous to the RAP-BP sequence, at least one of Z_1 and Z_3 being present in the fusion protein.

16

The term "evolutionarily related to", with respect to nucleic acid sequences encoding RAP-binding proteins, refers to nucleic acid sequences which have arisen naturally in an organism, including naturally occurring mutants. Moreover, the term also refers to nucleic acid sequences which, while initially derived from naturally-occurring isoforms of RAP-binding proteins, have been altered by mutagenesis, as for example, such combinatorial mutagenesis as described below, yet which still encode polypeptides that bind FKBP/rapamycin complexes, or that retain at least one activity of the parent RAP-binding protein, or which are antagonists of that protein's activities.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, an isolated nucleic acid encoding one of the subject RAP-binding proteins preferably includes no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks that particular RAP-BP gene in genomic DNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

As used herein, an "rapamycin-binding domain" refers to a polypeptide sequence which confers a binding activity for specifically interacting with an FKBP/rapamycin complex. Exemplary rapamycin-binding domains are represented within the polypeptides defined by Val26-Tyr160 of SEQ ID No. 2 (mRAPT1), Val2012-Tyr2144 of SEQ ID No. 12 (hRAPT1), Val41-Tyr173 of SEQ ID No. 14 (caRAPT1), Val1-Tyr133 of SEQ ID No. 16 (TOR1), and Val1-Arg133 of SEQ ID No. 18 (TOR2).

A "RAPT1-like polypeptide" refers to a eukaryotic cellular protein which is a direct binding target protein for an FKBP/rapamycin complex, and which shares some sequence homology with a mammalian RAPT1 protein of the present invention. Exemplary RAPT1-like polypeptides include the yeast TOR1 and TOR2 proteins.

A "soluble protein" refers to a polypeptide which does not precipitate (e.g. at least about 95-percent, more preferably at least 99-percent remains in the supernatant) from an aqueous buffer under physiologically isotonic conditions, as for example, 0.14M NaCl or sucrose, at a protein concentration of as much as 10 μ M, more preferably as much as 10 mM. These conditions specifically relate to the absence of detergents or other denaturants in effective concentrations.

As described below, one aspect of this invention pertains to an isolated nucleic acid comprising the nucleotide sequence encoding a RAP-binding protein, fragments thereof, and/or equivalents of such nucleic acids. The term nucleic acid as used herein is intended to include such fragments and equivalents, e.g., the term equivalent is understood to include
5 nucleotide sequences encoding functionally equivalent RAP-binding proteins or functionally equivalent peptides which, for example, retain the ability to bind to the FKBP/rapamycin complex, and which may additionally retain other activities of a RAP-binding protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will also
10 include sequences that differ from the nucleotide sequence of the mammalian RAPT1 genes represented in SEQ ID No: 1 or SEQ ID No. 11, or the nucleotide sequence of the fungal RAPT1 protein of SEQ ID No. 13, or the nucleotide sequence encoding the UBC enzyme represented in SEQ ID No. 23, due to the degeneracy of the genetic code. Equivalent nucleic acids will also include nucleotide sequences that hybridize under stringent conditions (i.e.,
15 equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to a nucleotide sequence of a RAPT1 protein comprising either the sequence shown in SEQ ID No: 2 or 12, or to a nucleotide sequence of the RAPT1 gene insert of pIC524 (ATCC accession no. 75787). Likewise, equivalent nucleic acids encoding homologs of the subject rap-UBC enzyme include nucleotide sequences that hybridize under
20 stringent conditions to a nucleotide sequence represented in SEQ ID No. 23, or to a nucleotide sequence of the rap-UBC gene insert of SMR4-15 (ATCC accession no. 75786). In one embodiment, equivalents will further include nucleic acid sequences derived from, and evolutionarily related to, a nucleotide sequence comprising that shown in either SEQ ID No. 1, or SEQ ID No. 11, or SEQ ID No. 13, or SEQ ID No. 23.

25 The amino acid sequence shown in SEQ ID No: 2, and the fragment represented in the ATCC clone 75787 represent biologically active portions of larger full-length forms of mammalian RAPT1 proteins. In preferred embodiments, the RAPT1 polypeptide includes a binding domain for binding to FKBP/rapamycin complexes, such as the rap-binding domains represented by residues 28-160 of SEQ ID No. 2, or residues 2012-2144 of SEQ ID No. 12.
30 In preferred embodiments, portions of the RAPT1 protein isolated from the full-length form will retain a specific binding affinity for an FKBP/rapamycin complex, e.g. an FKBP12/rapamycin complex, e.g. an affinity at least 50%, more preferably at least 75%, and even more preferably at least 90% that of the binding affinity of a naturally-occurring form of RAPT1 for such a rapamycin complex. A polypeptide is considered to possess a
35 biological activity of a RAPT1 protein if the polypeptide has one or more of the following properties: the ability to bind an FKBP/drug complex, e.g., an FKBP/macrolide complex, e.g., an FKBP/rapamycin complex; the ability to bind to an FKBP12/rapamycin complex; the ability to modulate assembly of FKBP/rapamycin-complexes; the ability to regulate cell

/8

proliferation, e.g., to regulate the cell-cycle, e.g., to regulate the progression of a cell through the G₁ phase. Moreover, based on sequence analysis, the biological function of the subject RAPT1 proteins can include a phosphatidyl inositol-kinase activity, such as a PI-3-kinase activity. A protein is also considered bioactive with respect to RAPT1 bioactivity if it is a specific agonist (mimetic) or antagonist of one of the above recited properties.

With respect to the rap-UBC enzyme, preferred embodiments of the subject the protein comprise at least a portion of the amino acid sequence of SEQ ID No. 24 (or of the rap-UBC gene insert of SMR4-15 described in Example 5) which possess either the ability to bind a FKBP/rapamycin complex or the ability to conjugating ubiquitin to a cellular protein, or both. Given that rapamycin causes a block in the cell-cycle during G₁ phase, it is probable that the spectrum of biological activity of the subject rap-UBC enzyme includes control of half-lives of certain cell cycle regulatory proteins, particularly relatively short lived proteins (e.g. proteins which have half-lives on the order of 30 minutes to 2 hours). For example, the subject UBC may have the ability to mediate ubiquitination of, for example, p53, myc and/or cyclins, and therefore affects the cellular half-life of a cell-cycle regulatory protein in proliferating cells. The binding of the rap-UBC to the FKBP/rapamycin complex may result in sequestering of the enzyme away from its substrate proteins. Thus, rapamycin may interfere with the ubiquitin-mediated degradation of p53 in a manner which causes cellular p53 levels to rise which in turn inhibits progression of the G₁ phase.

Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of the cloned RAP-binding proteins which function in a limited capacity as one of either a RAP-BP agonists or a RAP-BP antagonists, in order to either promote or inhibit only a subset of the biological activities of the naturally occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all RAP-BP related biological activities. For instance, RAPT1 analogs and rap-UBC analogs can be generated which do not bind in any substantial fashion to an FKBP/rapamycin complex, yet which retain most of the other biological functions ascribed to the naturally-occurring form of the protein. For example, the RAPT1 homolog might retain a kinase activity, such as a phosphatidyl inositol kinase activity, e.g. a PI-3-kinase activity. Conversely, the RAPT1 homolog may be engineered to lack a kinase activity, yet retain the ability to bind an FKBP/rapamycin complex. For instance, the FKBP/rapamycin binding portions of the RAPT1 homologs, such as the rapamycin-binding domains represented in SEQ ID Nos. 2 or 12, can be used to competitively inhibit binding to rapamycin complexes by the naturally-occurring form of RAPT1.

Homologs of the subject RAP-binding proteins can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For instance, mutation can give rise

to homologs which retain substantially the same, or merely a subset, of the biological activity of the RAP-BP from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein, such as by competitively binding to FKBP/rapamycin complexes.

5 The nucleotide sequence designated in SEQ ID No: 1 encodes a biologically active portion of the mouse RAPT1 protein, and in particular, includes a rapamycin-binding domain. Accordingly, one embodiment of the present invention provides a nucleic acid encoding a polypeptide comprising an amino acid sequence substantially homologous to that
10 portion of the RAPT1 protein represented by SEQ ID No: 2. Preferably, the nucleic acid is a cDNA molecule comprising at least a portion of the nucleotide sequence shown in SEQ ID No: 1. Yet another embodiment of the present invention provides a nucleic acid encoding a polypeptide comprising an amino acid sequence substantially homologous to a portion of the RAPT1 protein represented by SEQ ID No. 12 corresponding to a rapamycin-binding domain, e.g. Val2012 to Tyr 2144 of SEQ ID No: 12. In similar fashion, the present
15 invention provides a nucleic acid encoding at least a portion, e.g., a rapamycin-binding portion, of the *Candida* RAPT1 polypeptide of SEQ ID No. 14.

 Preferred nucleic acids encode a polypeptide including an amino acid sequence which is at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in one or more of SEQ ID Nos: 2, 12 or 14.
20 Nucleic acids encoding peptides, particularly peptides having an activity of a RAPT1 protein, and comprising an amino acid sequence which is at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homologous with a sequence shown in either SEQ ID No: 2, 12 or 14 are also within the scope of the invention, as of course are proteins which are identical to the aforementioned sequence listings. In one embodiment, the
25 nucleic acid is a cDNA encoding a peptide having at least one activity of a subject RAP-binding protein. Preferably, the nucleic acid is a cDNA molecule comprising at least a portion of the nucleotide sequence represented in one of SEQ ID Nos: 2, 12 or 14. A preferred portion of these cDNA molecules includes the coding region of the gene. For instance, a recombinant RAP-BP gene can include nucleotide sequences of a PCR fragment
30 generated by amplifying the coding sequences for one of the RAP-BP clones of ATCC deposit No: 75787.

 The nucleotide sequence shown in SEQ ID No: 23 encodes a biologically active human ubiquitin conjugating enzyme. Accordingly, in one embodiment of the present invention, the nucleic acid encodes a polypeptide including the rapamycin-binding domain of
35 the rap-UBC protein represented by SEQ ID No: 24. Preferably, the nucleic acid is a cDNA molecule comprising at least a portion of the nucleotide sequence shown in SEQ ID No: 23. Preferred nucleic acids encode a peptide comprising an amino acid sequence which is at least

2188061
20

60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in SEQ ID No: 24. Nucleic acids encoding polypeptides, particularly those having a ubiquitin conjugating activity, and comprising an amino acid sequence which is at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homologous with a sequence shown in SEQ ID No: 24 are also within the scope of the invention.

In a further embodiment of the invention, the recombinant RAP-BP genes can further include, in addition to the amino acid sequence shown in in the appended sequence listing, additional nucleotide sequences which encode amino acids at the C-terminus and N-terminus of the protein though not shown in those sequence listings. For instance, the recombinant RAPT1 gene can include nucleotide sequences of a PCR fragment generated by amplifying the RAPT1 coding sequence of pIC524 using sets of primers such described in Example 4. Additionally, in light of the present disclosure, it will be possible using no more than routine experimentation to isolate from, for example, a cDNA library, the remaining 5' sequences of RAPT1, such as by RACE PCR using primers designed from the sequences of the pIC524 clone, e.g., to generate the full-length sequence of SEQ ID No: 12. In particular, the invention contemplates a recombinant RAPT1 gene encoding the full-length RAPT1 protein. Yet another embodiment of the invention includes nucleic acids that encode isoforms of the mouse or human RAPT1, especially isoforms (e.g. splicing variants, allelic variants, etc.) that are capable of binding with the FKBP12/rapamycin complex. Such isoforms, as well as other members of the larger family of RAP-binding proteins, can be isolated using the drug-dependent interaction trap assays described in further detail below.

Another aspect of the invention provides a nucleic acid that hybridizes under high or low stringency conditions to a nucleic acid which encodes a peptide having at least a portion of an amino acid sequence represented by one of SEQ ID Nos.: 2, 12 or 14. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Nucleic acids having a sequence which differs from the nucleotide sequence shown in any of SEQ ID Nos: 1, 11 or 13 due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent peptides (i.e., a peptide having a biological activity of a RAP-binding protein) but that differ in sequence

from the appended sequence listings due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect the amino acid sequence of the RAP-binding protein.

5 However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject RAP-binding proteins will exist among vertebrates. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding polypeptides having an activity of a RAP-binding protein may exist among individuals of a given species due to natural allelic
10 variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of this invention.

The present invention also provides nucleic acid encoding only a portion of a RAPT1 protein, such as the rapamycin-binding domain. As used herein, a fragment of a nucleic acid encoding such a portion of a RAP-binding protein refers to a nucleotide sequence having
15 fewer nucleotides than the nucleotide sequence encoding the entire amino acid sequence of a full-length RAP-binding protein, yet which still includes enough of the coding sequence so as to encode a polypeptide which is capable of binding to an FKBP/rapamycin complex. Moreover, nucleic acid fragments within the scope of the invention include those fragments capable of hybridizing under high or low stringency conditions with nucleic acids from other
20 vertebrate species, particularly other mammals, and can be used in screening protocols to detect homologs, of the subject RAP-binding proteins. Nucleic acids within the scope of the invention may also contain linker sequences, modified restriction endonuclease sites and other sequences useful for molecular cloning, expression or purification of recombinant peptides derived from RAP-binding proteins.

25 As indicated by the examples set out below, a nucleic acid encoding a RAP-binding protein may be obtained from mRNA present in any of a number of cells from a vertebrate organism, particularly from mammals, e.g. mouse or human. It should also be possible to obtain nucleic acids encoding RAP-binding proteins from genomic DNA obtained from both adults and embryos. For example, a gene encoding a RAP-binding protein can be cloned
30 from either a cDNA or a genomic library in accordance with protocols herein described, as well as those generally known in the art. For instance, a cDNA encoding a RAPT1 protein, particularly other isoforms, e.g. paralogs or orthologs, of the RAPT1 proteins represented by either SEQ ID No. 2 or 12, can be obtained by isolating total mRNA from a mammalian cell, e.g. a human cell, generating double stranded cDNAs from the total mRNA, cloning the
35 cDNA into a suitable plasmid or bacteriophage vector, and isolating RAPT1 clones using any one of a number of known techniques, e.g. oligonucleotide probes or western blot analysis. Genes encoding proteins related to the subject RAP-binding proteins can also be cloned using

established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridizes (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a RAP-binding protein so as to inhibit expression of that protein, as for example by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a RAP-binding protein. Alternatively, the antisense construct can be an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of a RAP-BP gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by van der Krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of loads of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or

Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration, 5 penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral 10 administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which 15 they specifically bind. Such diagnostic tests are described in further detail below.

Likewise, the antisense constructs of the present invention, by antagonizing the normal biological activity of a RAP-binding protein, can be used in the manipulation of tissue, e.g. tissue proliferation and/or differentiation, both for *in vivo* and *ex vivo* tissue culture systems.

20 This invention also provides expression vectors containing a nucleic acid encoding a RAP-binding protein of the present invention, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of a 25 recombinant RAP-binding protein. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences-sequences that control the expression of a DNA sequence when 30 operatively linked to it may be used in these vectors to express DNA sequences encoding the RAP-binding proteins of this invention. Such useful expression control sequences, include, for example, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of 35 phage lambda, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the

24

promoters of the yeast α -mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In one embodiment, the expression vector includes a recombinant gene encoding a polypeptide which mimics or otherwise agonizes the action of a RAP-binding protein, or alternatively, which encodes a polypeptide that antagonizes the action of an authentic RAP-binding protein. Such expression vectors can be used to transfect cells and thereby produce polypeptides, including fusion proteins, encoded by nucleic acids as described herein.

Moreover, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids encoding either an agonistic or antagonistic form of one or more of the subject RAP-binding proteins. Thus, another aspect of the invention features expression vectors for *in vivo* transfection and expression of a RAP-binding protein in particular cell types so as to reconstitute the function of, or alternatively, abrogate the function of one or more of the subject RAP-binding proteins in a cell in which that protein or other transcriptional regulatory proteins to which it bind are misexpressed. For example, gene therapy can be used to deliver a gene encoding a rapamycin-insensitive RAP-binding protein in order to render a particular tissue or cell-type resistant to rapamycin induced cell-cycle arrest.

Expression constructs of the subject RAP-binding proteins, and mutants thereof, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the RAP-BP gene to cells *in vivo*. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramicidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or CaPO_4 precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of RAP-BP expression are also useful for *in vitro* transduction of cells, such as in diagnostic assays.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA, encoding the particular form of the RAP-binding protein desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded
5 within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the
10 transferred nucleic acids are stably integrated into the chromosomal DNA of the host. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy.
15 and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be constructed in which part of the retroviral coding sequence (*gag*, *pol*, *env*) has been replaced by nucleic acid encoding one of the subject receptors rendering the retrovirus replication defective.

20 The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.)
Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory
25 manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include ψ Crip, ψ Cre, ψ 2 and ψ Am. Retroviruses have been used to introduce a variety of genes into many different cell types, including lymphocytes, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science*
30 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464; Wilson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3014-3018; Armentano et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6141-6145; Huber et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8039-8043; Ferry et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8377-8381; Chowdhury et al. (1991) *Science* 254:1802-1805; van Beusechem et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:7640-
35 7644; Kay et al. (1992) *Human Gene Therapy* 3:641-647; Dai et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10892-10895; Hwu et al. (1993) *J. Immunol.* 150:4104-4115; U.S. Patent No.

4,868,116; U.S. Patent No. 4,980,286; PCT Application WO 89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234 and WO94/06920). For instance, strategies for the modification of the infection spectrum of retroviral vectors include: coupling antibodies specific for cell surface antigens to the viral *env* protein (Roux et al. (1989) *PNAS* 86:9079-9083; Julan et al. (1992) *J. Gen Virol* 73:3251-3255; and Goud et al. (1983) *Virology* 163:251-254); or coupling cell surface receptor ligands to the viral *env* proteins (Neda et al. (1991) *J Biol Chem* 266:14143-14146). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an ecotropic vector in to an amphotropic vector.

Moreover, use of retroviral gene delivery can be further enhanced by the use of tissue- or cell-specific transcriptional regulatory sequences which control expression of the RAP-BP gene of the retroviral vector.

Another viral gene delivery system useful in the present invention utilizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See for example Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain circumstances in that they are not capable of infecting nondividing cells and can be used to infect a wide variety of cell types. Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis *in situations* where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al. cited *supra*; Haj-Ahmand and Graham (1986) *J. Virol.* 57:267). Most replication-defective adenoviral vectors currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3

genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al. (1979) *Cell* 16:683; Berkner et al., *supra*; and Graham et al. in Methods in Molecular Biology, E.J. Murray, Ed. (Humana, Clifton, NJ, 1991) vol. 7. pp. 109-127). Expression of the inserted RAP-BP gene can be under control of, for example, the E1A promoter, the major
5 late promoter (MLP) and associated leader sequences, the E3 promoter, or exogenously added promoter sequences.

Yet another viral vector system useful for delivery of the subject RAP-BP gene is the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for
10 efficient replication and a productive life cycle. (For a review see Muzyczka et al. *Curr. Topics in Micro. and Immunol.* (1992) 158:97-129). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (see for example Flotte et al. (1992) *Am. J. Respir. Cell. Mol. Biol.* 7:349-356; Samulski et al. (1989) *J. Virol.* 63:3822-3828; and McLaughlin et al. (1989) *J. Virol.* 62:1963-1973).
15 Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector such as that described in Tratschin et al. (1985) *Mol. Cell. Biol.* 5:3251-3260 can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see for example Hermonat et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470;
20 Tratschin et al. (1985) *Mol. Cell. Biol.* 4:2072-2081; Wondisford et al. (1988) *Mol. Endocrinol.* 2:32-39; Tratschin et al. (1984) *J. Virol.* 51:611-619; and Flotte et al. (1993) *J. Biol. Chem.* 268:3781-3790).

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of an RAP-binding protein in the tissue of
25 an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject RAP-BP gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and
30 artificial viral envelopes.

In a representative embodiment, a gene encoding one of the subject RAP-binding proteins can be entrapped in liposomes bearing positive charges on their surface (e.g., lipofectins) and (optionally) which are tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al. (1992) *No Shinkei Geka* 20:547-551; PCT publication
35 WO91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075). For example, lipofection of cells can be carried out using liposomes tagged with monoclonal antibodies against any cell surface antigen present on, for example, T-cells.

In clinical settings, the gene delivery systems for the therapeutic RAP-BP gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) *PNAS* 91: 3054-3057).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

Another aspect of the present invention concerns recombinant RAP-binding proteins which are encoded by genes derived from eukaryotic cells, e.g. mammalian cells, e.g. cells from humans, mice, rats, rabbits, or pigs. The term "recombinant protein" refers to a protein of the present invention which is produced by recombinant DNA techniques, wherein generally DNA encoding, for example, the RAPT1 protein, is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant gene encoding the recombinant RAP-binding protein, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native RAP-binding protein, or an amino acid sequence similar thereto, which is generated by mutation so as to include substitutions and/or deletions relative to a naturally occurring form of the RAP-binding protein of a organism. Recombinant RAPT1 proteins preferred by the present invention, in addition to those having an amino acid sequence of a native RAPT1 protein, comprise amino acid sequences which are at least 70% homologous, more preferably 80% homologous and most preferably 90% homologous with an amino acid sequence shown in one of SEQ ID No: 2, 12 or 14. A polypeptide having a biological activity of a RAPT1 protein and which comprises an amino acid sequence that is at least about 95%, more preferably at least about 98%, and most preferably are identical to a sequence represented in one of SEQ ID No: 2, 12 or 14 are also within the scope of the invention.

Likewise, preferred embodiments of recombinant rap-UBC proteins include an amino acid sequence which is at least 70% homologous, more preferably 80% homologous, and most preferably 90% homologous with an amino acid sequence represented by SEQ ID No. 24. Recombinant rap-UBC proteins which are identical, or substantially identical (e.g. 95 to 5 98% homologous) with an amino acid sequence of SEQ ID No. 24 are also specifically contemplated by the present invention.

In addition, the invention expressly encompasses recombinant RAPT1 proteins produced from the ATCC deposited clones described in Example 4, e.g. from ATCC deposit number 75787, as well as recombinant ubiquitin-conjugating enzymes produced from ATCC 10 deposit number 75786, described in Example 5.

The present invention further pertains to recombinant forms of the subject RAP-binding proteins which are evolutionarily related to a RAP-binding protein represented in one of SEQ ID No: 2 or 12, that is, not identical, yet which are capable of functioning as an agonist or an antagonist of at least one biological activity of a RAP-binding protein. The 15 term "evolutionarily related to", with respect to amino acid sequences of recombinant RAP-binding proteins, refers to proteins which have amino acid sequences that have arisen naturally, as well as to mutational variants which are derived, for example, by recombinant mutagenesis.

Another aspect of the present invention pertains to methods of producing the subject 20 RAP-binding proteins. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject RAPT1 protein or rap-UBC can be cultured under appropriate conditions to allow expression of the peptide to occur. The peptide may be secreted and isolated from a mixture of cells and medium containing the recombinant protein. Alternatively, the peptide may be retained 25 cytoplasmically, as the naturally occurring forms of the subject RAP-binding proteins are believed to be, and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant RAP-binding proteins can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange 30 chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for a RAP-binding protein. In one embodiment, the RAP-binding protein is a fusion protein containing a domain which facilitates its purification, such as a RAPT1-GST fusion protein or a rapUBC-GST fusion protein.

35 The present invention also provides host cells transfected with a RAP-BP gene for expressing a recombinant form of a RAP-binding protein. The host cell may be any

prokaryotic or eukaryotic cell. Thus, a nucleotide sequence derived from the cloning of the RAP-binding proteins of the present invention, encoding all or a selected portion of a protein, can be used to produce a recombinant form of a RAP-BP via microbial or eukaryotic cellular processes. Ligating a polynucleotide sequence into a gene construct, such as an expression
5 vector, and transforming or transfecting host cells with the vector are standard procedures used in producing other well-known proteins, e.g. insulin, interferons, p53, myc, cyclins and the like. Similar procedures, or modifications thereof, can be employed to prepare recombinant RAP-binding proteins, or portions thereof, by microbial means or tissue-culture technology in accord with the subject invention. Host cells suitable for expression of a
10 recombinant RAP-binding protein can be selected, for example, from amongst eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial) cells.

The recombinant RAP-BP gene can be produced by ligating nucleic acid encoding a RAP-binding protein, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant
15 forms of RAP-binding proteins include plasmids and other vectors. For instance, suitable vectors for the expression of a RAP-BP include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli*.

A number of vectors exist for the expression of recombinant proteins in yeast. For
20 instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach *et al.* (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83). These vectors can replicate in *E. coli* due the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication
25 determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used.

Preferred mammalian expression vectors contain prokaryotic sequences to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription regulatory sequences that cause expression of a recombinant RAP-BP gene in eukaryotic cells. The
30 pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of
35 viruses such as the bovine papilloma virus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells.

31

Examples of other viral (including retroviral) expression systems can be found above in the description of gene therapy delivery systems.

In some instances, it may be desirable to express a recombinant RAP-binding protein by the use of a baculovirus expression system (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the β -gal containing pBlueBac III).

The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

When expression of a portion of one of the subject RAP-binding proteins is desired, i.e. a truncation mutant, such as the RAPT1 polypeptides of SEQ ID Nos: 2, 12 or 14, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli* (Ben-Bassat et al. (1987) *J. Bacteriol.* 169:751-757) and *Salmonella typhimurium* and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) *PNAS* 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing RAP-BP-derived polypeptides in a host which produces MAP (e.g., *E. coli* or CM89 or *S. cerevisiae*), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., *supra*).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene so as to be covalently linked in-frame with a second nucleotide sequence encoding a different polypeptide. This type of expression system can be useful, for instance, where it is desirable to produce an immunogenic fragment of a RAP-binding protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the RAPT1 polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequences corresponding to the portion of the RAPT1 protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising a portion of the protein RAPT1 as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions

can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of an RAPT1 protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No. 0259149; and Evans *et al.* (1989) *Nature* 339:385; Huang *et al.* (1988) *J. Virol.* 62:3855; and Schlienger *et al.* (1992) *J. Virol.* 66:2). The subject ubiquitin-conjugating enzyme can be manipulated as an immunogen in like fashion.

The Multiple Antigen Peptide system for peptide-based immunization can also be utilized, wherein a desired portion of a RAP-binding protein is obtained directly from organo-chemical synthesis of the peptide onto an oligomeric branching lysine core (see, for example, Posnett *et al.* (1988) *JBC* 263:1719 and Nardelli *et al.* (1992) *J. Immunol.* 148:914). Antigenic determinants of the RAP-binding proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression and purification of proteins, such as any one of the RAP-binding proteins of the present invention. For example, a RAP-binding protein can be generated as a glutathione-S-transferase (GST) fusion protein. Such GST fusion proteins can simplify purification of a RAP-binding protein, as for example by affinity purification using glutathione-derivatized matrices (see, for example, *Current Protocols in Molecular Biology*, eds. Ausabel *et al.* (N.Y.: John Wiley & Sons, 1991)). In another embodiment, a fusion gene coding for a purification leader sequence, such as a peptide leader sequence comprising a poly-(His)/enterokinase cleavage sequence, can be added to the N-terminus of the desired portion of a RAP-binding protein in order to permit purification of the poly(His)-fusion protein by affinity chromatography using a Ni²⁺ metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase (e.g., see Hochuli *et al.* (1987) *J. Chromatography* 411:177; and Janknecht *et al.* *PNAS* 88:8972).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which are subsequently annealed to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992).

The present invention also makes available purified, or otherwise isolated forms of the subject RAP-binding proteins which is isolated from, or otherwise substantially free of other cellular proteins, especially FKBP or other rapamycin binding proteins, as well as ubiquitin and ubiquitin-dependent enzymes, signal transduction, and cell-cycle regulatory proteins, which may be normally associated with the RAP-binding protein. The term "substantially free of other cellular or viral proteins" (also referred to herein as "contaminating proteins") or "substantially pure or purified preparations" are defined as encompassing preparations of RAP-binding proteins having less than 20% (by dry weight) contaminating protein, and preferably having less than 5% contaminating protein. Functional forms of the subject RAP-binding proteins can be prepared, for the first time, as purified preparations by using recombinant proteins as described herein. Alternatively, the subject RAP-binding proteins can be isolated by affinity purification using, for example, matrix bound FKBP/rapamycin protein. By "purified", it is meant, when referring to a peptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins (particularly FK506 binding proteins, as well as other contaminating proteins). The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions.

Furthermore, isolated peptidyl portions of the subject RAP-binding proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a RAP-binding protein of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a RAP-binding protein activity, such as by microinjection assays or *in vitro* protein binding assays. In an illustrative embodiment, peptidyl portions of a RAP-binding protein, such as RAPT1 or rapUBC, can be tested for FKBP/rapamycin-binding activity.

It will also be possible to modify the structure of a RAP-binding protein for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the RAP-binding protein described in more detail herein. Such modified peptide can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the folding of the protein, and may or may not have much of an effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur-containing = cysteine and methionine (see, for example, *Biochemistry*, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Alternatively, amino acid replacement can be based on steric criteria, e.g. isosteric replacements, without regard for polarity or charge of amino acid sidechains. Whether a change in the amino acid sequence of a peptide results in a functional RAP-BP homolog (e.g. functional in the sense that it acts to mimic or antagonize the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type RAP-BP or competitively inhibit such a response. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of RAP-binding proteins, e.g. of RAPT1 proteins and/or rap-UBC enzymes, as well as truncation mutants, thereof and is especially useful for identifying variant sequences (e.g. RAP-BP homologs) that are functional in regulating rapamycin-mediated effects, as well as other aspects of cell growth or differentiation. In similar fashion, RAP-BP homologs can be

generated by the present combinatorial approach which are antagonists in that they are able to interfere with the normal cellular functions of authentic forms of the protein.

One purpose for screening such combinatorial libraries is, for example, to isolate novel RAP-BP homologs from the library which function in the capacity as one of either an agonists or an antagonist of the biological activities of the wild-type ("authentic") protein, or alternatively, which possess novel biological activities all together. To illustrate, RAPT1 homologs can be engineered by the present method to provide homologs which are unable to bind to the FKBP/rapamycin complex, yet still retain at least a portion of the normal cellular activity associated with authentic RAPT1. Thus, combinatorially-derived homologs can be generated to provide rapamycin-resistance. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols.

Likewise, mutagenesis can give rise to RAP-BP homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of, the authentic RAP-binding protein. Such homologs, and the genes which encode them, can be utilized to alter the envelope of expression of a particular RAP-BP by modulating the half-life of the protein. For instance, a short half-life can give rise to more transient RAPT1 biological effects and, when part of an inducible expression system, can allow tighter control of recombinant RAPT1 levels within the cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

In an illustrative embodiment of this method, the amino acid sequences for a population of RAP-BP homologs, or other related proteins, are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, RAPT1 homologs from one or more species, e.g. a sequence alignment of the mouse and human RAPT1 proteins represented by SEQ ID Nos. 2 and 12, or different RAP-BP isoforms from the same species, e.g. different human RAPT1 isoforms. Amino acids which appear at each position of the sequence alignment can be selected to create a degenerate set of combinatorial sequences.

In a preferred embodiment, the combinatorial RAP-BP library is produced by way of a degenerate library of genes encoding a library of polypeptides which each include at least a portion of potential RAP-BP sequences, e.g. the portion of RAPT1 represented by SEQ ID No: 2 or 12, or the portion of rap-UBC represented by SEQ ID No. 24. A mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential RAP-BP sequences are expressible as individual polypeptides, or

36

alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the RAP-BP sequence library therein.

There are many ways by which the library of RAP-BP homologs can be generated from a degenerate oligonucleotide sequence. For instance, chemical synthesis of a degenerate gene sequence can be carried out in an automated DNA synthesizer, and the synthetic genes then ligated into an appropriate gene for expression. The purpose of a degenerate set of RAP-BP genes is to provide, in one mixture, all of the sequences encoding the desired set of potential RAP-BP sequences. The synthesis of degenerate oligonucleotides is well known in the art (see, for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura *et al.* (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott *et al.* (1990) *Science* 249:386-390; Roberts *et al.* (1992) *PNAS* 89:2429-2433; Devlin *et al.* (1990) *Science* 249: 404-406; Cwirla *et al.* (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial library. For example, RAP-BP homologs (both agonist and antagonist forms) can be generated and isolated from a library generated by using, for example, alanine scanning mutagenesis and the like (Ruf *et al.* (1994) *Biochemistry* 33:1565-1572; Wang *et al.* (1994) *J. Biol. Chem.* 269:3095-3099; Balint *et al.* (1993) *Gene* 137:109-118; Grodberg *et al.* (1993) *Eur. J. Biochem.* 218:597-601; Nagashima *et al.* (1993) *J. Biol. Chem.* 268:2888-2892; Lowman *et al.* (1991) *Biochemistry* 30:10832-10838; and Cunningham *et al.* (1989) *Science* 244:1081-1085), by linker scanning mutagenesis (Gustin *et al.* (1993) *Virology* 193:653-660; Brown *et al.* (1992) *Mol. Cell Biol.* 12:2644-2652; McKnight *et al.* (1982) *Science* 232:316); by saturation mutagenesis (Meyers *et al.* (1986) *Science* 232:613); by PCR mutagenesis (Leung *et al.* (1989) *Method Cell Mol Biol* 1:11-19); or by random mutagenesis (Miller *et al.* (1992) *A Short Course in Bacterial Genetics*, CSHL Press, Cold Spring Harbor, NY; and Greener *et al.* (1994) *Strategies in Mol Biol* 7:32-34).

A wide range of techniques are known in the art for screening gene products of variegated gene libraries made by combinatorial mutagenesis, especially for identifying individual gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of, for example, RAPT1 homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity

facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate RAP-BP sequences created by combinatorial mutagenesis techniques.

5 In one screening assay, the candidate RAP-BP gene products are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind the FKBP12/rapamycin complex via this gene product is detected in a "panning assay". For instance, the degenerate RAP-BP gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning
10 protocols (see, for example, Ladner *et al.*, WO 88/06630; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1371; and Goward *et al.* (1992) *TIBS* 18:136-140). In a similar fashion, fluorescently labeled molecules which bind the RAP-binding protein, such as fluorescently labeled rapamycin or FKBP12/rapamycin complexes, can be used to score for potentially functional RAP-BP homologs. Cells can be visually inspected and separated under a fluorescence
15 microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In an alternate embodiment, the gene library is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two
20 significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E.coli* filamentous phages M13, fd, and f1 are
25 most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner *et al.* PCT publication WO 90/02909; Garrard *et al.*, PCT publication WO 92/09690; Marks *et al.* (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Clackson *et al.* (1991) *Nature* 352:624-628; and Barbas *et al.* (1992) *PNAS*
30 89:4457-4461). In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening RAP-BP combinatorial libraries, and the RAP-BP phage library can be panned on glutathione-immobilized FKBP-GST/rapamycin complexes. Successive rounds of reinfection, phage amplification, and panning will greatly enrich for homologs which retain
35 FKBP/rapamycin binding and which can be subsequently screened for further biological activities in order to discern between agonists and antagonists.

Homologs of the human and mouse RAP-binding proteins can also be generated through the use of interaction trap assays to screen combinatorial libraries of RAP-BP mutants. As described in Example 10 below, the same two hybrid assay used to screen cDNA libraries for proteins which interact with FK506-binding proteins in a drug-dependent manner can also be used to sort through combinatorial libraries of, for example, RAPT1 mutants, to find both agonistic and antagonistic forms. By controlling the sensitivity of the assay for interactions, e.g. through the manipulation of the strength of the promoter sequence used to drive expression of the reporter construct, the assay can be generated to favor agonistic forms of RAPT1 with tighter binding affinities for rapamycin than the authentic form of the protein. Alternatively, as described in Example 10, the assay can be used to select for RAPT1 homologs which are now unable to bind rapamycin complexes and hence are versions of the RAPT1 protein which can render a cell insensitive to treatment with that macrolide.

The invention also provides for reduction of the rapamycin-binding domains of the subject RAP-binding proteins to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt binding of a polypeptide of the present invention with an FKBP/rapamycin complex. Thus, such mutagenic techniques as described above are also useful to map the determinants of RAP-binding proteins which participate in interactions involved in, for example, binding to an FKBP/rapamycin complex. To illustrate, the critical residues of a RAP-binding protein which are involved in molecular recognition of FKBP/rapamycin can be determined and used to generate RAP-BP-derived peptidomimetics that competitively inhibit binding of the RAP-BP to rapamycin complexes. By employing, for example, scanning mutagenesis to map the amino acid residues of a particular RAP-binding protein involved in binding FKBP/rapamycin complexes, peptidomimetic compounds can be generated which mimic those residues in binding to the rapamycin complex, and which, by inhibiting binding of the RAP-BP to FKBP/rapamycin, can interfere with the function of rapamycin in cell-cycle arrest. For instance, non-hydrolyzable peptide analogs of such residues can be generated using retro-inverse peptides (e.g., see U.S. Patents 5,116,947 and 5,218,089; and Pallai et al. (1983) *Int J Pept Protein Res* 21:84-92) benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β -turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β -aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res*

Commun 134:71). Utilizing side-by-side assays, peptidomimetics can be designed to specifically inhibit the interaction of human RAPT1 (or other mammalian homologs) with the FKBP12/rapamycin complex in mammalian cells, but which do not substantially affect the interaction of the yeast protein TOR1 or TOR2 with the FKBP1/rapamycin complex. Such a peptide analog could be used in conjunction with rapamycin treatment of mycotic infections to protect the host mammal from rapamycin side-effects, such as immunosuppression, without substantially reducing the efficacy of rapamycin as an anti-fungal agent.

Another aspect of the invention pertains to an antibody specifically reactive with one or more of the subject RAP-binding proteins. For example, by using immunogens derived from a RAP-binding protein, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a full length RAP-binding protein or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject RAP-binding proteins can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the RAP-binding proteins of the present invention, e.g. antigenic determinants of a protein represented in one of SEQ ID Nos: 2, 12 or a closely related human or non-human mammalian homolog thereof. For instance, a favored anti-RAP-BP antibody of the present invention does not substantially cross react (i.e. react specifically) with a protein which is less than 90 percent homologous to one of SEQ ID Nos: 2 or 12; though antibodies which do not substantially cross react with a protein which is less than 95 percent homologous with one of SEQ ID Nos: 2, 12 or 24, or even less than 98-99 percent homologous with one of SEQ ID Nos: 2 or 12, are specifically contemplated. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein (e.g. a yeast TOR1 or TOR2 protein) which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a mammalian RAPT1 protein, e.g., such as represented one of SEQ ID Nos: 2 or 12.

Following immunization, anti-RAP-BP antisera can be obtained and, if desired, polyclonal anti-RAP-BP antibodies isolated from the serum. To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized

40

animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) *Nature*, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) *Immunology Today*, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a RAP-binding protein of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

10 An antibody preparation of this invention prepared from a polypeptide as described above can be in dry form as obtained by lyophilization. However, the antibodies are normally used and supplied in an aqueous liquid composition in serum or a suitable buffer such as PBS.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with one of the subject RAP-binding protein. Antibodies can be fragmented using conventional techniques, including recombinant engineering, and the fragments screened for utility in the same manner as described above for whole antibodies. For example, F(ab')₂ fragments can be generated by treating antibody with pepsin. The resulting F(ab')₂ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having an anti-RAP-BP portion.

Both monoclonal and polyclonal antibodies (Ab) directed against a RAP-binding protein can be used to block the action of that protein and allow the study of the role of a particular RAP-binding protein in, for example, cell-cycle regulation generally, or in the etiology of proliferative and/or differentiative disorders specifically, or in the mechanism of action of rapamycin, e.g. by microinjection of anti-RAP-BP antibodies into cells.

Antibodies which specifically bind RAP-BP epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject RAP-binding proteins. Anti-RAP-BP antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate RAP-BP levels in tissue or bodily fluid as part of a clinical testing procedure. For instance, such measurements as the level of free RAP-BP to RAP-BP/FKBP/drug complexes can be useful in predictive valuations of the efficacy of a particular rapamycin analog, and can permit determination of the efficacy of a given treatment regimen for an individual. The level of a RAP-binding protein can be measured in cells found in bodily fluid, such as in cells from samples of blood, or can be measured in tissue, such as produced by biopsy.

4/

Another application of the subject antibodies is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λ gt18-23, λ ZAP, and λ ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λ gt11 will produce
5 fusion proteins whose amino termini consist of β -galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a RAP-binding protein can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-RAP-BP antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of RAP-BP homologs can be
10 detected and cloned from other animals, and alternate isoforms (including splicing variants) can be detected and cloned from human sources.

Moreover, the nucleotide sequence determined from the cloning of the subject RAP-binding proteins from a human cell line will further allow for the generation of probes designed for use in identifying homologs in other human cell types, as well as RAP-BP
15 homologs (e.g. orthologs) from other mammals. For example, by identifying highly conserved nucleotides sequence through comparison of the mammalian RAPT1 genes with the yeast TOR genes, it will be possible to design degenerate primers for isolating RAPT1 homologs from virtually any eukaryotic cell. For instance, alignment of the mouse RAPT1 gene sequence and the yeast DRR-1 and TOR2 sequences, we have determined that optimal
20 primers for isolating RAPT1 homologs from other mammalian homologs, as well as from pathogenic fungi, include the primers GRGAYTTRAWBGABGCHYAMGAWTGG, CAAGCBTGGGAYMTYMTYTATMAYGTBTTTCAG, and GAYYBGARTTGGCTGTBCCHGG.

Accordingly, the present invention also provides a probe/primer comprising a
25 substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or anti-sense sequence of one of SEQ ID Nos: 1 or 11, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from the
30 group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Such probes can also be used as a part of a diagnostic test kit for identifying transformed cells, such as for measuring a level of a RAP-BP nucleic acid in a sample of cells from a patient; e.g. detecting mRNA encoding a RAP-BP mRNA level; e.g. determining whether a genomic RAP-BP gene has been mutated or deleted.

35 In addition, nucleotide probes can be generated which allow for histological screening of intact tissue and tissue samples for the presence of a RAP-BP mRNA. Similar to the diagnostic uses of anti-RAP-BP antibodies, the use of probes directed to RAP-BP

mRNAs, or to genomic RAP-BP sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth) or abnormal differentiation of tissue. Used in conjunction with an antibody immunoassays, the nucleotide probes can help
5 facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of a RAP-binding protein. For instance, variation in synthesis of a RAP-binding protein can be distinguished from a mutation in the genes coding sequence.

Thus, the present invention provides a method for determining if a subject is at risk
10 for a disorder characterized by unwanted cell proliferation or abherent control of differentiation. In preferred embodiments, the subject method can be generally characterized as comprising detecting, in a tissue sample of the subject (e.g. a human patient), the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding one of the subject RAP-binding proteins or (ii) the mis-
15 expression of a RAP-BP gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a RAP-BP gene, (ii) an addition of one or more nucleotides to such a RAP-BP gene, (iii) a substitution of one or more nucleotides of a RAP-BP gene, (iv) a gross chromosomal rearrangement of one of the RAP-BP genes, (v) a gross alteration in the level of a messenger
20 RNA transcript of a RAP-BP gene, (vi) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a RAP-BP gene, and (vii) a non-wild type level of a RAP-binding protein. In one aspect of the invention there is provided a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of one of SEQ ID Nos: 1 or 11, or naturally
25 occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject RAP-BP genes. The probe is exposed to nucleic acid of a tissue sample; and the hybridization of the probe to the sample nucleic acid is detected. In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos: 4,683,195 and 4,683,202) or, alternatively,
30 in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science*, 241:1077-1080; and NaKazawa et al. (1944) *PNAS* 91:360-364) the later of which can be particularly useful for detecting point mutations in the RAP-BP gene. Alternatively, immunoassays can be employed to determine the level of RAP-binding protein and/or its participation in protein complexes, particularly transcriptional regulatory complexes such as those involving
35 FKBP/rapamycin.

Also, by inhibiting endogenous production of a particular RAP-binding protein, anti-sense techniques (e.g. microinjection of antisense molecules, or transfection with plasmids

43

whose transcripts are anti-sense with regard to a RAP-BP mRNA or gene sequence) can be used to investigate role of each of the subject RAP-BP in growth and differentiative events, such as those giving rise to Wilm's tumor, as well as normal cellular functions of each of the subject RAP-binding proteins, e.g. in regulation of transcription. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

Furthermore, by making available purified and recombinant RAP-binding proteins, the present invention provides for the generation of assays which can be used to screen for drugs which are either agonists or antagonists of the cellular function of each of the subject RAP-binding proteins, or of their role in the pathogenesis of proliferative and differentiative disorders. For instance, an assay can be generated according to the present invention which evaluates the ability of a compound to modulate binding between a RAP-binding protein and an FK506-binding protein. In particular, such assays can be used to design and screen novel rapamycin analogs, as well as test completely unrelated compounds for their ability to mediate formation of FKBP/RAP-BP complexes. Such assays can be used to generate more potent anti-proliferative agents having a similar mechanism of action as rapamycin, e.g. rapamycin analogs. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by skilled artisan.

One aspect of the present invention which facilitates the generation of drug screening assays, particularly the high-throughout assays described below, is the identification of the rapamycin binding domain of RAPT1-like proteins. For instance, the present invention provides portions of the RAPT1-like proteins which are easier to manipulate than the full length protein. The full length protein is, because of its size, more difficult to express as a recombinant protein or a fusion protein which would retain rapamycin-binding activity, and may very well be insoluble. Accordingly, the present invention provides soluble polypeptides which include a soluble portion of a RAPT1-like polypeptide that binds to said FKBP/rapamycin complex, such as the rapamycin-binding domain represented by an amino acid sequence selected from the group consisting Val26-Tyr160 of SEQ ID No. 2, Val2012-Tyr2144 of SEQ ID No. 12, Val41-Tyr173 of SEQ ID No. 14, Val1-Tyr133 of SEQ ID No. 16, and Val1-Arg133 of SEQ ID No. 18.

For instance, RAPT1 polypeptides useful in the subject screening assays may be represented by the general formula X-Y-Z, Y represents an amino acid sequence of a rapamycin-binding domain within residues 2012 to 2144 of SEQ ID No. 12, X is absent, or represents all or a C-terminal portion of the amino acid sequence between about residues 1700 and 2144 of SEQ ID No. 12 not represented by Y, and Z is absent, or represents all or an N-terminal portion of the amino acid sequence between residues 2012 and 2549 of SEQ ID No. 12 not represented by Y. Preferably, the polypeptide includes only about 50 to 200

residues of RAPT1 protein sequence, which portion includes a rapamycin-binding domain. Similar polypeptides can be generated for other RAPT1-like proteins.

In an alternative embodiment, the same formula can also be used to designate a bioactive fragment of the subject RAPT1 protein, wherein Y represents a rapamycin-binding domain within residues 2012 to 2144 of SEQ ID No. 12, X is absent or represents a polypeptide from 1 to about 500 amino acid residues of SEQ ID No. 12 immediately N-terminal to the rapamycin-binding domain, and Z is absent or represents from 1 to about 365 amino acid residues of SEQ ID No. 2 immediately C-terminal to the selected rapamycin-binding domain.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target when contacted with a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest (the "drug") is contacted with a mixture generated from an isolated and purified RAP-binding protein, such as RAPT1 or rapUBC, and an FK506-binding protein. Detection and quantification of drug-depedent FKBP/RAP-BP complexes provides a means for determining the compound's efficacy for mediating complex formation between the two proteins. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified RAP-BP is added to a composition containing the FK506-binding protein, and the formation of FKBPRAP-BP complexes is quantitated in the absence of the test compound.

Complex formation between the RAP-binding protein and an FKBP/drug complex may be detected by a variety of techniques. For instance, modulation in the formation of complexes can be quantitated using, for example, detectably labelled proteins (e.g. radiolabelled, fluorescently labelled, or enzymatically labelled), by immunoassay, or by chromatographic detection.

Typically, it will be desirable to immobilize either the FK506-binding protein or the RAP-binding protein to facilitate separation of drug-dependent protein complexes from

45

uncomplexed forms of one of the proteins, as well as to accommodate automation of the assay. In an illustrative embodiment, a fusion protein can be provided which adds a domain that permits the protein to be bound to an insoluble matrix. For example, glutathione-S-transferase/FKBP (FKBP-GST) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the RAP-binding protein, e.g. an ^{35}S -labeled RAP-binding protein, and the test compound and incubated under conditions conducive to complex formation (see, for instance, Example 9). Following incubation, the beads are washed to remove any unbound RAP-BP, and the matrix bead-bound radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the FKBP/RAP-BP complexes are dissociated, e.g. when microtitre plates are used. Alternatively, after washing away unbound protein, the complexes can be dissociated from the matrix, separated by SDS-PAGE gel, and the level of RAP-BP found in the matrix-bound fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, the FK506-binding protein can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated FKBP can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the FKBP can be derivatized to the wells of the plate, and FKBP trapped in the wells by antibody conjugation. As above, preparations of a RAP-binding protein and a test compound are incubated in the FKBP-presenting wells of the plate, and the amount of FKBP/RAP-BP complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the RAP-binding protein, or which are reactive with the FK506-binding protein and compete for binding with the RAP-BP; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the RAP-binding protein. In the instance of the latter, the enzymatic activity can be endogenous, such as a kinase (RAPT1) or ubiquitin ligase (rapUBC) activity, or can be an exogenous activity chemically conjugated or provided as a fusion protein with the RAP-binding protein. To illustrate, the RAP-binding protein can be chemically cross-linked with alkaline phosphatase, and the amount of RAP-BP trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. paranitrophenyl phosphate. Likewise, a fusion protein comprising the RAP-BP and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) *J Biol Chem* 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as the anti-RAP-BP antibodies described herein, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the RAP-BP or FKBP sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include *myc*-epitopes (e.g., see Ellison et al. (1991) *J Biol Chem* 266:21150-21157) which includes a 10-residue sequence from *c-myc*, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, NJ).

Additionally, the subject RAP-binding proteins can be used to generate a drug-dependent interaction trap assay, as described in the examples below, for detecting agents which induce complex formation between a RAP-binding protein and an FK506-binding protein. As described below, the interaction trap assay relies on reconstituting *in vivo* a functional transcriptional activator protein from two separate fusion proteins, one of which comprises the DNA-binding domain of a transcriptional activator fused to an FK506-binding protein (see also U.S. Patent No: 5,283,317; PCT publication WO94/10300; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J Biol Chem* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; and Iwabuchi et al. (1993) *Oncogene* 8:1693-1696). The second fusion protein comprises a transcriptional activation domain (e.g. able to initiate RNA polymerase transcription) fused to one of the subject RAP-binding proteins. When the FKBP and RAP-binding protein interact in the presence of an agent such as rapamycin, the two domains of the transcriptional activator protein are brought into sufficient proximity as to cause transcription of a reporter gene. In addition to the LexA interaction trap described in the examples below, yet another illustrative embodiment comprises *Saccharomyces cerevisiae* YPB2 cells transformed simultaneously with a plasmid encoding a GAL4db-FKBP fusion (db: DNA binding domain) and with a plasmid encoding the GAL4 activation domain (GAL4ad) fused to a subject RAP-BP. Moreover, the strain is transformed such that the GAL4-responsive promoter drives expression of a phenotypic marker. For example, the ability to grow in the absence of histidine can depend on the expression of the *HIS3* gene. When the *HIS3* gene is placed under the control of a GAL4-responsive promoter, relief of this auxotrophic phenotype indicates that a functional GAL4 activator has been reconstituted through the drug-dependent interaction of FKBP and the RAP-BP. Thus, agent able to promote RAP-BP interaction with an FKBP will result in yeast cells able to grow in the absence of histidine. Commercial kits which can be modified to develop two-hybrid assays with the subject RAP-binding proteins are presently available (e.g., MATCHMAKER^{*} kit, ClonTech catalog number K1605-1, Palo Alto, CA).

*Trade-mark

47

In a preferred embodiment, assays which employ the subject mammalian RAP-binding proteins can be used to identify rapamycin mimetics that have therapeutic indexes more favorable than rapamycin. For instance, rapamycin-like drugs can be identified by the present invention which have enhanced tissue-type or cell-type specificity relative to rapamycin. To illustrate, the subject assays can be used to generate compounds which preferentially inhibit IL-2 mediated proliferation/activation of lymphocytes without substantially interfering with other tissues, e.g. hepatocytes. Likewise, similar assays can be used to identify rapamycin-like drugs which inhibit proliferation of yeast cells or other lower eukaryotes, but which have a substantially reduced effect on mammalian cells, thereby improving therapeutic index of the drug as an anti-mycotic agent relative to rapamycin.

In one embodiment, the identification of such compounds is made possible by the use of differential screening assays which detect and compare drug-mediated formation of two or more different types of FKBP/RAP-BP complexes. To illustrate, the assay can be designed for side-by-side comparison of the effect of a test compound on the formation of tissue-type specific FKBP/RAPT1 complexes. Given the diversity of FKBP, and the substantial likelihood that RAPT1 represents a single member of a larger family of related proteins, it is probable that different functional FKBP/RAPT1 complexes exist and, in certain instances, are localized to particular tissue or cell types. As described in PCT publication WO93/23548, entitled "*Method of Detecting Tissue-Specific FK506 Binding Protein Messenger RNAs and Uses Thereof*", the tissue distribution of FKBP can vary from one species of the protein to the next. Thus, test compounds can be screened for agents able to mediate the tissue-specific formation of only a subset of the possible repertoire of FKBP/RAPT1 complexes. In an exemplary embodiment, an interaction trap assay can be derived using two or more different bait proteins, e.g. FKBP12 (SEQ ID Nos. 5 and 6), FKBP25 (GenBank Accession M90309), or FKBP52 (Genbank Accession M88279), while the fish protein is constant in each, e.g. a human RAPT1 construct. Running the ITS side-by-side permits the detection of agents which have a greater effect (e.g. statistically significant) on the formation of one of the FKBP/RAPT1 complexes than on the formation of the other FKBP complexes.

In similar fashion, differential screening assays can be used to exploit the difference in drug-mediated formation of mammalian FKBP/RAP-BP complexes and yeast FKBP/TOR complexes in order to identify agents which display a statistically significant increase in specificity for the yeast complexes relative to the mammalian complexes. Thus, lead compounds which act specifically on pathogens, such as fungus involved in mycotic infections, can be developed. By way of illustration, the present assays can be used to screen for agents which may ultimately be useful for inhibiting at least one fungus implicated in such mycosis as candidiasis, aspergillosis, mucormycosis, blastomycosis, geotrichosis, cryptococcosis, chromoblastomycosis, coccidioidomycosis, conidiosporosis, histoplasmosis,

maduromycosis, rhinosporidiosis, nocardiosis, para-actinomycosis, penicilliosis, moniliasis, or sporotrichosis. For example, if the mycotic infection to which treatment is desired is candidiasis, the present assay can comprise comparing the relative effectiveness of a test compound on mediating formation of a mammalian FKBP/RAP1 complex with its effectiveness towards mediating such complexes formed from genes cloned from yeast selected from the group consisting of *Candida albicans*, *Candida stellatoidea*, *Candida tropicalis*, *Candida parapsilosis*, *Candida krusei*, *Candida pseudotropicalis*, *Candida quilliermondii*, or *Candida rugosa*. Likewise, the present assay can be used to identify anti-fungal agents which may have therapeutic value in the treatment of aspergillosis by making use of the subject drug-dependent interaction trap assays derived from FKBP and TOR genes cloned from yeast such as *Aspergillus fumigatus*, *Aspergillus flavus*, *Aspergillus niger*, *Aspergillus nidulans*, or *Aspergillus terreus*. Where the mycotic infection is mucormycosis, the complexes can be derived from yeast such as *Rhizopus arrhizus*, *Rhizopus oryzae*, *Absidia corymbifera*, *Absidia ramosa*, or *Mucor pusillus*. Sources of other rapamycin-dependent complexes for comparison with a mammalian FKBP/RAP1 complex includes the pathogen *Pneumocystis carinii*. Exemplary FK506-binding proteins from human pathogens and other lower eukaryotes are provided by, for example, GenBank Accession numbers: M84759 (*Candida albican*); U01195, U01198, U01197, U01193, U01188, U01194, U01199 (*Neisseria* spp.); and M98428 (*Streptomyces chrysomallus*).

In an exemplary embodiment, the differential screening assay can be generated using at least the rapamycin-binding domain of the *Candida albican* RAP1 protein (see Example 11) and a *Candida* FK506-binding protein (such as RBP1, GenBank No. M84759, see also Ferrara et al. (1992) *Gene* 113:125-127), or a yeast FK506-binding protein (see Example 8 and Figure 3). Comparison of formation of human RAP1 complexes and *Candida* RAP1 complexes provides a means for identifying agents which are more selective for the formation of caRAP1 complexes and, accordingly, likely to be more specific as anti-mycotic agents relative to rapamycin.

Another aspect of the present invention concerns transgenic animals which are comprised of cells (of that animal) which contain a transgene of the present invention and which preferably (though optionally) express an exogenous RAP-binding protein in one or more cells in the animal. The RAP-BP transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs designed to inhibit expression of the endogenous gene. In preferred embodiments, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, through the use of cis-acting sequences that control expression in the desired pattern. In the present invention, such mosaic expression of the subject RAP-binding proteins can be essential for many forms of lineage analysis and can

5 additionally provide a means to assess the effects of loss-of-function mutations, which deficiency might grossly alter development in small patches of tissue within an otherwise normal embryo. Toward this end, tissue-specific regulatory sequences and conditional regulatory sequences can be used to control expression of the transgene in certain spatial patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

10 Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation *in vivo* are known to those skilled in the art. For instance, genetic systems are available which allow for the regulated expression of a recombinase that catalyzes the genetic recombination a target sequence. As used herein, the phrase "target sequence" refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of a subject RAP-binding protein. For example, excision of a target sequence which interferes with the expression of a recombinant RAP-BP gene can be designed to activate expression of that gene. This interference with expression of the protein can result from a variety of mechanisms, such as spatial separation of the gene from a promoter element or an internal stop codon. Moreover, the transgene can be made wherein the coding sequence of the gene is flanked by recombinase recognition sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In such an instance, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allow for promoter driven transcriptional activation.

25 In an illustrative embodiment, either the *cre/loxP* recombinase system of bacteriophage P1 (Lakso et al. (1992) *PNAS* 89:6232-6236; Orban et al. (1992) *PNAS* 89:6861-6865) or the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355; PCT publication WO 92/15694) can be used to generate *in vivo* site-specific genetic recombination systems. Cre recombinase catalyzes the site-specific recombination of an intervening target sequence located between *loxP* sequences. *loxP* sequences are 34 base pair nucleotide repeat sequences to which the Cre recombinase binds and are required for Cre recombinase mediated genetic recombination. The orientation of *loxP* sequences determines whether the intervening target sequence is excised or inverted when Cre recombinase is present (Abremski et al. (1984) *J. Biol. Chem.* 259:1509-1514); catalyzing the excision of the target sequence when the *loxP* sequences are oriented as direct repeats and catalyzes inversion of the target sequence when *loxP* sequences are oriented as inverted repeats.

Accordingly, genetic recombination of the target sequence is dependent on expression of the Cre recombinase. Expression of the recombinase can be regulated by promoter elements which are subject to regulatory control, e.g., tissue-specific, developmental stage-specific, inducible or repressible by externally added agents. This regulated control will result in genetic recombination of the target sequence only in cells where recombinase expression is mediated by the promoter element. Thus, the activation expression of a RAP-binding protein can be regulated via regulation of recombinase expression.

Use of the *cre/loxP* recombinase system to regulate expression of a recombinant RAP-binding protein, such as RAPT1 or rapUBC, requires the construction of a transgenic animal containing transgenes encoding both the Cre recombinase and the subject protein. Animals containing both the Cre recombinase and the recombinant RAP-BP genes can be provided through the construction of "double" transgenic animals. A convenient method for providing such animals is to mate two transgenic animals each containing a transgene, e.g., the RAP-BP gene in one animal and recombinase gene in the other.

One advantage derived from initially constructing transgenic animals containing a transgene in a recombinase-mediated expressible format derives from the likelihood that the subject protein will be deleterious upon expression in the transgenic animal. In such an instance, a founder population, in which the subject transgene is silent in all tissues, can be propagated and maintained. Individuals of this founder population can be crossed with animals expressing the recombinase in, for example, one or more tissues. Thus, the creation of a founder population in which, for example, an antagonistic RAP-BP transgene is silent will allow the study of progeny from that founder in which disruption of cell-cycle regulation in a particular tissue or at developmental stages would result in, for example, a lethal phenotype.

Similar conditional transgenes can be provided using prokaryotic promoter sequences which require prokaryotic proteins to be simultaneously expressed in order to facilitate expression of the transgene. Exemplary promoters and the corresponding trans-activating prokaryotic proteins are given in U.S. Patent No. 4,833,080. Moreover, expression of the conditional transgenes can be induced by gene therapy-like methods wherein a gene encoding the trans-activating protein, e.g. a recombinase or a prokaryotic protein, is delivered to the tissue and caused to be expressed using, for example, one of the gene therapy constructs described above. By this method, the RAP-BP transgene could remain silent into adulthood and its expression "turned on" by the introduction of the trans-activator.

In an exemplary embodiment, the "transgenic non-human animals" of the invention are produced by introducing transgenes into the germline of the non-human animal. Embryonal target cells at various developmental stages can be used to introduce transgenes.

Different methods are used depending on the stage of development of the embryonal target cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster et al. (1985) *PNAS* 82:4438-4442). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating transgenes in practicing the invention.

Retroviral infection can also be used to introduce a RAP-BP transgene into a non-human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, R. (1976) *PNAS* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al. (1985) *PNAS* 82:6927-6931; Van der Putten et al. (1985) *PNAS* 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart et al. (1987) *EMBO J.* 6:383-388). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner et al. (1982) *Nature* 298:623-628). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the cells which formed the transgenic non-human animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner et al. (1982) *supra*).

A third type of target cell for transgene introduction is the embryonal stem cell (ES). ES cells are obtained from pre-implantation embryos cultured *in vitro* and fused with embryos (Evans et al. (1981) *Nature* 292:154-156; Bradley et al. (1984) *Nature* 309:255-258; Gossler et al. (1986) *PNAS* 83: 9065-9069; and Robertson et al. (1986) *Nature* 322:445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) *Science* 240:1468-1474.

Methods of making knock-out or disruption transgenic animals are also generally known. See, for example, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination to insert recombinase target sequences, such
5 that tissue specific and/or temporal control of inactivation of a RAP-BP gene can be controlled as above.

Another aspect of the present invention concerns a novel *in vivo* method for the isolation of genes encoding proteins which physically interact with a "bait" protein/drug complex. The method relies on detecting the reconstitution of a transcriptional activator in
10 the presence of the drug, particularly wherein the drug is a non-peptidyl small organic molecule (e.g. <2500K), e.g. a macrolide, e.g. rapamycin, FK506 or cyclosporin. In particular, the method makes use of chimeric genes which express hybrid proteins. The first hybrid comprises the DNA-binding domain of a transcriptional activator fused to the bait protein. The second hybrid protein contains a transcriptional activation domain fused to a
15 "fish" protein, e.g. a test protein derived from a cDNA library. If the fish and bait proteins are able to interact in a drug-dependent manner, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the marker gene can be detected and used to score
20 for the interaction of the bait protein/drug complex with another protein.

One advantage of this method is that a multiplicity of proteins can be simultaneously tested to determine whether any interact with the drug/protein complex. For example, a DNA fragment encoding the DNA-binding domain can be fused to a DNA fragment encoding the bait protein in order to provide one hybrid. This hybrid is introduced into the cells carrying
25 the marker gene, and the cells are contacted with a drug which is known to bind the bait protein. For the second hybrid, a library of plasmids can be constructed which may include, for example, total mammalian complementary DNA (cDNA) fused to the DNA sequence encoding the activation domain. This library is introduced into the cells carrying the first hybrid. If any individual plasmid from the test library encodes a protein that is capable of
30 interacting with the drug/protein complex, a positive signal may be obtained by detecting expression of the reporter gene. In addition, when the interaction between the drug complex and a novel protein occurs, the gene for the newly identified protein is readily available.

As illustrated herein, the present interaction trap system is a valuable tool in the identification of novel genes encoding proteins which act at a point in a given signal
35 transduction pathway that is directly upstream or downstream from a particular protein/drug complex. For example, the subject assay can be used to identify the immediate downstream targets of an FKBP/rapamycin complex, or of an FKBP/FK506 complex, or of a

cyclophilin/cyclosporin complex. Proteins that interact in a drug-dependent manner with one of such complexes may be identified, and these proteins can be of both diagnostic and therapeutic value.

5 A first chimeric gene is provided which is capable of being expressed in the host cell, preferably a yeast cell, most preferably *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*. The host cell contains a detectable gene having a binding site for the DNA-binding domain of the transcriptional activator, such that the gene expresses a marker protein when the marker gene is transcriptionally activated. Such activation occurs when the transcriptional activation domain of a transcriptional activator is brought into sufficient
10 proximity to the DNA-binding domain of the transcriptional activator. The first chimeric gene may be present in a chromosome of the host cell. The gene encodes a chimeric protein which comprises a DNA-binding domain that recognizes the binding site on the marker gene in the host cell and a bait protein which is to be tested for drug-mediated interaction with a second test protein or protein fragment.

15 A second chimeric gene is provided which is capable of being expressed in the host cell. In one embodiment, both the first and the second chimeric genes are introduced into the host cell in the form of plasmids. Preferably, however, the first chimeric gene is present in a chromosome of the host cell and the second chimeric gene is introduced into the host cell as part of a plasmid. The second chimeric gene contains a DNA sequence that encodes a second
20 hybrid protein. The second hybrid protein contains a transcriptional activation domain. The second hybrid protein also contains a second test protein or a protein fragment which is to be tested for interaction with the first test protein or protein fragment. Preferably, the DNA-binding domain of the first hybrid protein and the transcriptional activation domain of the second hybrid protein are derived from transcriptional activators having separate DNA-
25 binding and transcriptional activation domains. These separate DNA-binding and transcriptional activation domains are also known to be found in the yeast GAL4 protein, and are also known to be found in the yeast GCN4 and ADR1 proteins. Many other proteins involved in transcription also have separable binding and transcriptional activation domains which make them useful for the present invention. In another embodiment, the DNA-binding
30 domain and the transcriptional activation domain may be from different transcriptional activators. The second hybrid protein is preferably encoded on a library of plasmids that contain genomic, cDNA or synthetically generated DNA sequences fused to the DNA sequence encoding the transcriptional activation domain.

The drug-mediated interaction between the first test protein and the second test
35 protein in the host cell, therefore, causes the transcriptional activation domain to activate transcription of the detectable gene. The method is carried out by introducing the first chimeric gene and the second chimeric gene into the host cell, and contacting the cell with

54

the drug of interest. The host cell is subjected to conditions under which the first hybrid protein and the second hybrid protein are expressed in sufficient quantity for the detectable gene to be activated. The cells are then tested for drug-dependent expression of the detectable gene.

5 Thus, interactions between a first test protein and a library of proteins can be tested in the presence of the drug of interest, in order to determine which members of the library are involved in the formation of drug-dependent complexes between the first and second protein. For example, the bait protein may be a protein which binds FK506, rapamycin, or cyclosporin, e.g. can be an FKBP or cyclophilin. The second test protein may be derived
10 from a cDNA library.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of
15 certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1

Construction Of The Bait Plasmids For The 2-Hybrid Screen

20 *A. LexA-FKBP12 bait:*

The bait protein and fish protein constructs used in the present drug-dependent interaction trap are essentially the same as constructs used for other 2 hybrid assays (see, for example, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J Biol Chem* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; and
25 Iwabuchi et al. (1993) *Oncogene* 8:1693-1696). Using the following oligonucleotides:

coding strand

G GGT TTG GAA TTC CTA ATA ATG TCT GTA CAA GTA GAA ACC
(SEQ ID No: 3)

30

non-coding strand

GGG TTT CGG GAT CCC GTC ATT CCA GTT TTA GAA G
(SEQ ID No:4)

PCR amplification was carried out from a lymphocyte cDNA library to isolated the coding
35 sequence for the FKBP12 protein. The sequence of the human FKBP12 cloned was confirmed as:

55

ATGTCCGTACAAGTAGAAACCATCTcCCCAGGAGACGGGCGCACCTTcCCCCA
 AGCGCGGCCAGACCTGCGTGGTGCCTACACCGGgATGCTTGAAGATGGAAA
 GAAATTTGATTCTCCCGTGACCGTAACAAGCCCTTTAAGTTtATgCTAGGC
 aAGCAGGAGGTGATCCGAGGCTGGGAAGAagGGGTTGcCCAGATGAGTGTGG
 5 gTCAGCGTGCCAAaCTgACTAtAtCTCcAGaTtATgCcTATGgTGCCACTGG
 GCACcCAGGCATCATCCCACCACATGCCACTCTCGTCTTCGATGTGGAGCTT
 CTAAAACTGGAATGA (SEQ ID No: 5)

The resulting PCR product containing the human FKBP12 coding sequences was then digested with EcoRI and BamHI, and cloned into the EcoRI + BamHI sites of pBTM116 creating an in-frame fusion between LexA and FKBP12. The resulting plasmid is referred to below as plC504.

B. *LexA-(gly)₆-FKBP12 bait:*

In order to generate an in frame fusion between LexA and FKBP12 separated by six glycine residues, the coding sequence from human FKBP12 was cloned by PCR as above, except that the sense oligonucleotide provided an additional 18 nucleotides which inserted 6 glycines in the open reading frame of the fusion protein. The oligos used for PCR were:

coding strand

TCG CCG GAA TTC GGG GGC GGA GGT GGA GGA GTA CAA
 GTA GAA ACC ATC (SEQ ID No: 7)

non-coding strand

GGG TTT CGG GAT CCC GTC ATT CCA GTT TTA GAA G
 (SEQ ID No: 8)

The PCR product containing the human FKBP12 coding sequences was then digested with EcoRI and BamHI and cloned into the EcoRI + BamHI sites of pBTM116 as above. The resulting plasmid is referred to below as plC506.

Example 2

Construction of the FKBP12 deletion strain

A 1.8 kb HindIII-EcoRI yeast genomic fragment containing FKB1 (the *S. Cerevisia* homolog of FKBP12) was cloned into the HindIII + EcoRI sites of pSP72 (Promega).

A one-step PCR strategy was used to create a precise deletion of the FKB1 coding sequences extending from the ATG start codon to the TGA stop codon. Simultaneously a unique BamHI site was introduced in lieu of the FKB1 coding sequences. The oligos used to generate the FKB1 deletion and introduction of the unique BamHI site were:

56

CGCGGATCCGCGCATTATTACTTGTTTTGATTGATTTTTTG
(SEQ ID No: 9)

CGCGGATCCGCGGTAAAAGCAAAGTACTATCAATTGAGCCG
(SEQ ID No: 10)

5

The yeast ADE2 gene on a 3.6 kb BamHI fragment was then cloned into the unique BamHI site of the plasmid described above to generate the plasmid pVB172. Flanking the ADE2 disruption marker of pVB172 in the 5' and 3' noncoding sequence of FKB1 are XhoI sites. pVB172 was digested with XhoI to release a linear fragment containing ADE2 flanked
10 by FKB1 noncoding sequences. This linear fragment was used to transform yeast strain L40 (Mat a his3 Δ 200 trp1-901 leu2-3,112 ade2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ GAL4 gal80) selecting for adenine prototrophy.

ADE+ yeast transformants were tested for rapamycin resistance to confirm that the wild type FKB1 allele was replaced by ADE2. This disruption allele of FKB1 is designated
15 L40-fkbl-2.

Example 3

Cloning Of Mammalian Rapamycin Target Genes

We used the drug-dependent interaction trap described in Example 1 above, with the
20 LexA binding-domain fusion constructs as bait to detect interaction with clones from cDNA libraries containing VP16 activation-domain fusions. The reporters used as "read-outs" signaling interaction in this system are the *S. cerevisiae* HIS3 and the *E. coli* LacZ genes. The yeast strain L40, the bait vector plasmid pBTM116 and the mouse embryonic PCR library in the vector pVP16 were used to construct the cDNA fusion protein library

25 The strain L40-fkbl-2, described above in Example 2, was transformed with each of two bait plasmids, plC504, encoding the LexA-FKBP12 fusion protein, or plC506, encoding the LexA-(gly)6-FKBP12 fusion protein. The transformants, L40-fkbl-2/plC504 (named ICY99) and L40-fkbl-2/plC506 (named ICY101) were maintained on yeast media lacking tryptophan which selects for cells harboring the bait plasmid.

30 A mouse embryo PCR library in pVP16 (designated pSH10.5), which was generated by standard protocols using random-primed synthesis of 10.5 day-post-coital CD1 mouse embryo polyA+ RNA and size-selected for inserts between 350bp and 700bp in length, was used to transform the yeast ICY99 and ICY101. The transformed yeast cells were plated onto media lacking tryptophan and leucine. Approximately 10^7 transformants from each strain
35 were pooled, thoroughly mixed, and stored frozen in aliquots in 50% glycerol at -80°C.

Prior to screening, cells were thawed, grown for 5 hours in liquid medium, and plated onto selective medium. Approximately 1.5×10^7 ICY99/pSH10.5 cells were plated onto phosphate-buffered (pH7) synthetic agar medium containing (i) all amino acids except tryptophan, leucine and histidine, (ii) Rapamycin at 125 ng/ml, (iii) the chromogenic substrate X-gal at 100 ng/ml, and (iv) 2% glucose as carbon source, at a plating density of approximately 10^6 per 15 cm plate. An identical protocol was used for screening ICY101/pSH10.5 transformants, except that a lower concentration of rapamycin was used, at 15.6 ng/ml.

Colonies which both grew on the selective medium and were blue were picked for further testing. These represent cells which do not require histidine for growth and which are expressing the β -galactosidase reporter. Candidate colonies appeared between 4-11 days after plating, and the blue color ranged from very light blue to deep blue. They were then subjected to the following tests.

i) Rapamycin-dependence

Each candidate was streaked onto media lacking histidine and containing either 125ng/ml (for ICY99/pSH10.5 candidates), 15.6 ng/ml (for ICY101/pSH10.5 candidates) rapamycin, or no rapamycin (for both). Candidate clones which grew in the presence of rapamycin and failed to grow on media without rapamycin were chosen for the next test.

For the ICY99/pSH10.5 screen, out of 107 His⁺ and LacZ⁺ candidates screened, 24 were rapamycin-dependent for growth on medium lacking histidine. For the ICY101/pSH10.5 screen, 20 out of 101 His⁺ and LacZ⁺ candidates screened were rapamycin-dependent.

ii) plasmid-linkage

To eliminate false positives caused by chromosomal mutations, each candidate was grown in non-selective medium (YPD) to permit loss of the bait (Trp⁺) and the cDNA (Leu⁺) plasmids. Cells which had lost the bait plasmid (Trp⁻), the cDNA plasmid (Leu⁻) or both plasmids (Trp⁻ and Leu⁻), as well as those which had retained both plasmids (Trp⁺ and Leu⁺), were streaked onto media containing rapamycin but lacking histidine. Those candidates for which only the derivatives containing both plasmids (Trp⁺ and Leu⁺) grew, while the other three derivatives did not, were chosen for further analysis.

For the ICY99/pSH10.5 screen, 23 out of 24 passed the test. For the ICY101/pSH10.5 screen, all 20 passed the test.

iii) Positive and negative interaction with control baits

Whereas the previous test asked if the interaction disappears when either or both members of the interaction (bait and fish constructs) are lost, the present test asks if the candidate cDNA plasmid (Leu⁺) can confer interaction when transformed into yeast strains harboring various baits. DNA samples were prepared from each candidate and used to transform *E. coli* strain B290 (auxotrophic for tryptophan and leucine). Since the yeast TRP1 and LEU2 genes can complement the bacterial auxotrophies, respectively, B290 cells containing the bait plasmid are Trp⁺ and can grow on medium lacking tryptophan, while B290 cells containing the cDNA plasmid are Leu⁺ and can grow on medium lacking leucine. Plasmid DNA samples were each containing a different bait: i) ICY99, the original strain used in the screen, containing the LexA::FKBP12 bait fusion; ii) ICY101, containing the LexA::(gly)₆::FKBP12 bait fusion, and iii) ICY102, containing a LexA fusion bait irrelevant for the present study and which serves as a negative control. The ideal candidate clone should confer His⁺ and LacZ⁺ to ICY99 and ICY101 in a rapamycin-dependent manner, but not to ICY102.

For the ICY99/pSH10.5 screen, 11 out of the 23 candidates fulfilled the above criteria. For the ICY101/pSH10.5 screen, 10 out of the 20 candidates fulfilled the above criteria.

The cDNA inserts of these candidate clones were sequenced in both strands using the ABI fluorescent sequencing system. All 11 candidates from the ICY99/pSH10.5 screen, and at least 4 out of 10 of the candidates from the ICY101/pSH10.5 screen contain overlapping fragments of an identical sequence. The 14 clones represent at least 5 independent cloning events from the library as judged by the insert/vector boundaries of each clone. The longest and the shortest inserts differ by approximately 70 bp at the amino-terminus and about 10 bp at the amino-terminus. The partial nucleotide sequence, and corresponding amino acid sequence, isolated from the mouse rapamycin/FKBP12 binding protein (RAPT1), is given in SEQ ID No: 1 and SEQ ID No: 2, respectively.

Surprisingly, a search of the GenBank database using the program BLAST, revealed that the peptide encoded by the above sequence shares some homology, though less than 60 percent absolute homology, to the *S. cerevisiae* TOR1 (and DRR1) and TOR2 gene products previously isolated from yeast.

Example 4

Cloning of Human Homologs of Rapamycin Target Genes

Having isolated a partial sequence for the gene encoding a rapamycin-target-protein from a mouse library, we proceeded to isolate the human gene using the mouse sequence as a probe. The plasmid clone pIC99.1.5, containing the longest insert of the RAPT1 clone, was chosen as probe for hybridization. The insert (500 bp) was separated from plasmid DNA by

59

digestion with Not I restriction endonuclease followed by agarose gel electrophoresis and fragment purification. The fragment was radiolabelled with α^{32} -labeled dCTP by random-incorporation with the Klenow fragment of DNA polymerase. The radiolabelled DNA probe was isolated away from free nucleotides by a G50 column, alkali-denatured, and added to the hybridization mix at 2×10^6 cpm/ml.

Approximately 3×10^6 phage of a human B cell cDNA library in λ -pACT (Figure 1) were screened by filter hybridization using the probe described above, in 30% formamide, 5XSSC, 5X Denhardts, 20 μ g/ml denatured salmon sperm DNA, and 1 % SDS, at 37°C. Following hybridization, the filters were washed at 0.5xSSC and 0.1% SDS, at 50°C. These represent conditions of medium stringency appropriate for mouse-to-human cross-species hybridizations. A number of positive plaques were obtained, and several were analyzed. A number of the isolated clones turned out to be various 3' fragments of the same gene, or very closely related genes, which, after sequence analysis, was determined to be the human RAPT1 gene. The clone containing the longest coding sequence fragment, comprising what is believed to be roughly half the full-length protein (C-terminus) and including the FKBP/rapamycin binding site and the putative PI-kinase activity, is designated as plasmid pIC524. A deposit of the pACT[™] plasmid form of pIC524 was made with the American Type Culture Collection (Rockville, MD) on May 27, 1994, under the terms of the Budapest Treaty. ATCC Accession number 75787 has been assigned to the deposit.

Figure 1 is a map of the human RAPT1 clone of pIC524 (inserted at the XhoI site). The insert is approximately 3.74 kb in length, and nucleotide RAPT1 coding sequence from the insert has been obtained and is represented by nucleotide residues 4717-7746 of SEQ ID No. 11. The corresponding amino acid sequence is represented by residues His1541-Trp2549 of SEQ ID No. 12. The region of the human RAPT1 clone corresponding to the mouse RAPT1 fragment is greater than 95% homologous at the amino acid level and 90% homologous at the nucleotide level. In addition to the pIC524 clone, further 5' sequence of the human RAPT1 gene was obtained from other overlapping clones, with the additional sequence of the 3' end of the ~5.4kb partial gene given in SEQ ID No. 11. Furthermore, SEQ ID No. 19 provides additional 3' non-coding sequence (obtained from another clone) which flanks the RAPT1 coding sequence.

It will be evident to those skilled in the art that, given the present sequence information, PCR primers can be designed to amplify all, or certain fragments of the RAPT1 gene sequence provided in pIC524. For example, the primers TGAAGATACCCACCAA-ACCC (SEQ ID No. 21) and TGCACAGTTGAAGTGAAC (SEQ ID No. 22) correspond to pACT[™] sequences flanking the XhoI site, and can be used to PCR amplify the entire RAPT1 sequence from pIC524. Alternatively, primers based on the nucleic acid sequence of SEQ ID No. 11 can be used to amplify fragments of the RAPT1 gene in pIC524. The PCR primers

can be subsequently sub-cloned into expression vectors, and used to produce recombinant forms of the subject RAPT1 protein. Thus, the present provides recombinant RAPT1 proteins encoded by recombinant genes comprising RAPT1 nucleotide sequences from ATCC deposit number 75787. Moreover, it is clear that primer/probes can be generated which include even those portion of pIC524 not yet sequenced by simply providing PCR primers based on the known sequences.

Furthermore, our preliminary data indicate that other proteins which are related to RAPT1, e.g. RAPT1 homologs, were also obtained from the present assay, suggesting that RAPT1 is a member of a larger family of related proteins.

Example 5

Cloning of Novel Human Ubiquitin Conjugating Enzyme

Constructs similar to those described above for the drug-dependent interaction trap assay were used to screen a WI38 (mixed G₀ and dividing fibroblast) cDNA library (Clontech, Palo Alto CA) in pGADGH (XhoI insert, Clontech). Briefly, the two hybrid assay was carried out as above, using GAL4 constructs instead of LexA, and in an HF7C yeast cell (Clontech) in which FKB1 gene was disrupted (see Example 1). Of the clones isolated, a novel human ubiquitin-conjugating enzyme (rap-UBC) has been identified. A deposit of the pGADGH plasmid (clone "SMR4-15") was made with the American Type Culture Collection (Rockville, MD) on May 27, 1994, under the terms of the Budapest Treaty. ATCC Accession number 75786 has been assigned to the deposit. The insert is approximately 1kB.

The sequence UBC-encoding portion of the SMR4-15 insert is given by SEQ ID No. 23 (nucleotide) and SEQ ID No. 24 (amino acid). The sequence for the 3' portion of the clone is provided by SEQ ID No. 25. As described above, primers based on the nucleic acid sequence of SEQ ID No. 23 (and 25) can be used to amplify fragments of the rap-UBC gene from SMR4-15. The PCR primers can be subsequently sub-cloned into expression vectors, and used to produce recombinant forms of the subject enzyme. Thus, the present provides recombinant rap-UBC proteins encoded by recombinant genes comprising rap-UBC nucleotide sequences from ATCC deposit number 75786.

Example 6

Construction of the Serine-to-Arginine RAPT1 mutation

The smallest mRAPT1 clone that interacted with the FKBP12/rapamycin complex was 399 bp, defining a rapamycin binding domain. The RAPT1 binding domain corresponds to a region in yeast TOR1/TOR2 located immediately upstream, but outside of the lipid

kinase consensus sequence. This region contains the serine residue which when mutated in yeast TOR1 confers resistance to rapamycin (Cafferkey et al. (1993) *Mol Cell Biol* 13:6012-6023). Both a mouse and human RAPT1 serine-to-arginine mutation was constructed by oligonucleotide mutagenesis. In the instance of the mRAPT1 mutant, coding and noncoding strand oligonucleotides containing the mutations were: GAAGAGGCAAGACGCTTGTAC (SEQ ID NO:26) and GTACAAGCGTCTTGCCTCTTC (SEQ ID NO:27). PCR reactions were performed using these oligonucleotides in combination with oligonucleotides GAGTTTGAGCAGATGTTTA (SEQ ID NO:28) and the M13 universal primer which are sequences in the pVP16 vector, 5' and 3' of the mRAPT1 insert, respectively. pVP16 containing mRAPT1 was used as the template for PCR. The PCR product, digested with BamHI and EcoRI, was cloned into the BamHI and EcoRI sites in pVP16. The resulting clone was sequenced to verify that the clone contained the serine-to-arginine mutation and no others.

The smallest mRAPT1 clone that interacted with the FKBP12/rapamycin complex was 399 bp, defining the RAPT1 binding domain. The RAPT1 binding domain corresponds to a region in yeast TOR located immediately upstream, but outside of the lipid kinase consensus sequence. This region contains the serine residue which when mutated in yeast TOR1 (also called DRR1) confers resistance to rapamycin (Cafferkey et al. (1993) *Mol. Cell Biol.* 13:6012-6023; Helliwell et al. (1994) *Mol. Cell Biol.* 5:105-118). The corresponding mutation was constructed in mRAPT1. The serine-to-arginine mutation abolishes interaction of mRAPT1 with the FKBP12/rapamycin complex (see Figure 3), activating neither *HIS3* nor *lacZ* expression on the two-hybrid assay, indicating that the serine is involved in the association of the FKBP12/rapamycin complex with mRAPT1.

Example 7

Northern Analysis

The multiple tissue Northern blots (containing 2 µg of human RNA per lane) were obtained from Clontech Labs., Inc. Hybridizations were at 42°C in 5X SSPE, 5X Denhardt's, 30% formamide, 1% SDS and 200 µg/ml denatured salmon sperm DNA. Washes were at 0.1X SSC and 0.1% SDS at 55°C. The blot was exposed for 5 days prior to autoradiography. The levels of RNA loaded in each lane were independently monitored by hybridizing the same blots with a human *G3PDH* probe and were found to be similar in all lanes, with the exception of skeletal muscle, which had approximately 2-3 fold the signal.

RAPT1 specifies a single transcript of approximately 9 kb that is present in all tissues examined, exhibiting the highest levels in testis. The transcript is sufficient to encode a protein equivalent to the size of yeast TOR which is 284 kDa. Assuming that RAPT1 is of

62

similar size. a small fragment of 133 amino acids has been cloned from within a large protein. but which fragment is sufficient to bind FKBP12/rapamycin complex.

Example 8

5 High throughput assay based on the two-hybrid system for identifying novel rapamycin analogs.

To develop a high throughput screen based on the two-hybrid system, we devised a procedure to quantitate protein-protein interaction mediated by a small molecule. Since protein-protein interaction in the two-hybrid system stimulates transcription of the *lacZ* reporter gene, the assay utilizes a substrate of β -galactosidase (the *lacZ* gene product *lacZ* gene product) which when cleaved produces a chemiluminescent signal that can be quantitated. This assay can be performed in microtiter plates, allowing thousands of compounds to be screened per week. The assay includes the following steps:

1. Inoculate yeast cells from a single colony into 50 ml of growth medium, synthetic complete medium lacking leucine and tryptophan (Sherman, F. (1991) *Methods Enzymol.* 194:3-20). Incubate the flask overnight at 30°C with shaking (~200 rpm).
2. Dilute the overnight culture to a final A_{600} of 0.02 in growth medium and incubate overnight as described in step 1.
3. Dilute the second overnight culture to a final A_{600} of 0.5 in growth medium. Using a Quadra 96™ pipettor (TomTec, Inc.), dispense 135 μ l aliquots of the cell suspension into wells of a round bottom microtiter plate pre-loaded with 15 μ l/well of the compound to be tested at various concentrations. (The compounds are dissolved in 5% dimethyl sulfoxide, so that the final DMSO concentration added to cells is 0.5% which does not perturb yeast cell growth.) Cover microtiter plates and incubate at 30°C for 4 hr with shaking at 300 rpm.
4. Centrifuge microtiter plate for 10 min at 2000 rpm. Remove the supernatant with the Quadra 96™ pipettor and wash with 225 μ l phosphate buffered saline.
5. Dispense 100 μ l of lysis buffer (100mM $_2$ HPO $_4$ pH 7.8; 0.2% Triton X-100; 1.0 mM dithiothriitol) into each well, cover, and incubate for 30 min at room temperature with shaking at 300 rpm.
6. Dispense into each well of a Microfluor™ plate (Dynatech Laboratories, Chantilly, VA), 50 μ l of the chemiluminescent substrate, Galacton Plus™ (Tropix, Inc., Bedford, MA) in diluent (100 mM Na $_2$ HPO $_4$, 1 mM MgCl $_2$, pH 8.0). To these wells, transfer 20 μ l of cell lysate and incubate in the dark for 60 min at room temperature.

63

7. Add to each well 75 μ l of Emerald™ accelerator. Cover plate and count in a Topcount™ scintillation counter (Packard, Inc.) for 0.01 min/well.

The rapamycin target proteins, isolated as described above, were incorporated into the quantitative assay, as was a variety of FKBP. The FKBP included in the screen were human FKBP12 and that from pathogenic fungi, FKBP13 (Jin et al. (1991) *Proc. Natl. Acad. Sci.* 88:6677) and FKBP25 (Jin et al. (1992) *J. Biol. Chem.* 267:2942; Galat et al. (1992) *Biochem.* 31:2427-2434). Yeast strains containing different FKBP-target pairs can be tested against libraries of rapamycin and FK506 analogs. Such a screen can yield different classes of compounds including (i) target-specific compounds, those that mediate interaction between a specific target and more than one FKBP, (ii) FKBP-specific compounds, those that mediate interaction between a particular FKBP and more than one target and, most ideally, (iii) FKBP/target-specific compounds, those that mediate interaction between a particular FKBP and target. The protein interactions mediated by the test compounds and measured in this assay can be correlated with immunosuppressive, antifungal, antiproliferative and toxicity profiles, as well as their K_i 's for inhibition of FKBP PPIase activity.

Using the quantitative chemiluminescence assay described above, the interaction of human LexA-FKBP12 and VP16-RAPT1 was analyzed in the presence and absence of rapamycin. Interaction between FKBP12 and RAPT1 was measured as a function of drug concentration. Addition of rapamycin from 0 to 500 ng/ml increased β -galactosidase activity approximately one thousand-fold. This effect was specific for rapamycin; FK506 over the same concentration range did not increase β -galactosidase activity significantly over background levels. If lexA-da, a control construct, is substituted for the lexA-FKBP12, β -galactosidase activity does not increase as a function of rapamycin addition. The basal levels of β -galactosidase in the negative controls are 0.1 per cent of the maximum levels detected in the yeast strain containing the FKBP12 and RAPT1 constructs, grown in media containing 500 ng/ml rapamycin. These results, illustrated in Figure 2, indicate that protein interactions mediated by a small molecule in the two-hybrid system can be quantitated and assayed in a microtiter format that can be used for high throughput screening. Employing various FKBP and RAPT1 proteins in the two-hybrid format (Figure 3) rapamycin-mediated interactions were measured in this quantitative assay.

Example 9

In vitro protein interactions mediated by rapamycin

Drug-mediated interactions of FK506-binding proteins and the RAPT1 proteins is analyzed *in vitro* using purified FKBP12 fused to glutathione-S-transferase (GST) and ^{35} S labeled RAPT1 proteins prepared by *in vitro* transcription and translation. For this purpose FKBP12 is fused in the frame of GST in pGEX™ (Pharmacia, Piscataway, NJ). GST-FKBP12

64

fusion proteins are expressed and purified from *E. coli* (Vojtek et al. (1993) *Cell* 74:205-214). RAPT1 coding sequences are cloned behind the CMV and T7 promoters in the mammalian expression vector, pX (Superti-Furga et al. (1991) *J. Immunol. Meths.* 151:237-244). RAPT1 sequences are transcribed from the T7 promoter and translated *in vitro* using commercially available reagents (Promega, Madison, WI) in a reaction containing ³⁵S-methionine. For *in vitro* binding (Toyoshima et al. (1994) *Cell* 78:67-74), 5 to 20 µl of the *in vitro* transcription/translation reactions are added to 200 µl of binding buffer (20mM HEPES[pH7.4], 150 mM NaCl, 10% glycerol, 0.05% NP-40TM). After addition of 10 µl of GST-FKBP12 bound to glutathione-agarose beads, the reaction is incubated at 4°C for 2 hr with rotation. Various contrations of drug are added to reactions, such as 0.1 to 10-fold that of FKBP12 on a molar basis. No drug is added to control reactions. The agarose beads are then precipitated and washed four times with binding buffer. Bound proteins iseluted by boiling in Laemmli sample buffer, resolved on 4-20% gradient SDS polyacrylamide gels, and visualized by autoradiography. Detection of ³⁵S-labelled RAPT1 protein from binding reactions containing drug demonstrates direct binding to FKBP12 as a function of drug.

Example 10

Effect of RAPT1 mutations on complex formation and rapamycin sensitivity

To more particularly map the rapamycin-binding domain of RAPT1 requires the isolation of mutants that fail to bind to a FKBP/rapamycin complex. As described in the Examples above, association with the FKBP/rapamycin can be tested in the LexA two-hybrid system in which FKBP12 is expressed as a fusion to LexA and RAPT1 proteins are expressed as fusions to the VP16 activation domain. Accordingly, a library of mutant RAPT1 proteins is generated by mutagenizing coding sequences through PCR-generated random mutagenesis (Cadwell and Joyce (1992) *PCR Methods Appl* 2:28-33). The 5' and 3' oligos for PCR contain BamHI and EcoRI restriction sites, respectively, that allow subsequent cloning of the PCR products into pVP16 creating an in-frame fusion. In addition, the 3' oligo contains a 27 bp HA epitope sequence followed by an in frame stop codon. The addition of the HA epitope tag to the C-terminal end of the fusion proteins allows the characterization of the mutant RAPT1 proteins (see below).

Upon completion of the mutagenesis, the EcoRI-BamHI digested PCR products are inserted into pVP16. The library of mutant RAPT1 proteins is amplified by transformation into *E. coli*. To identify those mutations that impair the ability of a RAPT1 to interact with an FKBP/rapamycin complex, the mutagenized RAPT1 library is introduced into a yeast strain containing the LexA-FKBP bait protein. Each transformed cell carries one individual mutant RAPT1 fused to the transcriptional activator VP16. Interaction between the FKBP and wild type RAPT1 occurs when cells are grown in media containing rapamycin, inducing

65

5 *lacZ* expression and turning colonies blue on X-GAL indicator plates. Colonies in which the interaction between an FKBP/rapamycin complex and the RAPT1 mutant is impaired are light blue or white. Two classes of mutations can produce this phenotype: nonsense mutations resulting in truncated version of RAPT1 or sense mutations that affect the binding
10 of RAPT1 to the FKBP/rapamycin complex. To distinguish between these two types of mutations, total protein extracts made from these colonies is subjected to Western blot analysis using an anti-HA antibody. Nonsense mutations that give rise to shorter, truncated proteins do not contain the HA epitope at their C-terminus and thus are not detected by the anti-HA antibody. Conversely, full-length proteins with an incorporated sense mutations are
10 detected with this antibody.

The library plasmids from the light blue or white colonies that express full-length RAPT1 protein with the HA epitope are rescued by retransformation into *E. Coli*. The position of the mutation is determined by sequence analysis, and the phenotype verified by retransformation of these plasmids back into the yeast strain containing LexA-FKBP12.

15 Mutants that retest can also be cloned into the mammalian expression vector, pX. pX-RAPT1 or pX lacking RAPT1 sequences, are then introduced into the lymphoid (CTLL and Kit225) and nonlymphoid cells (MG63 and RH30) sensitive to rapamycin. The effect of the mutation on rapamycin sensitivity is measured in terms of inhibition of DNA synthesis monitored by BrdU incorporation. Mutants that confer resistance of rapamycin by virtue of
20 being unable to bind to the FKBP12/rapamycin complex indicate which mutations mediate drug sensitivity in lymphoid and nonlymphoid cells. Of particular interest is whether different RAPT1s mediate drug sensitivity in different cell types.

Example 11

25 *Cloning of a RAPT1-like polypeptide from Candida albicans*

In order to clone homologs of the RAPT1 genes from human pathogen *Candida*, degenerate oligonucleotides based on the conserved regions of the RAPT1 and TOR proteins were designed and used to amplify *C. albicans* cDNA in λ ZAP (strain 3153A). The amplification consisted of 30 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 1
30 minute with the PCR amplimers GGNAARGCNCA YCCNCARGC and ATNGCNGGRTAYTGYTG DATNTC. The PCR reactions were separated on a 2.5% low melting agarose gel, that identified a sizable fragment. The fragment was eluted and cloned into pCRII (TA cloning system, Invitrogen corporation).

35 The *C. albicans* DNA probes were ³²P-labeled by nick translation and used on Southern blots to confirm the species identity of the fragments and to further screen *C. albicans* cDNA libraries. Sequencing of the larger cDNAs confirmed the identity of the

66

clones. The partial sequence of a *C. albicans* RAPT1-like polypeptide, with the open-reading frame designated, is provided by SEQ ID Nos. 13 and 14.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

67
SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT:
(A) NAME: Mitotix, Inc.
(B) STREET: One Kendall Square, Building 600
(C) CITY: Cambridge
(D) STATE: MA
10 (E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02139
(G) TELEPHONE: (617) 225-0001
(H) TELEFAX: (617) 225-0005
- 15 (ii) TITLE OF INVENTION: Immunosuppressant Target Proteins
- (iii) NUMBER OF SEQUENCES: 25
- (iv) COMPUTER READABLE FORM:
20 (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: ASCII (text)
- 25 (vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/250,795
(B) FILING DATE: 27-MAY-1994
- 30 (vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/250,795
(B) FILING DATE: 20-DEC-1994

(2) INFORMATION FOR SEQ ID NO:1:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
40 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
45 (A) NAME/KEY: CDS
(B) LOCATION: 1..486
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

50
CTC ACC CGT CAC AAT GCA GCC AAC AAG ATC TTG AAG AAC ATG TGT GAA 48
Leu Thr Arg His Asn Ala Ala Asn Lys Ile Leu Lys Asn Met Cys Glu
1 5 10 15

55
CAC AGC AAC ACG CTG GTC CAG CAG GCC ATG ATG GTG AGT GAA GAG CTG 96
His Ser Asn Thr Leu Val Gln Gln Ala Met Met Val Ser Glu Glu Leu
20 25 30

68

	ATT CGG GTA GCC ATC CTC TGG CAT GAG ATG TGG CAT GAA GGC CTG GAA	144
	Ile Arg Val Ala Ile Leu Trp His Glu Met Trp His Glu Gly Leu Glu	
	35 40 45	
5	GAG GCA TCT CGC TTG TAC TTT GGG GAG AGG AAC GTG AAA GGC ATG TTT	192
	Glu Ala Ser Arg Leu Tyr Phe Gly Glu Arg Asn Val Lys Gly Met Phe	
	50 55 60	
10	GAG GTG CTG GAG CCC CTG CAT GCT ATG ATG GAA CGG GGT CCC CGG ACT	240
	Glu Val Leu Glu Pro Leu His Ala Met Met Glu Arg Gly Pro Arg Thr	
	65 70 75 80	
15	CTG AAG GAA ACA TCC TTT AAT CAG GCA TAT GGC CGA GAT TTA ATG GAG	288
	Leu Lys Glu Thr Ser Phe Asn Gln Ala Tyr Gly Arg Asp Leu Met Glu	
	85 90 95	
20	GCA CAA GAA TGG TGT CGA AAG TAC ATG AAG TCG GGG AAC GTC AAG GAC	336
	Ala Gln Glu Trp Cys Arg Lys Tyr Met Lys Ser Gly Asn Val Lys Asp	
	100 105 110	
25	CTC ACG CAA GCC TGG GAC CTC TAC TAT CAC GTG TTC AGA CGG ATC TCA	384
	Leu Thr Gln Ala Trp Asp Leu Tyr Tyr His Val Phe Arg Arg Ile Ser	
	115 120 125	
30	AAG CAG CTA CCC CAG CTC ACA TCC CTG GAG CTG CAG TAT GTG TCC CCC	432
	Lys Gln Leu Pro Gln Leu Thr Ser Leu Glu Leu Gln Tyr Val Ser Pro	
	130 135 140	
35	AAA CTT CTG ATG TGC CGA GAC CTT GAG TTG GCT GTG CCA GGA ACA TAC	480
	Lys Leu Leu Met Cys Arg Asp Leu Glu Leu Ala Val Pro Gly Thr Tyr	
	145 150 155 160	
35	GAC CCC	486
	Asp Pro	

(2) INFORMATION FOR SEQ ID NO:2:

40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 162 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

50	Leu Thr Arg His Asn Ala Ala Asn Lys Ile Leu Lys Asn Met Cys Glu	1 5 10 15
	His Ser Asn Thr Leu Val Gln Gln Ala Met Met Val Ser Glu Glu Leu	20 25 30
55	Ile Arg Val Ala Ile Leu Trp His Glu Met Trp His Glu Gly Leu Glu	35 40 45
	Glu Ala Ser Arg Leu Tyr Phe Gly Glu Arg Asn Val Lys Gly Met Phe	

69

50

55

60

Glu Val Leu Glu Pro Leu His Ala Met Met Glu Arg Gly Pro Arg Thr
65 70 75 80

Leu Lys Glu Thr Ser Phe Asn Gln Ala Tyr Gly Arg Asp Leu Met Glu
85 90 95

Ala Gln Glu Trp Cys Arg Lys Tyr Met Lys Ser Gly Asn Val Lys Asp
100 105 110

Leu Thr Gln Ala Trp Asp Leu Tyr Tyr His Val Phe Arg Arg Ile Ser
115 120 125

Lys Gln Leu Pro Gln Leu Thr Ser Leu Glu Leu Gln Tyr Val Ser Pro
130 135 140

Lys Leu Leu Met Cys Arg Asp Leu Glu Leu Ala Val Pro Gly Thr Tyr
145 150 155 160

Asp Pro

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGTTTGGAA TTCCTAATAA TGTCTGTACA AGTAGAAACC

40

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGTTTCGGG ATCCCGTCAT TCCAGTTTCA CAAC

34

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:

70

- (A) LENGTH: 348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

10

- (A) NAME/KEY: CDS
(B) LOCATION: 14..325

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

15

GGAATTCCTA ATA ATG TCC GTA CAA GTA GAA ACC ATC TCC CCA GGA GAC 49
Met Ser Val Gln Val Glu Thr Ile Ser Pro Gly Asp
1 5 10

20

GGG CGC ACC TTC CCC AAG CGC GGC CAG ACC TGC GTG GTG CAC TAC ACC 97
Gly Arg Thr Phe Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr
15 20 25

25

GGG ATG CTT GAA GAT GGA AAG AAA TTT GAT TCC TCC CGT GAC CGT AAC 145
Gly Met Leu Glu Asp Gly Lys Lys Phe Asp Ser Ser Arg Asp Arg Asn
30 35 40

30

AAG CCC TTT AAG TTT ATG CTA GGC AAG CAG GAG GTG ATC CGA GGC TGG 193
Lys Pro Phe Lys Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp
45 50 55 60

35

GAA GAA GGG GTT GCC CAG ATG AGT GTG GGT CAG CGT GCC AAA CTG ACT 241
Glu Glu Gly Val Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr
65 70 75

ATA TCT CCA GAT TAT GCC TAT GGT GCC ACT GGG CAC CCA GGC ATC ATC 289
Ile Ser Pro Asp Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile
80 85 90

40

CCA CCA CAT GCC ACT CTC GTC TTC GAT GTG GAG CTT CTAAACTGG 335
Pro Pro His Ala Thr Leu Val Phe Asp Val Glu Leu
95 100

45

AATGACGGGA TCC 348

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ser Val Gln Val Glu Thr Ile Ser Pro Gly Asp Gly Arg Thr Phe

71

1

5

10

15

Pro Lys Arg Gly Gln Thr Cys Val Val His Tyr Thr Gly Met Leu Glu
20 25 30

5

Asp Gly Lys Lys Phe Asp Ser Ser Arg Asp Arg Asn Lys Pro Phe Lys
35 40 45

10

Phe Met Leu Gly Lys Gln Glu Val Ile Arg Gly Trp Glu Glu Gly Val
50 55 60

Ala Gln Met Ser Val Gly Gln Arg Ala Lys Leu Thr Ile Ser Pro Asp
65 70 75 80

15

Tyr Ala Tyr Gly Ala Thr Gly His Pro Gly Ile Ile Pro Pro His Ala
85 90 95

Thr Leu Val Phe Asp Val Glu Leu
100

20

(2) INFORMATION FOR SEQ ID NO:7:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: other nucleic acid

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCGCCGGAAT TCGGGGGCGG AGGTGGAGGA GTACAAGTAG AAACCATC

48

(2) INFORMATION FOR SEQ ID NO:8:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: other nucleic acid

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGTTTCGGG ATCCCGTCAT TCCAGTTTGA GAAG

34

(2) INFORMATION FOR SEQ ID NO:9:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: nucleic acid

72

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGCGGATCCG CGCATTATTA CTTGTTTTGA TTGATTTTTT G

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CGCGGATCCG CGTAAAAGCA AAGTACTATC AATTGAGCCG

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7824 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 97..7743

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAGGCGGGCG GTGGGGCACG GGGCCTGAAG CGGCGGTACC GGTGCTGGCG GCGGCAGCTG 60

AGGCCTTGGC CGAAGCCGCG CGAACCTCAG GGCAAG ATG CTT GGA ACC GGA CCT 114

Met Leu Gly Thr Gly Pro
1 5

GCC GCC GCC ACC ACC GCT GCC ACC ACA TCT AGC AAT GTG AGC GTC CTG 162

Ala Ala Ala Thr Thr Ala Ala Thr Thr Ser Ser Asn Val Ser Val Leu
10 15 20

CAG CAG TTT GCC AGT GGC CTA AAG AGC CGG AAT GAG GAA ACC AGG GCC 210

Gln Gln Phe Ala Ser Gly Leu Lys Ser Arg Asn Glu Glu Thr Arg Ala

73

	25	30	35	
5	AAA GCC GCC AAG GAG CTC CAG CAC TAT GTC ACC ATG GAA CTC CGA GAG Lys Ala Ala Lys Glu Leu Gln His Tyr Val Thr Met Glu Leu Arg Glu	258		
	40	45	50	
10	ATG AGT CAA GAG GAG TCT ACT CGC TTC TAT GAC CAA CTG AAC CAT CAC Met Ser Gln Glu Glu Ser Thr Arg Phe Tyr Asp Gln Leu Asn His His	306		
	55	60	65	70
15	ATT TTT GAA TTG GTT TCC AGC TCA GAT GCC AAT GAG AGG AAA GGT GGC Ile Phe Glu Leu Val Ser Ser Ser Asp Ala Asn Glu Arg Lys Gly Gly	354		
	75	80	85	
20	ATC TTG GCC ATA GCT AGC CTC ATA GGA GTG GAA GGT GGG AAT GCC ACC Ile Leu Ala Ile Ala Ser Leu Ile Gly Val Glu Gly Gly Asn Ala Thr	402		
	90	95	100	
25	CGA ATT GGC AGA TTT GCC AAC TAT CTT CGG AAC CTC CTC CCC TCC AAT Arg Ile Gly Arg Phe Ala Asn Tyr Leu Arg Asn Leu Leu Pro Ser Asn	450		
	105	110	115	
30	GAC CCA GTT GTC ATG GAA ATG GCA TCC AAG GCC ATT GGC CGT CTT GCC Asp Pro Val Val Met Glu Met Ala Ser Lys Ala Ile Gly Arg Leu Ala	498		
	120	125	130	
35	ATG GCA GGG GAC ACT TTT ACC GCT GAG TAC GTG GAA TTT GAG GTG AAG Met Ala Gly Asp Thr Phe Thr Ala Glu Tyr Val Glu Phe Glu Val Lys	546		
	135	140	145	150
40	CGA GCC CTG GAA TGG CTG GGT GCT GAC CGC AAT GAG GGC CGG AGA CAT Arg Ala Leu Glu Trp Leu Gly Ala Asp Arg Asn Glu Gly Arg Arg His	594		
	155	160	165	
45	GCA GCT GTC CTG GTT CTC CGT GAG CTG GCC ATC AGC GTC CCT ACC TTC Ala Ala Val Leu Val Leu Arg Glu Leu Ala Ile Ser Val Pro Thr Phe	642		
	170	175	180	
50	TTC TTC CAG CAA GTG CAA CCC TTC TTT GAC AAC ATT TTT GTG GCC GTG Phe Phe Gln Gln Val Gln Pro Phe Phe Asp Asn Ile Phe Val Ala Val	690		
	185	190	195	
55	TGG GAC CCC AAA CAG GCC ATC CGT GAG GGA GCT GTA GCC GCC CTT CGT Trp Asp Pro Lys Gln Ala Ile Arg Glu Gly Ala Val Ala Ala Leu Arg	738		
	200	205	210	
60	GCC TGT CTG ATT CTC ACA ACC CAG CGT GAG CCG AAG GAG ATG CAG AAG Ala Cys Leu Ile Leu Thr Thr Gln Arg Glu Pro Lys Glu Met Gln Lys	786		
	215	220	225	230
65	CCT CAG TGG TAC AGG CAC ACA TTT GAA GAA GCA GAG AAG GGA TTT GAT Pro Gln Trp Tyr Arg His Thr Phe Glu Glu Ala Glu Lys Gly Phe Asp	834		
	235	240	245	
70	GAG ACC TTG GCC AAA GAG AAG GGC ATG AAT CGG GAT GAT CGG ATC CAT Glu Thr Leu Ala Lys Glu Lys Gly Met Asn Arg Asp Asp Arg Ile His	882		
	250	255	260	

74

	GGA	GCC	TTG	TTG	ATC	CTT	AAC	GAG	CTG	GTC	CGA	ATC	AGC	AGC	ATG	GAG	930
	Gly	Ala	Leu	Leu	Ile	Leu	Asn	Glu	Leu	Val	Arg	Ile	Ser	Ser	Met	Glu	
			265					270					275				
5	GGA	GAG	CGT	CTG	AGA	GAA	GAA	ATG	GAA	GAA	ATC	ACA	CAG	CAG	CAG	CTG	978
	Gly	Glu	Arg	Leu	Arg	Glu	Glu	Met	Glu	Glu	Ile	Thr	Gln	Gln	Gln	Leu	
		280					285				290						
10	GTA	CAC	GAC	AAG	TAC	TGC	AAA	GAT	CTC	ATG	GGC	TTC	GGA	ACA	AAA	CCT	1026
	Val	His	Asp	Lys	Tyr	Cys	Lys	Asp	Leu	Met	Gly	Phe	Gly	Thr	Lys	Pro	
	295					300					305					310	
15	CGT	CAC	ATT	ACC	CCC	TTC	ACC	AGT	TTC	CAG	GCT	GTA	CAG	CCC	CAG	CAG	1074
	Arg	His	Ile	Thr	Pro	Phe	Thr	Ser	Phe	Gln	Ala	Val	Gln	Pro	Gln	Gln	
				315						320					325		
20	TCA	AAT	GCC	TTG	GTG	GGG	CTG	CTG	GGG	TAC	AGC	TCT	CAC	CAA	GGC	CTC	1122
	Ser	Asn	Ala	Leu	Val	Gly	Leu	Leu	Gly	Tyr	Ser	Ser	His	Gln	Gly	Leu	
			330						335					340			
25	ATG	GGA	TTT	GGG	ACC	TCC	CCC	AGT	CCA	GCT	AAG	TCC	ACC	CTG	GTG	GAG	1170
	Met	Gly	Phe	Gly	Thr	Ser	Pro	Ser	Pro	Ala	Lys	Ser	Thr	Leu	Val	Glu	
		345					350				355						
30	AGC	CGG	TGT	TGC	AGA	GAC	TTG	ATG	GAG	GAG	AAA	TTT	GAT	CAG	GTG	TGC	1218
	Ser	Arg	Cys	Cys	Arg	Asp	Leu	Met	Glu	Glu	Lys	Phe	Asp	Gln	Val	Cys	
	360					365					370						
35	CAG	TGG	GTG	CTG	AAA	TGC	AGG	AAT	AGC	AAG	AAC	TCG	CTG	ATC	CAA	ATG	1266
	Gln	Trp	Val	Leu	Lys	Cys	Arg	Asn	Ser	Lys	Asn	Ser	Leu	Ile	Gln	Met	
	375				380					385						390	
40	ACA	ATC	CTT	AAT	TTG	TTG	CCC	CGC	TTG	GCT	GCA	TTC	CGA	CCT	TCT	GCC	1314
	Thr	Ile	Leu	Asn	Leu	Leu	Pro	Arg	Leu	Ala	Ala	Phe	Arg	Pro	Ser	Ala	
				395					400						405		
45	TTC	ACA	GAT	ACC	CAG	TAT	CTC	CAA	GAT	ACC	ATG	AAC	CAT	GTC	CTA	AGC	1362
	Phe	Thr	Asp	Thr	Gln	Tyr	Leu	Gln	Asp	Thr	Met	Asn	His	Val	Leu	Ser	
			410					415					420				
50	TGT	GTC	AAG	AAG	GAG	AAG	GAA	CGT	ACA	GCG	GCC	TTC	CAA	GCC	CTG	GGG	1410
	Cys	Val	Lys	Lys	Glu	Lys	Glu	Arg	Thr	Ala	Ala	Phe	Gln	Ala	Leu	Gly	
		425					430				435						
55	CTA	CTT	TCT	GTG	GCT	GTG	AGG	TCT	GAG	TTT	AAG	GTC	TAT	TTG	CCT	CGC	1458
	Leu	Leu	Ser	Val	Ala	Val	Arg	Ser	Glu	Phe	Lys	Val	Tyr	Leu	Pro	Arg	
	440					445					450						
60	GTG	CTG	GAC	ATC	ATC	CGA	GCG	GCC	CTG	CCC	CCA	AAG	GAC	TTC	GCC	CAT	1506
	Val	Leu	Asp	Ile	Ile	Arg	Ala	Ala	Leu	Pro	Pro	Lys	Asp	Phe	Ala	His	
	455					460					465					470	
65	AAG	AGG	CAG	AAG	GCA	ATG	CAG	GTG	GAC	GCC	ACA	GTC	TTC	ACT	TGC	ATC	1554
	Lys	Arg	Gln	Lys	Ala	Met	Gln	Val	Asp	Ala	Thr	Val	Phe	Thr	Cys	Ile	
				475				480							485		
70	AGC	ATG	CTG	GCT	CGA	GCA	ATG	GGG	CCA	GGC	ATC	CAG	CAG	GAT	ATC	AAG	1602
	Ser	Met	Leu	Ala	Arg	Ala	Met	Gly	Pro	Gly	Ile	Gln	Gln	Asp	Ile	Lys	

		490		495	75		500	
5	GAG CTG CTG GAG CCC ATG CTG GCA GTG GGA CTA AGC CCT GCC CTC ACT							1650
	Glu Leu Leu Glu Pro Met Leu Ala Val Gly Leu Ser Pro Ala Leu Thr							
		505		510			515	
10	GCA GTG CTC TAC GAC CTG AGC CGT CAG ATT CCA CAG CTA AAG AAG GAC							1698
	Ala Val Leu Tyr Asp Leu Ser Arg Gln Ile Pro Gln Leu Lys Lys Asp							
		520		525			530	
15	ATT CAA GAT GGG CTA CTG AAA ATG CTG TCC CTG GTC CTT ATG CAC AAA							1746
	Ile Gln Asp Gly Leu Leu Lys Met Leu Ser Leu Val Leu Met His Lys							
		535		540			545	550
20	CCC CTT CGC CAC CCA GGC ATG CCC AAG GGC CTG GCC CAT CAG CTG GCC							1794
	Pro Leu Arg His Pro Gly Met Pro Lys Gly Leu Ala His Gln Leu Ala							
			555			560		565
25	TCT CCT GGC CTC ACG ACC CTC CCT GAG GCC AGC GAT GTG GGC AGC ATC							1842
	Ser Pro Gly Leu Thr Thr Leu Pro Glu Ala Ser Asp Val Gly Ser Ile							
		570				575		580
30	ACT CTT GCC CTC CGA ACG CTT GGC AGC TTT GAA TTT GAA GGC CAC TCT							1890
	Thr Leu Ala Leu Arg Thr Leu Gly Ser Phe Glu Phe Glu Gly His Ser							
		585				590		595
35	CTG ACC CAA TTT GTT CGC CAC TGT GCG GAT CAT TTC CTG AAC AGT GAG							1938
	Leu Thr Gln Phe Val Arg His Cys Ala Asp His Phe Leu Asn Ser Glu							
		600		605			610	
40	CAC AAG GAG ATC CGC ATG GAG GCT GCC CGC ACC TGC TCC CGC CTG CTC							1986
	His Lys Glu Ile Arg Met Glu Ala Ala Arg Thr Cys Ser Arg Leu Leu							
		615		620			625	630
45	ACA CCC TCC ATC CAC CTC ATC AGT GGC CAT GCT CAT GTG GTT AGC CAG							2034
	Thr Pro Ser Ile His Leu Ile Ser Gly His Ala His Val Val Ser Gln							
			635			640		645
50	ACC GCA GTG CAA GTG GTG GCA GAT GTG CTT AGC AAA CTG CTC GTA GTT							2082
	Thr Ala Val Gln Val Val Ala Asp Val Leu Ser Lys Leu Leu Val Val							
		650				655		660
55	GGG ATA ACA GAT CCT GAC CCT GAC ATT CGC TAC TGT GTC TTG GCG TCC							2130
	Gly Ile Thr Asp Pro Asp Pro Asp Ile Arg Tyr Cys Val Leu Ala Ser							
		665				670		675
60	CTG GAC GAG CGC TTT GAT GCA CAC CTG GCC CAG GCG GAG AAC TTG CAG							2178
	Leu Asp Glu Arg Phe Asp Ala His Leu Ala Gln Ala Glu Asn Leu Gln							
		680				685		690
65	GCC TTG TTT GTG GCT CTG AAT GAC CAG GTG TTT GAG ATC CGG GAG CTG							2226
	Ala Leu Phe Val Ala Leu Asn Asp Gln Val Phe Glu Ile Arg Glu Leu							
		695		700			705	710
70	GCC ATC TGC ACT GTG GGC CGA CTC AGT AGC ATG AAC CCT GCC TTT GTC							2274
	Ala Ile Cys Thr Val Gly Arg Leu Ser Ser Met Asn Pro Ala Phe Val							
			715			720		725

	ATG	CCT	TTC	CTG	CGC	AAG	ATG	CTC	ATC	CAG	ATT	TTG	ACA	GAG	TTG	GAG	2322
	Met	Pro	Phe	Leu	Arg	Lys	Met	Leu	Ile	Gln	Ile	Leu	Thr	Glu	Leu	Glu	
				730					735					740			
5	CAC	AGT	GGG	ATT	GGA	AGA	ATC	AAA	GAG	CAG	AGT	GCC	CGC	ATG	CTG	GGG	2370
	His	Ser	Gly	Ile	Gly	Arg	Ile	Lys	Glu	Gln	Ser	Ala	Arg	Met	Leu	Gly	
			745					750					755				
10	CAC	CTG	GTC	TCC	AAT	GCC	CCC	CGA	CTC	ATC	CGC	CCC	TAC	ATG	GAG	CCT	2418
	His	Leu	Val	Ser	Asn	Ala	Pro	Arg	Leu	Ile	Arg	Pro	Tyr	Met	Glu	Pro	
		760					765					770					
15	ATT	CTG	AAG	GCA	TTA	ATT	TTG	AAA	CTG	AAA	GAT	CCA	GAC	CCT	GAT	CCA	2466
	Ile	Leu	Lys	Ala	Leu	Ile	Leu	Lys	Leu	Lys	Asp	Pro	Asp	Pro	Asp	Pro	
	775					780				785					790		
20	AAC	CCA	GGT	GTG	ATC	AAT	AAT	GTC	CTG	GCA	ACA	ATA	GGA	GAA	TTG	GCA	2514
	Asn	Pro	Gly	Val	Ile	Asn	Asn	Val	Leu	Ala	Thr	Ile	Gly	Glu	Leu	Ala	
				795					800					805			
25	CAG	GTT	AGT	GGC	CTG	GAA	ATG	AGG	AAA	TGG	GTT	GAT	GAA	CTT	TTT	ATT	2562
	Gln	Val	Ser	Gly	Leu	Glu	Met	Arg	Lys	Trp	Val	Asp	Glu	Leu	Phe	Ile	
			810						815				820				
30	ATC	ATC	ATG	GAC	ATG	CTC	CAG	GAT	TCC	TCT	TTG	TTG	GCC	AAA	AGG	CAG	2610
	Ile	Ile	Met	Asp	Met	Leu	Gln	Asp	Ser	Ser	Leu	Leu	Ala	Lys	Arg	Gln	
			825					830					835				
35	GTG	GCT	CTG	TGG	ACC	CTG	GGA	CAG	TTG	GTG	GCC	AGC	ACT	GGC	TAT	GTA	2658
	Val	Ala	Leu	Trp	Thr	Leu	Gly	Gln	Leu	Val	Ala	Ser	Thr	Gly	Tyr	Val	
		840					845					850					
40	GTA	GAG	CCC	TAC	AGG	AAG	TAC	CCT	ACT	TTG	CTT	GAG	GTG	CTA	CTG	AAT	2706
	Val	Glu	Pro	Tyr	Arg	Lys	Tyr	Pro	Thr	Leu	Leu	Glu	Val	Leu	Leu	Asn	
	855					860					865				870		
45	TTT	CTG	AAG	ACT	GAG	CAG	AAC	CAG	GGT	ACA	CGC	AGA	GAG	GCC	ATC	CGT	2754
	Phe	Leu	Lys	Thr	Glu	Gln	Asn	Gln	Gly	Thr	Arg	Arg	Glu	Ala	Ile	Arg	
				875					880					885			
50	GTG	TTA	GGG	CTT	TTA	GGG	GCT	TTG	GAT	CCT	TAC	AAG	CAC	AAA	GTG	AAC	2802
	Val	Leu	Gly	Leu	Leu	Gly	Ala	Leu	Asp	Pro	Tyr	Lys	His	Lys	Val	Asn	
			890						895				900				
55	ATT	GGC	ATG	ATA	GAC	CAG	TCC	CGG	GAT	GCC	TCT	GCT	GTC	AGC	CTG	TCA	2850
	Ile	Gly	Met	Ile	Asp	Gln	Ser	Arg	Asp	Ala	Ser	Ala	Val	Ser	Leu	Ser	
			905					910					915				
60	GAA	TCC	AAG	TCA	AGT	CAG	GAT	TCC	TCT	GAC	TAT	AGC	ACT	AGT	GAA	ATG	2898
	Glu	Ser	Lys	Ser	Ser	Gln	Asp	Ser	Ser	Asp	Tyr	Ser	Thr	Ser	Glu	Met	
		920					925					930					
65	CTG	GTC	AAC	ATG	GGA	AAC	TTG	CCT	CTG	GAT	GAG	TTC	TAC	CCA	GCT	GTG	2946
	Leu	Val	Asn	Met	Gly	Asn	Leu	Pro	Leu	Asp	Glu	Phe	Tyr	Pro	Ala	Val	
	935					940					945				950		
70	TCC	ATG	GTG	GCC	CTG	ATG	CGG	ATC	TTC	CGA	GAC	CAG	TCA	CTC	TCT	CAT	2994
	Ser	Met	Val	Ala	Leu	Met	Arg	Ile	Phe	Arg	Asp	Gln	Ser	Leu	Ser	His	

955

960

965

5	CAT CAC ACC ATG GTT GTC CAG GCC ATC ACC TTC ATC TTC AAG TCC CTG His His Thr Met Val Val Gln Ala Ile Thr Phe Ile Phe Lys Ser Leu 970 975 980	3042
10	GGA CTC AAA TGT GTG CAG TTC CTG CCC CAG GTC ATG CCC ACG TTC CTT Gly Leu Lys Cys Val Gln Phe Leu Pro Gln Val Met Pro Thr Phe Leu 985 990 995	3090
15	AAT GTC ATT CGA GTC TGT GAT GGG GCC ATC CGG GAA TTT TTG TTC CAG Asn Val Ile Arg Val Cys Asp Gly Ala Ile Arg Glu Phe Leu Phe Gln 1000 1005 1010	3138
20	CAG CTG GGA ATG TTG GTG TCC TTT GTG AAG AGC CAC ATC AGA CCT TAT Gln Leu Gly Met Leu Val Ser Phe Val Lys Ser His Ile Arg Pro Tyr 1015 1020 1025 1030	3186
25	ATG GAT GAA ATA GTC ACC CTC ATG AGA GAA TTC TGG GTC ATG AAC ACC Met Asp Glu Ile Val Thr Leu Met Arg Glu Phe Trp Val Met Asn Thr 1035 1040 1045	3234
30	TCA ATT CAG AGC ACG ATC ATT CTT CTC ATT GAG CAA ATT GTG GTA GCT Ser Ile Gln Ser Thr Ile Ile Leu Leu Ile Glu Gln Ile Val Val Ala 1050 1055 1060	3282
35	CTT GGG GGT GAA TTT AAG CTC TAC CTG CCC CAG CTG ATC CCA CAC ATG Leu Gly Gly Glu Phe Lys Leu Tyr Leu Pro Gln Leu Ile Pro His Met 1065 1070 1075	3330
40	CTG CGT GTC TTC ATG CAT GAC AAC AGC CCA GGC CGC ATT GTC TCT ATC Leu Arg Val Phe Met His Asp Asn Ser Pro Gly Arg Ile Val Ser Ile 1080 1085 1090	3378
45	AAG TTA CTG GCT GCA ATC CAG CTG TTT GGC GCC AAC CTG GAT GAC TAC Lys Leu Leu Ala Ala Ile Gln Leu Phe Gly Ala Asn Leu Asp Asp Tyr 1095 1100 1105 1110	3426
50	CTG CAT TTA CTG CTG CCT CCT ATT GTT AAG TTG TTT GAT GCC CCT GAA Leu His Leu Leu Leu Pro Pro Ile Val Lys Leu Phe Asp Ala Pro Glu 1115 1120 1125	3474
55	GCT CCA CTG CCA TCT CGA AAG GCA GCG CTA GAG ACT GTG GAC CGC CTG Ala Pro Leu Pro Ser Arg Lys Ala Ala Leu Glu Thr Val Asp Arg Leu 1130 1135 1140	3522
60	ACG GAG TCC CTG GAT TTC ACT GAC TAT GCC TCC CGG ATC ATT CAC CCT Thr Glu Ser Leu Asp Phe Thr Asp Tyr Ala Ser Arg Ile Ile His Pro 1145 1150 1155	3570
65	ATT GTT CGA ACA CTG GAC CAG AGC CCA GAA CTG CGC TCC ACA GCC ATG Ile Val Arg Thr Leu Asp Gln Ser Pro Glu Leu Arg Ser Thr Ala Met 1160 1165 1170	3618
70	GAC ACG CTG TCT TCA CTT GTT TTT CAG CTG GGG AAG AAG TAC CAA ATT Asp Thr Leu Ser Ser Leu Val Phe Gln Leu Gly Lys Lys Tyr Gln Ile 1175 1180 1185 1190	3666

78

	TTC ATT CCA ATG GTG AAT AAA GTT CTG GTG CGA CAC CGA ATC AAT CAT	3714
	Phe Ile Pro Met Val Asn Lys Val Leu Val Arg His Arg Ile Asn His	
	1195 1200 1205	
5	CAG CGC TAT GAT GTG CTC ATC TGC AGA ATT GTC AAG GGA TAC ACA CTT	3762
	Gln Arg Tyr Asp Val Leu Ile Cys Arg Ile Val Lys Gly Tyr Thr Leu	
	1210 1215 1220	
10	GCT GAT GAA GAG GAG GAT CCT TTG ATT TAC CAG CAT CGG ATG CTT AGG	3810
	Ala Asp Glu Glu Glu Asp Pro Leu Ile Tyr Gln His Arg Met Leu Arg	
	1225 1230 1235	
15	AGT GGC CAA GGG GAT GCA TTG GCT AGT GGA CCA GTG GAA ACA GGA CCC	3858
	Ser Gly Gln Gly Asp Ala Leu Ala Ser Gly Pro Val Glu Thr Gly Pro	
	1240 1245 1250	
20	ATG AAG AAA CTG CAC GTC AGC ACC ATC AAC CTC CAA AAG GCC TGG GGC	3906
	Met Lys Lys Leu His Val Ser Thr Ile Asn Leu Gln Lys Ala Trp Gly	
	1255 1260 1265 1270	
25	GCT GCC AGG AGG GTC TCC AAA GAT GAC TGG CTG GAA TGG CTG AGA CGG	3954
	Ala Ala Arg Arg Val Ser Lys Asp Asp Trp Leu Glu Trp Leu Arg Arg	
	1275 1280 1285	
30	CTG AGC CTG GAG CTG CTG AAG GAC TCA TCA TCG CCC TCC CTG CGC TCC	4002
	Leu Ser Leu Glu Leu Leu Lys Asp Ser Ser Ser Pro Ser Leu Arg Ser	
	1290 1295 1300	
35	TGC TGG GCC CTG GCA CAG GCC TAC AAC CCG ATG GCC AGG GAT CTC TTC	4050
	Cys Trp Ala Leu Ala Gln Ala Tyr Asn Pro Met Ala Arg Asp Leu Phe	
	1305 1310 1315	
40	AAT GCT GCA TTT GTG TCC TGC TGG TCT GAA CTG AAT GAA GAT CAA CAG	4098
	Asn Ala Ala Phe Val Ser Cys Trp Ser Glu Leu Asn Glu Asp Gln Gln	
	1320 1325 1330	
45	GAT GAG CTC ATC AGA AGC ATC GAG TTG GCC CTC ACC TCA CAA GAC ATC	4146
	Asp Glu Leu Ile Arg Ser Ile Glu Leu Ala Leu Thr Ser Gln Asp Ile	
	1335 1340 1345 1350	
50	GCT GAA GTC ACA CAG ACC CTC TTA AAC TTG GCT GAA TTC ATG GAA CAC	4194
	Ala Glu Val Thr Gln Thr Leu Leu Asn Leu Ala Glu Phe Met Glu His	
	1355 1360 1365	
55	AGT GAC AAG GGC CCC CTG CCA CTG AGA GAT GAC AAT GGC ATT GTT CTG	4242
	Ser Asp Lys Gly Pro Leu Pro Leu Arg Asp Asp Asn Gly Ile Val Leu	
	1370 1375 1380	
55	CTG GGT GAG AGA GCT GCC AAG TGC CGA GCA TAT GCC AAA GCA CTA CAC	4290
	Leu Gly Glu Arg Ala Ala Lys Cys Arg Ala Tyr Ala Lys Ala Leu His	
	1385 1390 1395	
55	TAC AAA GAA CTG GAG TTC CAG AAA GGC CCC ACC CCT GCC ATT CTA GAA	4338
	Tyr Lys Glu Leu Glu Phe Gln Lys Gly Pro Thr Pro Ala Ile Leu Glu	
	1400 1405 1410	
55	TCT CTC ATC AGC ATT AAT AAT AAG CTA CAG CAG CCG GAG GCA GCG GCC	4386
	Ser Leu Ile Ser Ile Asn Asn Lys Leu Gln Gln Pro Glu Ala Ala Ala	

	1415	1420	1425	1430	
5	GGA GTG TTA GAA TAT GCC ATG AAA CAC TTT GGA GAG CTG GAG ATC CAG Gly Val Leu Glu Tyr Ala Met Lys His Phe Gly Glu Leu Glu Ile Gln	1435	1440	1445	4434
10	GCT ACC TGG TAT GAG AAA CTG CAC GAG TGG GAG GAT GCC CTT GTG GCC Ala Thr Trp Tyr Glu Lys Leu His Glu Trp Glu Asp Ala Leu Val Ala	1450	1455	1460	4482
15	TAT GAC AAG AAA ATG GAC ACC AAC AAG GAC GAC CCA GAG CTG ATG CTG Tyr Asp Lys Lys Met Asp Thr Asn Lys Asp Asp Pro Glu Leu Met Leu	1465	1470	1475	4530
20	GGC CGC ATG CGC TGC CTC GAG GCC TTG GGG GAA TGG GGT CAA CTC CAC Gly Arg Met Arg Cys Leu Glu Ala Leu Gly Glu Trp Gly Gln Leu His	1480	1485	1490	4578
25	CAG CAG TGC TGT GAA AAG TGG ACC CTG GTT AAT GAT GAG ACC CAA GCC Gln Gln Cys Cys Glu Lys Trp Thr Leu Val Asn Asp Glu Thr Gln Ala	1495	1500	1505	4626
30	AAG ATG GCC CGG ATG GCT GCT GCA GCT GCA TGG GGT TTA GGT CAG TGG Lys Met Ala Arg Met Ala Ala Ala Ala Trp Gly Leu Gly Gln Trp	1515	1520	1525	4674
35	GAC AGC ATG GAA GAA TAC ACC TGT ATG ATC CCT CGG GAC ACC CAT GAT Asp Ser Met Glu Glu Tyr Thr Cys Met Ile Pro Arg Asp Thr His Asp	1530	1535	1540	4722
40	GGG GCA TTT TAT AGA GCT GTG CTG GCA CTG CAT CAG GAC CTC TTC TCC Gly Ala Phe Tyr Arg Ala Val Leu Ala Leu His Gln Asp Leu Phe Ser	1545	1550	1555	4770
45	TTG GCA CAA CAG TGC ATT GAC AAG GCC AGG GAC CTG CTG GAT GCT GAA Leu Ala Gln Gln Cys Ile Asp Lys Ala Arg Asp Leu Leu Asp Ala Glu	1560	1565	1570	4818
50	TTA ACT GCA ATG GCA GGA GAG AGT TAC AGT CGG GCA TAT GGG GCC ATG Leu Thr Ala Met Ala Gly Glu Ser Tyr Ser Arg Ala Tyr Gly Ala Met	1575	1580	1585	4866
55	GTT TCT TGC CAC ATG CTG TCC GAG CTG GAG GAG GTT ATC CAG TAC AAA Val Ser Cys His Met Leu Ser Glu Leu Glu Glu Val Ile Gln Tyr Lys	1595	1600	1605	4914
60	CTT GTC CCC GAG CGA CGA GAG ATC ATC CGC CAG ATC TGG TGG GAG AGA Leu Val Pro Glu Arg Arg Glu Ile Ile Arg Gln Ile Trp Trp Glu Arg	1610	1615	1620	4962
65	CTG CAG GGC TGC CAG CGT ATC GTA GAG GAC TGG CAG AAA ATC CTT ATG Leu Gln Gly Cys Gln Arg Ile Val Glu Asp Trp Gln Lys Ile Leu Met	1625	1630	1635	5010
70	GTG CGG TCC CTT GTG GTC AGC CCT CAT GAA GAC ATG AGA ACC TGG CTC Val Arg Ser Leu Val Val Ser Pro His Glu Asp Met Arg Thr Trp Leu	1640	1645	1650	5058

	AAG	TAT	GCA	AGC	CTG	TGC	GGC	AAG	AGT	GGC	AGG	CTG	GCT	CTT	GCT	CAT	5106
	Lys	Tyr	Ala	Ser	Leu	Cys	Gly	Lys	Ser	Gly	Arg	Leu	Ala	Leu	Ala	His	
	1655					1660					1665					1670	
5	AAA	ACT	TTA	GTG	TTG	CTC	CTG	GGA	GTT	GAT	CCG	TCT	CGG	CAA	CTT	GAC	5154
	Lys	Thr	Leu	Val	Leu	Leu	Leu	Gly	Val	Asp	Pro	Ser	Arg	Gln	Leu	Asp	
					1675					1680					1685		
10	CAT	CCT	CTG	CCA	ACA	GTT	CAC	CCT	CAG	GTG	ACC	TAT	GCC	TAC	ATG	AAA	5202
	His	Pro	Leu	Pro	Thr	Val	His	Pro	Gln	Val	Thr	Tyr	Ala	Tyr	Met	Lys	
				1690					1695					1700			
15	AAC	ATG	TGG	AAG	AGT	GCC	CGC	AAG	ATC	GAT	GCC	TTC	CAG	CAC	ATG	CAG	5250
	Asn	Met	Trp	Lys	Ser	Ala	Arg	Lys	Ile	Asp	Ala	Phe	Gln	His	Met	Gln	
			1705					1710					1715				
20	CAT	TTT	GTC	CAG	ACC	ATG	CAG	CAA	CAG	GCC	CAG	CAT	GCC	ATC	GCT	ACT	5298
	His	Phe	Val	Gln	Thr	Met	Gln	Gln	Gln	Ala	Gln	His	Ala	Ile	Ala	Thr	
		1720					1725					1730					
25	GAG	GAC	CAG	CAG	CAT	AAG	CAG	GAA	CTG	CAC	AAG	CTC	ATG	GCC	CGA	TGC	5346
	Glu	Asp	Gln	Gln	His	Lys	Gln	Glu	Leu	His	Lys	Leu	Met	Ala	Arg	Cys	
	1735					1740					1745					1750	
30	TTC	CTG	AAA	CTT	GGA	GAG	TGG	CAG	CTG	AAT	CTA	CAG	GGC	ATC	AAT	GAG	5394
	Phe	Leu	Lys	Leu	Gly	Glu	Trp	Gln	Leu	Asn	Leu	Gln	Gly	Ile	Asn	Glu	
					1755				1760						1765		
35	AGC	ACA	ATC	CCC	AAA	GTG	CTG	CAG	TAC	TAC	AGC	GCC	GCC	ACA	GAG	CAC	5442
	Ser	Thr	Ile	Pro	Lys	Val	Leu	Gln	Tyr	Tyr	Ser	Ala	Ala	Thr	Glu	His	
				1770					1775					1780			
40	GAC	CGC	AGC	TGG	TAC	AAG	GCC	TGG	CAT	GCG	TGG	GCA	GTG	ATG	AAC	TTC	5490
	Asp	Arg	Ser	Trp	Tyr	Lys	Ala	Trp	His	Ala	Trp	Ala	Val	Met	Asn	Phe	
			1785					1790					1795				
45	GAA	GCT	GTG	CTA	CAC	TAC	AAA	CAT	CAG	AAC	CAA	GCC	CGC	GAT	GAG	AAG	5538
	Glu	Ala	Val	Leu	His	Tyr	Lys	His	Gln	Asn	Gln	Ala	Arg	Asp	Glu	Lys	
		1800					1805					1810					
50	AAG	AAA	CTG	CGT	CAT	GCC	AGC	GGG	GCC	AAC	ATC	ACC	AAC	GCC	ACC	ACT	5586
	Lys	Lys	Leu	Arg	His	Ala	Ser	Gly	Ala	Asn	Ile	Thr	Asn	Ala	Thr	Thr	
	1815					1820					1825					1830	
55	GCC	GCC	ACC	ACG	GCC	GCC	ACT	GCC	ACC	ACC	ACT	GCC	AGC	ACC	GAG	GGC	5634
	Ala	Ala	Thr	Thr	Ala	Ala	Thr	Ala	Thr	Thr	Thr	Ala	Ser	Thr	Glu	Gly	
					1835					1840					1845		
60	AGC	AAC	AGT	GAG	AGC	GAG	GCC	GAG	AGC	ACC	GAG	AAC	AGC	CCC	ACC	CCA	5682
	Ser	Asn	Ser	Glu	Ser	Glu	Ala	Glu	Ser								

	1880	1885	1890	
5	TCA CGA GGC AAC AAC CTC CAG GAT ACA CTC AGA GTT CTC ACC TTA TGG Ser Arg Gly Asn Asn Leu Gln Asp Thr Leu Arg Val Leu Thr Leu Trp 1895 1900 1905 1910			5826
10	TTT GAT TAT GGT CAC TGG CCA GAT GTC AAT GAG GCC TTA GTG GAG GGG Phe Asp Tyr Gly His Trp Pro Asp Val Asn Glu Ala Leu Val Glu Gly 1915 1920 1925			5874
	GTG AAA GCC ATC CAG ATT GAT ACC TGG CTA CAG GTT ATA CCT CAG CTC Val Lys Ala Ile Gln Ile Asp Thr Trp Leu Gln Val Ile Pro Gln Leu 1930 1935 1940			5922
15	ATT GCA AGA ATT GAT ACG CCC AGA CCC TTG GTG GGA CGT CTC ATT CAC Ile Ala Arg Ile Asp Thr Pro Arg Pro Leu Val Gly Arg Leu Ile His 1945 1950 1955			5970
20	CAG CTT CTC ACA GAC ATT GGT CGG TAC CAC CCC CAG GCC CTC ATC TAC Gln Leu Leu Thr Asp Ile Gly Arg Tyr His Pro Gln Ala Leu Ile Tyr 1960 1965 1970			6018
25	CCA CTG ACA GTG GCT TCT AAG TCT ACC ACG ACA GCC CGG CAC AAT GCA Pro Leu Thr Val Ala Ser Lys Ser Thr Thr Thr Ala Arg His Asn Ala 1975 1980 1985 1990			6066
30	GCC AAC AAG ATT CTG AAG AAC ATG TGT GAG CAC AGC AAC ACC CTG GTC Ala Asn Lys Ile Leu Lys Asn Met Cys Glu His Ser Asn Thr Leu Val 1995 2000 2005			6114
	CAG CAG GCC ATG ATG GTG AGC GAG GAG CTG ATC CGA GTG GCC ATC CTC Gln Gln Ala Met Met Val Ser Glu Glu Leu Ile Arg Val Ala Ile Leu 2010 2015 2020			6162
35	TGG CAT GAG ATG TGG CAT GAA GGC CTG GAA GAG GCA TCT CGT TTG TAC Trp His Glu Met Trp His Glu Gly Leu Glu Glu Ala Ser Arg Leu Tyr 2025 2030 2035			6210
40	TTT GGG GAA AGG AAC GTG AAA GGC ATG TTT GAG GTG CTG GAG CCC TTG Phe Gly Glu Arg Asn Val Lys Gly Met Phe Glu Val Leu Glu Pro Leu 2040 2045 2050			6258
45	CAT GCT ATG ATG GAA CGG GGC CCC CAG ACT CTG AAG GAA ACA TCC TTT His Ala Met Met Glu Arg Gly Pro Gln Thr Leu Lys Glu Thr Ser Phe 2055 2060 2065 2070			6306
50	AAT CAG GCC TAT GGT CGA GAT TTA ATG GAG GCC CAA GAG TGG TGC AGG Asn Gln Ala Tyr Gly Arg Asp Leu Met Glu Ala Gln Glu Trp Cys Arg 2075 2080 2085			6354
	AAG TAC ATG AAA TCA GGG AAT GTC AAG GAC CTC ACC CAA GCC TGG GAC Lys Tyr Met Lys Ser Gly Asn Val Lys Asp Leu Thr Gln Ala Trp Asp 2090 2095 2100			6402
55	CTC TAT TAT CAT GTG TTC CGA CGA ATC TCA AAG CAG CTG CCT CAG CTC Leu Tyr Tyr His Val Phe Arg Arg Ile Ser Lys Gln Leu Pro Gln Leu 2105 2110 2115			6450

	ACA	TCC	TTA	GAG	CTG	CAA	TAT	GTT	TCC	CCA	AAA	CTT	CTG	ATG	TGC	CGG	6498
	Thr	Ser	Leu	Glu	Leu	Gln	Tyr	Val	Ser	Pro	Lys	Leu	Leu	Met	Cys	Arg	
	2120					2125					2130						
5	GAC	CTT	GAA	TTG	GCT	GTG	CCA	GGA	ACA	TAT	GAC	CCC	AAC	CAG	CCA	ATC	6546
	Asp	Leu	Glu	Leu	Ala	Val	Pro	Gly	Thr	Tyr	Asp	Pro	Asn	Gln	Pro	Ile	
	2135					2140					2145					2150	
10	ATT	CGC	ATT	CAG	TCC	ATA	GCA	CCG	TCT	TTG	CAA	GTC	ATC	ACA	TCC	AAG	6594
	Ile	Arg	Ile	Gln	Ser	Ile	Ala	Pro	Ser	Leu	Gln	Val	Ile	Thr	Ser	Lys	
				2155					2160						2165		
15	CAG	AGG	CCC	CGG	AAA	TTG	ACA	CTT	ATG	GGC	AGC	AAC	GGA	CAT	GAG	TTT	6642
	Gln	Arg	Pro	Arg	Lys	Leu	Thr	Leu	Met	Gly	Ser	Asn	Gly	His	Glu	Phe	
				2170					2175					2180			
20	GTT	TTC	CTT	CTA	AAA	GGC	CAT	GAA	GAT	CTG	CGC	CAG	GAT	GAG	CGT	GTG	6690
	Val	Phe	Leu	Leu	Lys	Gly	His	Glu	Asp	Leu	Arg	Gln	Asp	Glu	Arg	Val	
		2185					2190					2195					
	ATG	CAG	CTC	TTC	GGC	CTG	GTT	AAC	ACC	CTT	CTG	GCC	AAT	GAC	CCA	ACA	6738
	Met	Gln	Leu	Phe	Gly	Leu	Val	Asn	Thr	Leu	Leu	Ala	Asn	Asp	Pro	Thr	
	2200					2205					2210						
25	TCT	CTT	CGG	AAA	AAC	CTC	AGC	ATC	CAG	AGA	TAC	GCT	GTC	ATC	CCT	TTA	6786
	Ser	Leu	Arg	Lys	Asn	Leu	Ser	Ile	Gln	Arg	Tyr	Ala	Val	Ile	Pro	Leu	
	2215					2220					2225					2230	
30	TCG	ACC	AAC	TCG	GGC	CTC	ATT	GGC	TGG	GTT	CCC	CAC	TGT	GAC	ACA	CTG	6834
	Ser	Thr	Asn	Ser	Gly	Leu	Ile	Gly	Trp	Val	Pro	His	Cys	Asp	Thr	Leu	
				2235					2240					2245			
35	CAC	GCC	CTC	ATC	CGG	GAC	TAC	AGG	GAG	AAG	AAG	AAG	ATC	CTT	CTC	AAC	6882
	His	Ala	Leu	Ile	Arg	Asp	Tyr	Arg	Glu	Lys	Lys	Lys	Ile	Leu	Leu	Asn	
				2250					2255					2260			
40	ATC	GAG	CAT	CGC	ATC	ATG	TTG	CGG	ATG	GCT	CCG	GAC	TAT	GAC	CAC	TTG	6930
	Ile	Glu	His	Arg	Ile	Met	Leu	Arg	Met	Ala	Pro	Asp	Tyr	Asp	His	Leu	
		2265					2270					2275					
	ACT	CTG	ATG	CAG	AAG	GTG	GAG	GTG	TTT	GAG	CAT	GCC	GTC	AAT	AAT	ACA	6978
	Thr	Leu	Met	Gln	Lys	Val	Glu	Val	Phe	Glu	His	Ala	Val	Asn	Asn	Thr	
	2280					2285					2290						
45	GCT	GGG	GAC	GAC	CTG	GCC	AAG	CTG	CTG	TGG	CTG	AAA	AGC	CCC	AGC	TCC	7026
	Ala	Gly	Asp	Asp	Leu	Ala	Lys	Leu	Leu	Trp	Leu	Lys	Ser	Pro	Ser	Ser	
	2295					2300					2305					2310	
50	GAG	GTG	TGG	TTT	GAC	CGA	AGA	ACC	AAT	TAT	ACC	CGT	TCT	TTA	GCG	GTC	7074
	Glu	Val	Trp	Phe	Asp	Arg	Arg	Thr	Asn	Tyr	Thr	Arg	Ser	Leu	Ala	Val	
				2315					2320						2325		
55	ATG	TCA	ATG	GTT	GGG	TAT	ATT	TTA	GGC	CTG	GGA	GAT	AGA	CAC	CCA	TCC	7122
	Met	Ser	Met	Val	Gly	Tyr	Ile	Leu	Gly	Leu	Gly	Asp	Arg	His	Pro	Ser	
				2330					2335					2340			
	AAC	CTG	ATG	CTG	GAC	CGT	CTG	AGT	GGG	AAG	ATC	CTG	CAC	ATT	GAC	TTT	7170
	Asn	Leu	Met	Leu	Asp	Arg	Leu	Ser	Gly	Lys	Ile	Leu	His	Ile	Asp	Phe	

	2345	2350	2355	
5	GGG GAC TGC TTT GAG GTT GCT ATG ACC CGA GAG AAG TTT CCA GAG AAG Gly Asp Cys Phe Glu Val Ala Met Thr Arg Glu Lys Phe Pro Glu Lys 2360 2365 2370	7218		
10	ATT CCA TTT AGA CTA ACA AGA ATG TTG ACC AAT GCT ATG GAG GTT ACA Ile Pro Phe Arg Leu Thr Arg Met Leu Thr Asn Ala Met Glu Val Thr 2375 2380 2385 2390	7266		
	GGC CTG GAT GGC AAC TAC AGA ATC ACA TGC CAC ACA GTG ATG GAG GTG Gly Leu Asp Gly Asn Tyr Arg Ile Thr Cys His Thr Val Met Glu Val 2395 2400 2405	7314		
15	CTG CGA GAG CAC AAG GAC AGT GTC ATG GCC GTG CTG GAA GCC TTT GTC Leu Arg Glu His Lys Asp Ser Val Met Ala Val Leu Glu Ala Phe Val 2410 2415 2420	7362		
20	TAT GAC CCC TTG CTG AAC TGG AGG CTG ATG GAC ACA AAT ACC AAA GGC Tyr Asp Pro Leu Leu Asn Trp Arg Leu Met Asp Thr Asn Thr Lys Gly 2425 2430 2435	7410		
25	AAC AAG CGA TCC CGA ACG AGG ACG GAT TCC TAC TCT GCT GGC CAG TCA Asn Lys Arg Ser Arg Thr Arg Thr Asp Ser Tyr Ser Ala Gly Gln Ser 2440 2445 2450	7458		
30	GTC GAA ATT TTG GAC GGT GTG GAA CTT GGA GAG CCA GCC CAT AAG AAA Val Glu Ile Leu Asp Gly Val Glu Leu Gly Glu Pro Ala His Lys Lys 2455 2460 2465 2470	7506		
	ACG GGG ACC ACA GTG CCA GAA TCT ATT CAT TCT TTC ATT GGA GAC GGT Thr Gly Thr Thr Val Pro Glu Ser Ile His Ser Phe Ile Gly Asp Gly 2475 2480 2485	7554		
35	TTG GTG AAA CCA GAG GCC CTA AAT AAG AAA GCT ATC CAG ATT ATT AAC Leu Val Lys Pro Glu Ala Leu Asn Lys Lys Ala Ile Gln Ile Ile Asn 2490 2495 2500	7602		
40	AGG GTT CGA GAT AAG CTC ACT GGT CGG GAC TTC TCT CAT GAT GAC ACT Arg Val Arg Asp Lys Leu Thr Gly Arg Asp Phe Ser His Asp Asp Thr 2505 2510 2515	7650		
45	TTG GAT GTT CCA ACG CAA GTT GAG CTG CTC ATC AAA CAA GCG ACA TCC Leu Asp Val Pro Thr Gln Val Glu Leu Leu Ile Lys Gln Ala Thr Ser 2520 2525 2530	7698		
50	CAT GAA AAC CTC TGC CAG TGC TAT ATT GGC TGG TGC CCT TTC TGG His Glu Asn Leu Cys Gln Cys Tyr Ile Gly Trp Cys Pro Phe Trp 2535 2540 2545	7743		
	TAAGTGGAGG CCCAGATGTG CCCATCACGT TTTTCTGAG GCTTTTGTAC TTTAGTAAAT	7803		
55	GCTTCCACTA AACTGAAAAA A	7824		

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

84

(A) LENGTH: 2549 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

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

85

Arg Ile Ser Ser Met Glu Gly Glu Arg Leu Arg Glu Glu Met Glu Glu
275 280 285

5 Ile Thr Gln Gln Gln Leu Val His Asp Lys Tyr Cys Lys Asp Leu Met
290 295 300

Gly Phe Gly Thr Lys Pro Arg His Ile Thr Pro Phe Thr Ser Phe Gln
305 310 315 320

10 Ala Val Gln Pro Gln Gln Ser Asn Ala Leu Val Gly Leu Leu Gly Tyr
325 330 335

Ser Ser His Gln Gly Leu Met Gly Phe Gly Thr Ser Pro Ser Pro Ala
15 340 345 350

Lys Ser Thr Leu Val Glu Ser Arg Cys Cys Arg Asp Leu Met Glu Glu
355 360 365

20 Lys Phe Asp Gln Val Cys Gln Trp Val Leu Lys Cys Arg Asn Ser Lys
370 375 380

Asn Ser Leu Ile Gln Met Thr Ile Leu Asn Leu Leu Pro Arg Leu Ala
25 385 390 395 400

Ala Phe Arg Pro Ser Ala Phe Thr Asp Thr Gln Tyr Leu Gln Asp Thr
405 410 415

Met Asn His Val Leu Ser Cys Val Lys Lys Glu Lys Glu Arg Thr Ala
30 420 425 430

Ala Phe Gln Ala Leu Gly Leu Leu Ser Val Ala Val Arg Ser Glu Phe
435 440 445

35 Lys Val Tyr Leu Pro Arg Val Leu Asp Ile Ile Arg Ala Ala Leu Pro
450 455 460

Pro Lys Asp Phe Ala His Lys Arg Gln Lys Ala Met Gln Val Asp Ala
40 465 470 475 480

Thr Val Phe Thr Cys Ile Ser Met Leu Ala Arg Ala Met Gly Pro Gly
485 490 495

Ile Gln Gln Asp Ile Lys Glu Leu Leu Glu Pro Met Leu Ala Val Gly
45 500 505 510

Leu Ser Pro Ala Leu Thr Ala Val Leu Tyr Asp Leu Ser Arg Gln Ile
515 520 525

50 Pro Gln Leu Lys Lys Asp Ile Gln Asp Gly Leu Leu Lys Met Leu Ser
530 535 540

Leu Val Leu Met His Lys Pro Leu Arg His Pro Gly Met Pro Lys Gly
545 550 555 560

55 Leu Ala His Gln Leu Ala Ser Pro Gly Leu Thr Thr Leu Pro Glu Ala
565 570 575

86

	Ser	Asp	Val	Gly	Ser	Ile	Thr	Leu	Ala	Leu	Arg	Thr	Leu	Gly	Ser	Phe	
				580					585					590			
5	Glu	Phe	Glu	Gly	His	Ser	Leu	Thr	Gln	Phe	Val	Arg	His	Cys	Ala	Asp	
			595					600					605				
	His	Phe	Leu	Asn	Ser	Glu	His	Lys	Glu	Ile	Arg	Met	Glu	Ala	Ala	Arg	
		610					615					620					
10	Thr	Cys	Ser	Arg	Leu	Leu	Thr	Pro	Ser	Ile	His	Leu	Ile	Ser	Gly	His	
	625					630					635					640	
	Ala	His	Val	Val	Ser	Gln	Thr	Ala	Val	Gln	Val	Val	Ala	Asp	Val	Leu	
					645					650					655		
15	Ser	Lys	Leu	Leu	Val	Val	Gly	Ile	Thr	Asp	Pro	Asp	Pro	Asp	Ile	Arg	
				660					665					670			
	Tyr	Cys	Val	Leu	Ala	Ser	Leu	Asp	Glu	Arg	Phe	Asp	Ala	His	Leu	Ala	
20			675					680					685				
	Gln	Ala	Glu	Asn	Leu	Gln	Ala	Leu	Phe	Val	Ala	Leu	Asn	Asp	Gln	Val	
		690					695					700					
25	Phe	Glu	Ile	Arg	Glu	Leu	Ala	Ile	Cys	Thr	Val	Gly	Arg	Leu	Ser	Ser	
	705					710					715					720	
	Met	Asn	Pro	Ala	Phe	Val	Met	Pro	Phe	Leu	Arg	Lys	Met	Leu	Ile	Gln	
					725					730					735		
30	Ile	Leu	Thr	Glu	Leu	Glu	His	Ser	Gly	Ile	Gly	Arg	Ile	Lys	Glu	Gln	
				740					745					750			
	Ser	Ala	Arg	Met	Leu	Gly	His	Leu	Val	Ser	Asn	Ala	Pro	Arg	Leu	Ile	
35			755					760					765				
	Arg	Pro	Tyr	Met	Glu	Pro	Ile	Leu	Lys	Ala	Leu	Ile	Leu	Lys	Leu	Lys	
		770					775					780					
40	Asp	Pro	Asp	Pro	Asp	Pro	Asn	Pro	Gly	Val	Ile	Asn	Asn	Val	Leu	Ala	
	785					790					795					800	
	Thr	Ile	Gly	Glu	Leu	Ala	Gln	Val	Ser	Gly	Leu	Glu	Met	Arg	Lys	Trp	
				805						810					815		
45	Val	Asp	Glu	Leu	Phe	Ile	Ile	Ile	Met	Asp	Met	Leu	Gln	Asp	Ser	Ser	
				820					825					830			
	Leu	Leu	Ala	Lys	Arg	Gln	Val	Ala	Leu	Trp	Thr	Leu	Gly	Gln	Leu	Val	
50			835					840					845				
	Ala	Ser	Thr	Gly	Tyr	Val	Val	Glu	Pro	Tyr	Arg	Lys	Tyr	Pro	Thr	Leu	
		850					855					860					
55	Leu	Glu	Val	Leu	Leu	Asn	Phe	Leu	Lys	Thr	Glu	Gln	Asn	Gln	Gly	Thr	
	865					870					875					880	
	Arg	Arg	Glu	Ala	Ile	Arg	Val	Leu	Gly	Leu	Leu	Gly	Ala	Leu	Asp	Pro	

885 890 895

5 Tyr Lys His Lys Val Asn Ile Gly Met Ile Asp Gln Ser Arg Asp Ala
900 905 910

Ser Ala Val Ser Leu Ser Glu Ser Lys Ser Ser Gln Asp Ser Ser Asp
915 920 925

10 Tyr Ser Thr Ser Glu Met Leu Val Asn Met Gly Asn Leu Pro Leu Asp
930 935 940

Glu Phe Tyr Pro Ala Val Ser Met Val Ala Leu Met Arg Ile Phe Arg
945 950 955 960

15 Asp Gln Ser Leu Ser His His His Thr Met Val Val Gln Ala Ile Thr
965 970 975

Phe Ile Phe Lys Ser Leu Gly Leu Lys Cys Val Gln Phe Leu Pro Gln
980 985 990

20 Val Met Pro Thr Phe Leu Asn Val Ile Arg Val Cys Asp Gly Ala Ile
995 1000 1005

25 Arg Glu Phe Leu Phe Gln Gln Leu Gly Met Leu Val Ser Phe Val Lys
1010 1015 1020

Ser His Ile Arg Pro Tyr Met Asp Glu Ile Val Thr Leu Met Arg Glu
1025 1030 1035 1040

30 Phe Trp Val Met Asn Thr Ser Ile Gln Ser Thr Ile Ile Leu Leu Ile
1045 1050 1055

Glu Gln Ile Val Val Ala Leu Gly Gly Glu Phe Lys Leu Tyr Leu Pro
1060 1065 1070

35 Gln Leu Ile Pro His Met Leu Arg Val Phe Met His Asp Asn Ser Pro
1075 1080 1085

40 Gly Arg Ile Val Ser Ile Lys Leu Leu Ala Ala Ile Gln Leu Phe Gly
1090 1095 1100

Ala Asn Leu Asp Asp Tyr Leu His Leu Leu Leu Pro Pro Ile Val Lys
1105 1110 1115 1120

45 Leu Phe Asp Ala Pro Glu Ala Pro Leu Pro Ser Arg Lys Ala Ala Leu
1125 1130 1135

Glu Thr Val Asp Arg Leu Thr Glu Ser Leu Asp Phe Thr Asp Tyr Ala
1140 1145 1150

50 Ser Arg Ile Ile His Pro Ile Val Arg Thr Leu Asp Gln Ser Pro Glu
1155 1160 1165

55 Leu Arg Ser Thr Ala Met Asp Thr Leu Ser Ser Leu Val Phe Gln Leu
1170 1175 1180

Gly Lys Lys Tyr Gln Ile Phe Ile Pro Met Val Asn Lys Val Leu Val
1185 1190 1195 1200

Arg His Arg Ile Asn His Gln Arg Tyr Asp Val Leu Ile Cys Arg Ile
1205 1210 1215

5 Val Lys Gly Tyr Thr Leu Ala Asp Glu Glu Glu Asp Pro Leu Ile Tyr
1220 1225 1230

Gln His Arg Met Leu Arg Ser Gly Gln Gly Asp Ala Leu Ala Ser Gly
1235 1240 1245

10 Pro Val Glu Thr Gly Pro Met Lys Lys Leu His Val Ser Thr Ile Asn
1250 1255 1260

Leu Gln Lys Ala Trp Gly Ala Ala Arg Arg Val Ser Lys Asp Asp Trp
15 1265 1270 1275 1280

Leu Glu Trp Leu Arg Arg Leu Ser Leu Glu Leu Leu Lys Asp Ser Ser
1285 1290 1295

20 Ser Pro Ser Leu Arg Ser Cys Trp Ala Leu Ala Gln Ala Tyr Asn Pro
1300 1305 1310

Met Ala Arg Asp Leu Phe Asn Ala Ala Phe Val Ser Cys Trp Ser Glu
1315 1320 1325

25 Leu Asn Glu Asp Gln Gln Asp Glu Leu Ile Arg Ser Ile Glu Leu Ala
1330 1335 1340

Leu Thr Ser Gln Asp Ile Ala Glu Val Thr Gln Thr Leu Leu Asn Leu
30 1345 1350 1355 1360

Ala Glu Phe Met Glu His Ser Asp Lys Gly Pro Leu Pro Leu Arg Asp
1365 1370 1375

35 Asp Asn Gly Ile Val Leu Leu Gly Glu Arg Ala Ala Lys Cys Arg Ala
1380 1385 1390

Tyr Ala Lys Ala Leu His Tyr Lys Glu Leu Glu Phe Gln Lys Gly Pro
1395 1400 1405

40 Thr Pro Ala Ile Leu Glu Ser Leu Ile Ser Ile Asn Asn Lys Leu Gln
1410 1415 1420

Gln Pro Glu Ala Ala Ala Gly Val Leu Glu Tyr Ala Met Lys His Phe
45 1425 1430 1435 1440

Gly Glu Leu Glu Ile Gln Ala Thr Trp Tyr Glu Lys Leu His Glu Trp
1445 1450 1455

50 Glu Asp Ala Leu Val Ala Tyr Asp Lys Lys Met Asp Thr Asn Lys Asp
1460 1465 1470

Asp Pro Glu Leu Met Leu Gly Arg Met Arg Cys Leu Glu Ala Leu Gly
1475 1480 1485

55 Glu Trp Gly Gln Leu His Gln Gln Cys Cys Glu Lys Trp Thr Leu Val
1490 1495 1500

Asn Asp Glu Thr Gln Ala Lys Met Ala Arg Met Ala Ala Ala Ala Ala
1505 1510 1515 1520

5 Trp Gly Leu Gly Gln Trp Asp Ser Met Glu Glu Tyr Thr Cys Met Ile
1525 1530 1535

Pro Arg Asp Thr His Asp Gly Ala Phe Tyr Arg Ala Val Leu Ala Leu
1540 1545 1550

10 His Gln Asp Leu Phe Ser Leu Ala Gln Gln Cys Ile Asp Lys Ala Arg
1555 1560 1565

Asp Leu Leu Asp Ala Glu Leu Thr Ala Met Ala Gly Glu Ser Tyr Ser
1570 1575 1580

15 Arg Ala Tyr Gly Ala Met Val Ser Cys His Met Leu Ser Glu Leu Glu
1585 1590 1595 1600

20 Glu Val Ile Gln Tyr Lys Leu Val Pro Glu Arg Arg Glu Ile Ile Arg
1605 1610 1615

Gln Ile Trp Trp Glu Arg Leu Gln Gly Cys Gln Arg Ile Val Glu Asp
1620 1625 1630

25 Trp Gln Lys Ile Leu Met Val Arg Ser Leu Val Val Ser Pro His Glu
1635 1640 1645

Asp Met Arg Thr Trp Leu Lys Tyr Ala Ser Leu Cys Gly Lys Ser Gly
1650 1655 1660

30 Arg Leu Ala Leu Ala His Lys Thr Leu Val Leu Leu Leu Gly Val Asp
1665 1670 1675 1680

Pro Ser Arg Gln Leu Asp His Pro Leu Pro Thr Val His Pro Gln Val
1685 1690 1695

Thr Tyr Ala Tyr Met Lys Asn Met Trp Lys Ser Ala Arg Lys Ile Asp
1700 1705 1710

40 Ala Phe Gln His Met Gln His Phe Val Gln Thr Met Gln Gln Gln Ala
1715 1720 1725

Gln His Ala Ile Ala Thr Glu Asp Gln Gln His Lys Gln Glu Leu His
1730 1735 1740

45 Lys Leu Met Ala Arg Cys Phe Leu Lys Leu Gly Glu Trp Gln Leu Asn
1745 1750 1755 1760

Leu Gln Gly Ile Asn Glu Ser Thr Ile Pro Lys Val Leu Gln Tyr Tyr
1765 1770 1775

Ser Ala Ala Thr Glu His Asp Arg Ser Trp Tyr Lys Ala Trp His Ala
1780 1785 1790

55 Trp Ala Val Met Asn Phe Glu Ala Val Leu His Tyr Lys His Gln Asn
1795 1800 1805

Gln Ala Arg Asp Glu Lys Lys Lys Leu Arg His Ala Ser Gly Ala Asn

90

		1810				1815						1820					
		Ile	Thr	Asn	Ala	Thr	Thr	Ala	Ala	Thr	Thr	Ala	Ala	Thr	Ala	Thr	Thr
		1825						1830				1835					1840
5		Thr	Ala	Ser	Thr	Glu	Gly	Ser	Asn	Ser	Glu	Ser	Glu	Ala	Glu	Ser	Thr
						1845					1850					1855	
		Glu	Asn	Ser	Pro	Thr	Pro	Ser	Pro	Leu	Gln	Lys	Lys	Val	Thr	Glu	Asp
10					1860					1865					1870		
		Leu	Ser	Lys	Thr	Leu	Leu	Met	Tyr	Thr	Val	Pro	Ala	Val	Gln	Gly	Phe
				1875				1880						1885			
15		Phe	Arg	Ser	Ile	Ser	Leu	Ser	Arg	Gly	Asn	Asn	Leu	Gln	Asp	Thr	Leu
			1890					1895					1900				
		Arg	Val	Leu	Thr	Leu	Trp	Phe	Asp	Tyr	Gly	His	Trp	Pro	Asp	Val	Asn
20			1905				1910					1915					1920
		Glu	Ala	Leu	Val	Glu	Gly	Val	Lys	Ala	Ile	Gln	Ile	Asp	Thr	Trp	Leu
						1925					1930					1935	
		Gln	Val	Ile	Pro	Gln	Leu	Ile	Ala	Arg	Ile	Asp	Thr	Pro	Arg	Pro	Leu
25					1940					1945					1950		
		Val	Gly	Arg	Leu	Ile	His	Gln	Leu	Leu	Thr	Asp	Ile	Gly	Arg	Tyr	His
			1955					1960					1965				
30		Pro	Gln	Ala	Leu	Ile	Tyr	Pro	Leu	Thr	Val	Ala	Ser	Lys	Ser	Thr	Thr
			1970					1975					1980				
		Thr	Ala	Arg	His	Asn	Ala	Ala	Asn	Lys	Ile	Leu	Lys	Asn	Met	Cys	Glu
35			1985				1990					1995					2000
		His	Ser	Asn	Thr	Leu	Val	Gln	Gln	Ala	Met	Met	Val	Ser	Glu	Glu	Leu
						2005					2010					2015	
		Ile	Arg	Val	Ala	Ile	Leu	Trp	His	Glu	Met	Trp	His	Glu	Gly	Leu	Glu
40					2020					2025					2030		
		Glu	Ala	Ser	Arg	Leu	Tyr	Phe	Gly	Glu	Arg	Asn	Val	Lys	Gly	Met	Phe
				2035					2040					2045			
45		Glu	Val	Leu	Glu	Pro	Leu	His	Ala	Met	Met	Glu	Arg	Gly	Pro	Gln	Thr
			2050					2055					2060				
		Leu	Lys	Glu	Thr	Ser	Phe	Asn	Gln	Ala	Tyr	Gly	Arg	Asp	Leu	Met	Glu
50			2065				2070					2075					2080
		Ala	Gln	Glu	Trp	Cys	Arg	Lys	Tyr	Met	Lys	Ser	Gly	Asn	Val	Lys	Asp
						2085					2090					2095	
		Leu	Thr	Gln	Ala	Trp	Asp	Leu	Tyr	Tyr	His	Val	Phe	Arg	Arg	Ile	Ser
55					2100					2105					2110		
		Lys	Gln	Leu	Pro	Gln	Leu	Thr	Ser	Leu	Glu	Leu	Gln	Tyr	Val	Ser	Pro
				2115					2120					2125			

91

Lys Leu Leu Met Cys Arg Asp Leu Glu Leu Ala Val Pro Gly Thr Tyr
2130 2135 2140

5 Asp Pro Asn Gln Pro Ile Ile Arg Ile Gln Ser Ile Ala Pro Ser Leu
2145 2150 2155 2160

Gln Val Ile Thr Ser Lys Gln Arg Pro Arg Lys Leu Thr Leu Met Gly
2165 2170 2175

10 Ser Asn Gly His Glu Phe Val Phe Leu Leu Lys Gly His Glu Asp Leu
2180 2185 2190

Arg Gln Asp Glu Arg Val Met Gln Leu Phe Gly Leu Val Asn Thr Leu
15 2195 2200 2205

Leu Ala Asn Asp Pro Thr Ser Leu Arg Lys Asn Leu Ser Ile Gln Arg
2210 2215 2220

20 Tyr Ala Val Ile Pro Leu Ser Thr Asn Ser Gly Leu Ile Gly Trp Val
2225 2230 2235 2240

Pro His Cys Asp Thr Leu His Ala Leu Ile Arg Asp Tyr Arg Glu Lys
2245 2250 2255

25 Lys Lys Ile Leu Leu Asn Ile Glu His Arg Ile Met Leu Arg Met Ala
2260 2265 2270

Pro Asp Tyr Asp His Leu Thr Leu Met Gln Lys Val Glu Val Phe Glu
30 2275 2280 2285

His Ala Val Asn Asn Thr Ala Gly Asp Asp Leu Ala Lys Leu Leu Trp
2290 2295 2300

35 Leu Lys Ser Pro Ser Ser Glu Val Trp Phe Asp Arg Arg Thr Asn Tyr
2305 2310 2315 2320

Thr Arg Ser Leu Ala Val Met Ser Met Val Gly Tyr Ile Leu Gly Leu
2325 2330 2335

40 Gly Asp Arg His Pro Ser Asn Leu Met Leu Asp Arg Leu Ser Gly Lys
2340 2345 2350

Ile Leu His Ile Asp Phe Gly Asp Cys Phe Glu Val Ala Met Thr Arg
45 2355 2360 2365

Glu Lys Phe Pro Glu Lys Ile Pro Phe Arg Leu Thr Arg Met Leu Thr
2370 2375 2380

50 Asn Ala Met Glu Val Thr Gly Leu Asp Gly Asn Tyr Arg Ile Thr Cys
2385 2390 2395 2400

His Thr Val Met Glu Val Leu Arg Glu His Lys Asp Ser Val Met Ala
2405 2410 2415

55 Val Leu Glu Ala Phe Val Tyr Asp Pro Leu Leu Asn Trp Arg Leu Met
2420 2425 2430

92

Asp Thr Asn Thr Lys Gly Asn Lys Arg Ser Arg Thr Arg Thr Asp Ser
2435 2440 2445

Tyr Ser Ala Gly Gln Ser Val Glu Ile Leu Asp Gly Val Glu Leu Gly
2450 2455 2460

Glu Pro Ala His Lys Lys Thr Gly Thr Thr Val Pro Glu Ser Ile His
2465 2470 2475 2480

Ser Phe Ile Gly Asp Gly Leu Val Lys Pro Glu Ala Leu Asn Lys Lys
2485 2490 2495

Ala Ile Gln Ile Ile Asn Arg Val Arg Asp Lys Leu Thr Gly Arg Asp
2500 2505 2510

Phe Ser His Asp Asp Thr Leu Asp Val Pro Thr Gln Val Glu Leu Leu
2515 2520 2525

Ile Lys Gln Ala Thr Ser His Glu Asn Leu Cys Gln Cys Tyr Ile Gly
2530 2535 2540

Trp Cys Pro Phe Trp
2545

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1794 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1686

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTG GTT TAC CCT TTG ACA GTT GCT ATT ACT TCC GAA TCA ACG AGC CGT 48
Leu Val Tyr Pro Leu Thr Val Ala Ile Thr Ser Glu Ser Thr Ser Arg
1 5 10 15

AAA AAG GCA GCT CAA TCC ATT ATT GAA AAA ATG CGA GTA CAT TCT CCT 96
Lys Lys Ala Ala Gln Ser Ile Ile Glu Lys Met Arg Val His Ser Pro
20 25 30

AGC TTG GTG GAT CAA GCA GAA TTA GTG AGT CGA GAA CTC ATC CGA GTT 144
Ser Leu Val Asp Gln Ala Glu Leu Val Ser Arg Glu Leu Ile Arg Val
35 40 45

GCA GTT TTA TGG CAC GAA CAA TGG CAC GAT GCT TTG GAA GAT GCT AGC 192
Ala Val Leu Trp His Glu Gln Trp His Asp Ala Leu Glu Asp Ala Ser
50 55 60

93

5	AGG	TTT	TTC	TTT	GGT	GAA	CAC	AAC	ACA	GAA	AAG	ATG	TTT	GAA	ACA	TTG	240
	Arg	Phe	Phe	Phe	Gly	Glu	His	Asn	Thr	Glu	Lys	Met	Phe	Glu	Thr	Leu	
	65					70					75					80	
10	GAA	CCA	TTA	CAT	CAA	ATG	TTG	CAA	AAG	GGA	CCA	GAA	ACG	ATG	AGG	GAA	288
	Glu	Pro	Leu	His	Gln	Met	Leu	Gln	Lys	Gly	Pro	Glu	Thr	Met	Arg	Glu	
					85					90					95		
15	CAA	GCC	TTT	GCA	AAT	GCT	TTT	GGC	AGG	GAG	TTG	ACA	GAT	GCA	TAC	GAG	336
	Gln	Ala	Phe	Ala	Asn	Ala	Phe	Gly	Arg	Glu	Leu	Thr	Asp	Ala	Tyr	Glu	
				100				105						110			
20	TGG	GTG	CTC	AAC	TTT	AGA	AGA	ACT	AAA	GAC	ATA	ACC	AAT	TTG	AAT	CAA	384
	Trp	Val	Leu	Asn	Phe	Arg	Arg	Thr	Lys	Asp	Ile	Thr	Asn	Leu	Asn	Gln	
			115					120					125				
25	GCA	TGG	GAT	ATA	TAC	TAC	AAT	GTC	TTT	AGA	AGA	GTA	AGC	AAA	CAG	GTG	432
	Ala	Trp	Asp	Ile	Tyr	Tyr	Asn	Val	Phe	Arg	Arg	Val	Ser	Lys	Gln	Val	
		130					135					140					
30	CAG	CTG	TTA	GCT	AGT	CTT	GAG	TTG	CAG	TAT	GTA	TCT	CCG	GAC	TTA	GAG	480
	Gln	Leu	Leu	Ala	Ser	Leu	Glu	Leu	Gln	Tyr	Val	Ser	Pro	Asp	Leu	Glu	
	145					150				155					160		
35	CAT	GCT	CAA	GAT	TTG	GAA	TTG	GCT	GTA	CCA	GGT	ACT	TAC	CAA	GCA	GGC	528
	His	Ala	Gln	Asp	Leu	Glu	Leu	Ala	Val	Pro	Gly	Thr	Tyr	Gln	Ala	Gly	
				165				170						175			
40	AAA	CCT	GTG	ATC	AGA	ATA	ATC	AAA	TTT	GAT	CCT	ACT	TTT	TCG	ATT	ATT	576
	Lys	Pro	Val	Ile	Arg	Ile	Ile	Lys	Phe	Asp	Pro	Thr	Phe	Ser	Ile	Ile	
				180				185						190			
45	TCA	TCT	AAA	CAA	AGA	CCG	AGA	AAA	TTA	TCG	TGC	AGA	GGA	AGT	GAT	GGT	624
	Ser	Ser	Lys	Gln	Arg	Pro	Arg	Lys	Leu	Ser	Cys	Arg	Gly	Ser	Asp	Gly	
			195					200					205				
50	AAA	GAC	TAC	CAA	TAT	GCG	TTG	AAA	GGA	CAT	GAA	GAT	ATC	AGA	CAA	GAT	672
	Lys	Asp	Tyr	Gln	Tyr	Ala	Leu	Lys	Gly	His	Glu	Asp	Ile	Arg	Gln	Asp	
		210					215					220					
55	AAC	TTA	GTG	ATG	CAA	TTG	TTT	GGT	TTG	GTT	AAT	ACG	TTG	TTG	GTA	AAT	720
	Asn	Leu	Val	Met	Gln	Leu	Phe	Gly	Leu	Val	Asn	Thr	Leu	Leu	Val	Asn	
	225					230					235				240		
60	GAT	CCG	GTA	TGT	TTC	AAG	AGA	CAT	TTG	GAT	ATA	CAA	CAA	TAT	CCT	GCT	768
	Asp	Pro	Val	Cys	Phe	Lys	Arg	His	Leu	Asp	Ile	Gln	Gln	Tyr	Pro	Ala	
				245					250					255			
65	ATT	CCA	TTA	TCA	CCA	AAA	GTG	GGA	TTG	CTT	GGT	TGG	GTT	CCA	AAT	AGT	816
	Ile	Pro	Leu	Ser	Pro	Lys	Val	Gly	Leu	Leu	Gly	Trp	Val	Pro	Asn	Ser	
				260				265						270			
70	GAC	ACT	TTC	CAT	GTA	TTG	ATC	AAA	GGC	TAT	CGC	GAA	TCA	AGA	AGT	ATA	864
	Asp	Thr	Phe	His	Val	Leu	Ile	Lys	Gly	Tyr	Arg	Glu	Ser	Arg	Ser	Ile	
			275					280					285				
75	ATG	TTG	AAT	ATT	GAA	CAC	AGG	CTT	TTG	TTG	CAA	ATG	GCA	CCT	GAT	TAT	912

94

	Met	Leu	Asn	Ile	Glu	His	Arg	Leu	Leu	Leu	Gln	Met	Ala	Pro	Asp	Tyr	
	290						295					300					
5	GAT	TTC	TTG	ACA	TTA	TTG	CAA	AAA	GTT	GAA	GTG	TTC	ACA	AGT	GCA	ATG	960
	Asp	Phe	Leu	Thr	Leu	Leu	Gln	Lys	Val	Glu	Val	Phe	Thr	Ser	Ala	Met	
	305					310					315					320	
10	GAT	AAT	TGT	AAG	GGA	CAG	GAT	TTG	TAC	AAA	GTG	TTA	TGG	CTC	AAA	TCT	1008
	Asp	Asn	Cys	Lys	Gly	Gln	Asp	Leu	Tyr	Lys	Val	Leu	Trp	Leu	Lys	Ser	
					325					330					335		
15	AAA	TCA	TCC	GAG	GCG	TGG	TTG	GAC	CGT	AGA	ACA	ACA	TAC	ACG	AGA	TCA	1056
	Lys	Ser	Ser	Glu	Ala	Trp	Leu	Asp	Arg	Arg	Thr	Thr	Tyr	Thr	Arg	Ser	
				340					345					350			
20	TTA	GCT	GTA	ATG	TCT	ATG	GTT	GGG	TAT	ATA	TTA	GGT	TTG	GGG	GAT	AGG	1104
	Leu	Ala	Val	Met	Ser	Met	Val	Gly	Tyr	Ile	Leu	Gly	Leu	Gly	Asp	Arg	
			355					360					365				
25	CAC	CCA	TCA	AAT	TTG	ATG	TTG	GAC	CGT	ATT	ACT	GGG	AAA	GTC	ATC	CAT	1152
	His	Pro	Ser	Asn	Leu	Met	Leu	Asp	Arg	Ile	Thr	Gly	Lys	Val	Ile	His	
		370					375					380					
30	ATT	GAT	TTC	GGA	GAC	TGT	TTT	GAA	GCA	GCA	ATA	TTA	CGT	GAG	AAG	TAT	1200
	Ile	Asp	Phe	Gly	Asp	Cys	Phe	Glu	Ala	Ala	Ile	Leu	Arg	Glu	Lys	Tyr	
	385					390					395					400	
35	CCA	GAG	AGA	GTT	CCG	TTT	AGA	TTG	ACG	AGA	ATG	CTT	AAT	TAT	GCC	ATG	1248
	Pro	Glu	Arg	Val	Pro	Phe	Arg	Leu	Thr	Arg	Met	Leu	Asn	Tyr	Ala	Met	
					405					410					415		
40	GAA	GTT	AGT	GGA	ATA	GAG	GGC	TCG	TTC	AGA	ATC	ACA	TGT	GAA	CAT	GTT	1296
	Glu	Val	Ser	Gly	Ile	Glu	Gly	Ser	Phe	Arg	Ile	Thr	Cys	Glu	His	Val	
				420					425					430			
45	ATG	AGG	GTG	TTG	CGT	GAT	AAT	AAA	GAG	TCT	TTA	ATG	GCA	ATA	TTA	GAG	1344
	Met	Arg	Val	Leu	Arg	Asp	Asn	Lys	Glu	Ser	Leu	Met	Ala	Ile	Leu	Glu	
			435					440					445				
50	GCC	TTT	GCT	TAC	GAT	CCC	TTG	ATA	AAT	TGG	GGG	TTT	GAT	TTC	CCA	ACA	1392
	Ala	Phe	Ala	Tyr	Asp	Pro	Leu	Ile	Asn	Trp	Gly	Phe	Asp	Phe	Pro	Thr	
		450					455					460					
55	AAG	GCG	TTG	GCT	GAA	TCA	ACG	GGT	ATA	CGT	GTT	CCA	CAA	GTC	AAC	ACT	1440
	Lys	Ala	Leu	Ala	Glu	Ser	Thr	Gly	Ile	Arg	Val	Pro	Gln	Val	Asn	Thr	
	465					470					475					480	
60	GCA	GAA	TTA	TTA	CGC	AGA	GGA	CAG	ATT	GAC	GAA	AAA	GAA	GCT	GTA	AGA	1488
	Ala	Glu	Leu	Leu	Arg	Arg	Gly	Gln	Ile	Asp	Glu	Lys	Glu	Ala	Val	Arg	
					485					490					495		
65	TTG	CAA	AAG	CAA	AAT	GAA	TTG	GAA	ATA	AGA	AAC	GCT	AGA	GCT	GCA	TTA	1536
	Leu	Gln	Lys	Gln	Asn	Glu	Leu	Glu	Ile	Arg	Asn	Ala	Arg	Ala	Ala	Leu	
				500					505					510			
70	GTG	TTG	AAA	CGT	ATT	ACC	GAT	AAG	TTA	ACT	GGT	AAC	GAT	ATC	AAA	CGG	1584
	Val	Leu	Lys	Arg	Ile	Thr	Asp	Lys	Leu	Thr	Gly	Asn	Asp	Ile	Lys	Arg	
			515					520					525				

95

TTG	AGA	GGA	TTA	GAT	GTG	CCT	ACT	CAA	GTC	GAT	AAA	TTG	ATT	CAA	CAA	1632
Leu	Arg	Gly	Leu	Asp	Val	Pro	Thr	Gln	Val	Asp	Lys	Leu	Ile	Gln	Gln	
530						535					540					

GCC	ACC	AGT	GTT	GAG	AAT	TTG	TGT	CAG	CAT	TAC	ATT	GGT	TGG	TGT	TCG	1680
Ala	Thr	Ser	Val	Glu	Asn	Leu	Cys	Gln	His	Tyr	Ile	Gly	Trp	Cys	Ser	
545					550					555					560	

TGT TGG TAGGTTGATT ATCGTCATGT GTCGATAAGT ATGGTATTGT GGTAACCTATT 1736
Cys Trp

TTATAAAGGG AAATATTAAA GAATTGTATA TTATTAAAAA AAAAAAAAAA AACTCGAG 1794

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 562 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Leu Val Tyr Pro Leu Thr Val Ala Ile Thr Ser Glu Ser Thr Ser Arg
1 5 10 15

Lys Lys Ala Ala Gln Ser Ile Ile Glu Lys Met Arg Val His Ser Pro
20 25 30

Ser Leu Val Asp Gln Ala Glu Leu Val Ser Arg Glu Leu Ile Arg Val
35 40 45

Ala Val Leu Trp His Glu Gln Trp His Asp Ala Leu Glu Asp Ala Ser
50 55 60

Arg Phe Phe Phe Gly Glu His Asn Thr Glu Lys Met Phe Glu Thr Leu
65 70 75 80

Glu Pro Leu His Gln Met Leu Gln Lys Gly Pro Glu Thr Met Arg Glu
85 90 95

Gln Ala Phe Ala Asn Ala Phe Gly Arg Glu Leu Thr Asp Ala Tyr Glu
100 105 110

Trp Val Leu Asn Phe Arg Arg Thr Lys Asp Ile Thr Asn Leu Asn Gln
115 120 125

Ala Trp Asp Ile Tyr Tyr Asn Val Phe Arg Arg Val Ser Lys Gln Val
130 135 140

Gln Leu Leu Ala Ser Leu Glu Leu Gln Tyr Val Ser Pro Asp Leu Glu
145 150 155 160

His Ala Gln Asp Leu Glu Leu Ala Val Pro Gly Thr Tyr Gln Ala Gly

96

			165					170					175			
	Lys	Pro	Val	Ile	Arg	Ile	Ile	Lys	Phe	Asp	Pro	Thr	Phe	Ser	Ile	Ile
				180					185					190		
5	Ser	Ser	Lys	Gln	Arg	Pro	Arg	Lys	Leu	Ser	Cys	Arg	Gly	Ser	Asp	Gly
			195					200					205			
10	Lys	Asp	Tyr	Gln	Tyr	Ala	Leu	Lys	Gly	His	Glu	Asp	Ile	Arg	Gln	Asp
		210					215					220				
	Asn	Leu	Val	Met	Gln	Leu	Phe	Gly	Leu	Val	Asn	Thr	Leu	Leu	Val	Asn
	225					230					235					240
15	Asp	Pro	Val	Cys	Phe	Lys	Arg	His	Leu	Asp	Ile	Gln	Gln	Tyr	Pro	Ala
				245						250					255	
	Ile	Pro	Leu	Ser	Pro	Lys	Val	Gly	Leu	Leu	Gly	Trp	Val	Pro	Asn	Ser
20				260					265					270		
	Asp	Thr	Phe	His	Val	Leu	Ile	Lys	Gly	Tyr	Arg	Glu	Ser	Arg	Ser	Ile
			275					280					285			
25	Met	Leu	Asn	Ile	Glu	His	Arg	Leu	Leu	Leu	Gln	Met	Ala	Pro	Asp	Tyr
		290					295					300				
	Asp	Phe	Leu	Thr	Leu	Leu	Gln	Lys	Val	Glu	Val	Phe	Thr	Ser	Ala	Met
	305					310					315					320
30	Asp	Asn	Cys	Lys	Gly	Gln	Asp	Leu	Tyr	Lys	Val	Leu	Trp	Leu	Lys	Ser
				325						330					335	
	Lys	Ser	Ser	Glu	Ala	Trp	Leu	Asp	Arg	Arg	Thr	Thr	Tyr	Thr	Arg	Ser
35				340					345					350		
	Leu	Ala	Val	Met	Ser	Met	Val	Gly	Tyr	Ile	Leu	Gly	Leu	Gly	Asp	Arg
			355					360					365			
40	His	Pro	Ser	Asn	Leu	Met	Leu	Asp	Arg	Ile	Thr	Gly	Lys	Val	Ile	His
		370					375					380				
	Ile	Asp	Phe	Gly	Asp	Cys	Phe	Glu	Ala	Ala	Ile	Leu	Arg	Glu	Lys	Tyr
	385					390					395					400
45	Pro	Glu	Arg	Val	Pro	Phe	Arg	Leu	Thr	Arg	Met	Leu	Asn	Tyr	Ala	Met
				405						410					415	
	Glu	Val	Ser	Gly	Ile	Glu	Gly	Ser	Phe	Arg	Ile	Thr	Cys	Glu	His	Val
50				420					425					430		
	Met	Arg	Val	Leu	Arg	Asp	Asn	Lys	Glu	Ser	Leu	Met	Ala	Ile	Leu	Glu
			435					440					445			
55	Ala	Phe	Ala	Tyr	Asp	Pro	Leu	Ile	Asn	Trp	Gly	Phe	Asp	Phe	Pro	Thr
		450					455					460				
	Lys	Ala	Leu	Ala	Glu	Ser	Thr	Gly	Ile	Arg	Val	Pro	Gln	Val	Asn	Thr
	465					470					475					480

97

Ala Glu Leu Leu Arg Arg Gly Gln Ile Asp Glu Lys Glu Ala Val Arg
485 490 495

5 Leu Gln Lys Gln Asn Glu Leu Glu Ile Arg Asn Ala Arg Ala Ala Leu
500 505 510

Val Leu Lys Arg Ile Thr Asp Lys Leu Thr Gly Asn Asp Ile Lys Arg
515 520 525

10 Leu Arg Gly Leu Asp Val Pro Thr Gln Val Asp Lys Leu Ile Gln Gln
530 535 540

15 Ala Thr Ser Val Glu Asn Leu Cys Gln His Tyr Ile Gly Trp Cys Ser
545 550 555 560

Cys Trp

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..399

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

40 GTT AGT CAC GAG TTG ATC AGA GTA GCC GTT CTA TGG CAC GAA TTA TGG 48
Val Ser His Glu Leu Ile Arg Val Ala Val Leu Trp His Glu Leu Trp
1 5 10 15

45 TAT GAA GGA CTG GAA GAT GCG AGC CGC CAA TTT TTC GTT GAA CAT AAC 96
Tyr Glu Gly Leu Glu Asp Ala Ser Arg Gln Phe Phe Val Glu His Asn
20 25 30

50 ATA GAA AAA ATG TTT TCT ACT TTA GAA CCT TTA CAT AAA CAC TTA GGC 144
Ile Glu Lys Met Phe Ser Thr Leu Glu Pro Leu His Lys His Leu Gly
35 40 45

AAT GAG CCT CAA ACG TTA AGT GAG GTA TCG TTT CAG AAA TCA TTT GGT 192
Asn Glu Pro Gln Thr Leu Ser Glu Val Ser Phe Gln Lys Ser Phe Gly
50 55 60

55 AGA GAT TTG AAC GAT GCC TAC GAA TGG TTG AAT AAC TAC AAA AAG TCA 240
Arg Asp Leu Asn Asp Ala Tyr Glu Trp Leu Asn Asn Tyr Lys Lys Ser
65 70 75 80

98

AAA GAC ATC AAT AAT TTG AAC CAA GCT TGG GAT ATT TAT TAT AAC GTC	288
Lys Asp Ile Asn Asn Leu Asn Gln Ala Trp Asp Ile Tyr Tyr Asn Val	
859095	
5TTC AGA AAA ATA ACA CGT CAA ATA CCA CAG TTA CAA ACC TTA GAC TTA	336
Phe Arg Lys Ile Thr Arg Gln Ile Pro Gln Leu Gln Thr Leu Asp Leu	
100105110	
CAG CAT GTT TCT CCC CAG CTT CTG GCT ACT CAT GAT CTC GAA TTG GCT	384
Gln His Val Ser Pro Gln Leu Leu Ala Thr His Asp Leu Glu Leu Ala	
115120125	
GTT CCT GGG ACA TAT	399
Val Pro Gly Thr Tyr	
15130	

(2) INFORMATION FOR SEQ ID NO:16:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Ser His Glu Leu Ile Arg Val Ala Val Leu Trp His Glu Leu Trp	
151015	
Tyr Glu Gly Leu Glu Asp Ala Ser Arg Gln Phe Phe Val Glu His Asn	
202530	
Ile Glu Lys Met Phe Ser Thr Leu Glu Pro Leu His Lys His Leu Gly	
354045	
Asn Glu Pro Gln Thr Leu Ser Glu Val Ser Phe Gln Lys Ser Phe Gly	
505560	
Arg Asp Leu Asn Asp Ala Tyr Glu Trp Leu Asn Asn Tyr Lys Lys Ser	
65707580	
Lys Asp Ile Asn Asn Leu Asn Gln Ala Trp Asp Ile Tyr Tyr Asn Val	
859095	
Phe Arg Lys Ile Thr Arg Gln Ile Pro Gln Leu Gln Thr Leu Asp Leu	
100105110	
Gln His Val Ser Pro Gln Leu Leu Ala Thr His Asp Leu Glu Leu Ala	
115120125	
Val Pro Gly Thr Tyr	
130	

55

(2) INFORMATION FOR SEQ ID NO:17:

99

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 1..399

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

5	GTC AGC CAC GAA TTG ATA CGT ATG GCG GTG CTT TGG CAT GAG CAA TGG	48
	Val Ser His Glu Leu Ile Arg Met Ala Val Leu Trp His Glu Gln Trp	
	1 5 10 15	
10	TAT GAG GGT CTG GAT GAC GCC AGT AGG CAG TTT TTT GGA GAA CAT AAT	96
	Tyr Glu Gly Leu Asp Asp Ala Ser Arg Gln Phe Phe Gly Glu His Asn	
	20 25 30	
25	ACC GAA AAA ATG TTT GCT GCT TTA GAG CCT CTG TAC GAA ATG CTG AAG	144
	Thr Glu Lys Met Phe Ala Ala Leu Glu Pro Leu Tyr Glu Met Leu Lys	
	35 40 45	
30	AGA GGA CCG GAA ACT TTG AGG GAA ATA TCG TTC CAA AAT TCT TTT GGT	192
	Arg Gly Pro Glu Thr Leu Arg Glu Ile Ser Phe Gln Asn Ser Phe Gly	
	50 55 60	
35	AGG GAC TTG AAT GAC GCT TAC GAA TGG CTG ATG AAT TAC AAA AAA TCT	240
	Arg Asp Leu Asn Asp Ala Tyr Glu Trp Leu Met Asn Tyr Lys Lys Ser	
	65 70 75 80	
40	AAA GAT GTT AGT AAT TTA AAC CAA GCG TGG GAC ATT TAC TAT AAT GTT	288
	Lys Asp Val Ser Asn Leu Asn Gln Ala Trp Asp Ile Tyr Tyr Asn Val	
	85 90 95	
45	TTC AGG AAA ATT GGT AAA CAG TTG CCA CAA TTA CAA ACT CTT GAA CTA	336
	Phe Arg Lys Ile Gly Lys Gln Leu Pro Gln Leu Gln Thr Leu Glu Leu	
	100 105 110	
50	CAA CAT GTG TCG CCA AAA CTA CTA TCT GCG CAT GAT TTG GAA TTG GCT	384
	Gln His Val Ser Pro Lys Leu Leu Ser Ala His Asp Leu Glu Leu Ala	
	115 120 125	
55	GTC CCC GGG ACC CGT	399
	Val Pro Gly Thr Arg	
	130	

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 133 amino acids
(B) TYPE: amino acid

100

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Val Ser His Glu Leu Ile Arg Met Ala Val Leu Trp His Glu Gln Trp
1 5 10 15
Tyr Glu Gly Leu Asp Asp Ala Ser Arg Gln Phe Phe Gly Glu His Asn
20 25 30
Thr Glu Lys Met Phe Ala Ala Leu Glu Pro Leu Tyr Glu Met Leu Lys
35 40 45
15 Arg Gly Pro Glu Thr Leu Arg Glu Ile Ser Phe Gln Asn Ser Phe Gly
50 55 60
Arg Asp Leu Asn Asp Ala Tyr Glu Trp Leu Met Asn Tyr Lys Lys Ser
20 65 70 75 80
Lys Asp Val Ser Asn Leu Asn Gln Ala Trp Asp Ile Tyr Tyr Asn Val
85 90 95
25 Phe Arg Lys Ile Gly Lys Gln Leu Pro Gln Leu Gln Thr Leu Glu Leu
100 105 110
Gln His Val Ser Pro Lys Leu Leu Ser Ala His Asp Leu Glu Leu Ala
115 120 125
30 Val Pro Gly Thr Arg
130

35 (2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 531 base pairs
(B) TYPE: nucleic acid
40 (C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TGACCCTCAC CCCTTCCACC TATCCCAAAA ACCTCACTGG GTCTGTGGAC AAACAACANA 60
50 AATNTTTTCC ANANAGGCCC CAAATGAGNC CCANGGGTCT NTCTTCCATC AGACCCAGTG 120
ATTCTGCGAC TCACACNCTT CAATTCAAGA CCTGACCNCT AGTAGGGAGG TTTANTCAGA 180
55 TCGCTGGCAN CCTCGGCTGA NCAGATNCAN AGNGGGGNTC GCTGTTCAGT GGGNCCACCC 240
TCNCTGGCCT TCTTCANCAG GGGTCTGGGA TGTTTTTCAGT GGNCNAANA CNCTGTTTAG 300

101

AGCCAGGGCT CAGNAAACAG AAAANCTNTC ATGGNGGTTT TGGACACAGG GNAGGTCTGG 360
NACATATTGG GGATTATGAN CAGNACCAAN ACNCCACTAA ATNCCCCAAG NANAAAGTGT 420
AACCATNTCT ANACNCCATN TTNTATCAGN ANAAATTTTN TTCCNATAAA TGACATCAGN 480
ANTTTNAACA TNAAAAAAAAA AAAAAAAAAA AAAANAAAAA AAAAAAAAAA A 531

10 (2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 231 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(ix) FEATURE:

- 25 (A) NAME/KEY: misc_feature
(B) LOCATION: 128
(D) OTHER INFORMATION: /label= XhoI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

30 GCGTATAACG CGTTTGGAAT CACTACAGGG ATGTTTAATA CCACTACAAT GGATGATGTA 60
TATAACTATC TATTCGATGA TGAAGATACC CCACCAAACC CAAAAAAGA GATCTGGAAT 120
TCGGATCCTC GAGAGATCTA TGAATCGTAG ATACTGAAAA ACCCCGCAAG TTCACCTCAA 180
35 CTGTGCATCG TGCACCATCT CAATTTCTTT CATTTATACA TCGTTTTGCC T 231

(2) INFORMATION FOR SEQ ID NO:21:

40 (i) SEQUENCE CHARACTERISTICS;

- (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
45 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TGAAGATACC CCACCAAACC C 21

(2) INFORMATION FOR SEQ ID NO:22:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs

102

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: other nucleic acid

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGCACAGTTG AAGTGAAC 18

15 (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 907 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25 (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 34..507

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCCGGGGCTG CGGCCGCCCCG AGGGACTTTG AAC ATG TCG GGG ATC GCC CTC AGC 54
Met Ser Gly Ile Ala Leu Ser
1 5

35 AGA CTC GCC CAG GAG AGG AAA GCA TGG AGG AAA GAC CAC CCA TTT GGT 102
Arg Leu Ala Gln Glu Arg Lys Ala Trp Arg Lys Asp His Pro Phe Gly
10 15 20

40 TTC GTG GCT GTC CCA ACA AAA AAT CCC GAT GGC ACG ATG AAC CTC ATG 150
Phe Val Ala Val Pro Thr Lys Asn Pro Asp Gly Thr Met Asn Leu Met
25 30 35

45 AAC TGG GAG TGC GCC ATT CCA GGA AAG AAA GGG ACT CCG TGG GAA GGA 198
Asn Trp Glu Cys Ala Ile Pro Gly Lys Lys Gly Thr Pro Trp Glu Gly
40 45 50 55

50 GGC TTG TTT AAA CTA CGG ATG CTT TTC AAA GAT GAT TAT CCA TCT TCG 246
Gly Leu Phe Lys Leu Arg Met Leu Phe Lys Asp Asp Tyr Pro Ser Ser
60 65 70

CCA CCA AAA TGT AAA TTC GAA CCA CCA TTA TTT CAC CCG AAT GTG TAC 294
Pro Pro Lys Cys Lys Phe Glu Pro Pro Leu Phe His Pro Asn Val Tyr
75 80 85

55 CCT TCG GGG ACA GTG TGC CTG TCC ATC TTA GAG GAG GAC AAG GAC TGG 342
Pro Ser Gly Thr Val Cys Leu Ser Ile Leu Glu Glu Asp Lys Asp Trp
90 95 100

103

5 AGG CCA GCC ATC ACA ATC AAA CAG ATC CTA TTA GGA ATA CAG GAA CTT 390
Arg Pro Ala Ile Thr Ile Lys Gln Ile Leu Leu Gly Ile Gln Glu Leu
105 110 115

CTA AAT GAA CCA AAT ATC CAA GAC CCA GCT CAA GCA GAG GCC TAC ACG 438
Leu Asn Glu Pro Asn Ile Gln Asp Pro Ala Gln Ala Glu Ala Tyr Thr
120 125 130 135

10 ATT TAC TGC CAA AAC AGA GTG GAG TAC GAG AAA AGG GTC CGA GCA CAA 486
Ile Tyr Cys Gln Asn Arg Val Glu Tyr Glu Lys Arg Val Arg Ala Gln
140 145 150

15 GCC AAG AAG TTT GCG CCC TCA TAAGCAGCGA CCTTGTGGCA TCGTCAAAAG 537
Ala Lys Lys Phe Ala Pro Ser
155

GAAGGGATTG GTTTGGCAAG AACTTGTTTA CAACATTTTT GGCAAATCTA AAGTTGCTCC 597

20 ATACAATGAC TAGTCACCTG GGGGGGTTGG GCGGGCGCCA TCTTCCATTG CCGCCGCGGG 657

TGTGCGGTCT CGATTCGCTG AATTGCCCGT TTCCATACAG GGTCTCTTCC TTCGGTCTTT 717

25 TGGTATTTTT GGATTGTTAT GTAAAACTCG CTTTTATTTT AATATTGATG TCAGTATTTT 777

AACTGCTGTA AAATTATAAA CTTTATATACT GGGTAAGTCC CCCAGGGGCG AGTTNCCTCG 837

CTCTGGGATG CAGGCATGCT TCTCACCGTG CAGAGCTGCA CTTGNCCTCA GCTGNCTGNA 897

30 TGGAAATGCA 907

(2) INFORMATION FOR SEQ ID NO:24:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 158 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

45 Met Ser Gly Ile Ala Leu Ser Arg Leu Ala Gln Glu Arg Lys Ala Trp
1 5 10 15

Arg Lys Asp His Pro Phe Gly Phe Val Ala Val Pro Thr Lys Asn Pro
20 25 30

50 Asp Gly Thr Met Asn Leu Met Asn Trp Glu Cys Ala Ile Pro Gly Lys
35 40 45

Lys Gly Thr Pro Trp Glu Gly Gly Leu Phe Lys Leu Arg Met Leu Phe
50 55 60

55 Lys Asp Asp Tyr Pro Ser Ser Pro Pro Lys Cys Lys Phe Glu Pro Pro
65 70 75 80

104

Leu Phe His Pro Asn Val Tyr Pro Ser Gly Thr Val Cys Leu Ser Ile
85 90 95

5 Leu Glu Glu Asp Lys Asp Trp Arg Pro Ala Ile Thr Ile Lys Gln Ile
100 105 110

Leu Leu Gly Ile Gln Glu Leu Leu Asn Glu Pro Asn Ile Gln Asp Pro
115 120 125

10 Ala Gln Ala Glu Ala Tyr Thr Ile Tyr Cys Gln Asn Arg Val Glu Tyr
130 135 140

Glu Lys Arg Val Arg Ala Gln Ala Lys Lys Phe Ala Pro Ser
145 150 155

15

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
20 (A) LENGTH: 207 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

30 CCCTCCCTCC TGCCGCTCCT CTCTAGAACC TTCTAGAACC TGGGCTGTGC TGCTTTTGAG 60

CCTCAGACCC CAGGGCAGCA TCTCGGTTCT GCGCCACTTC CTTTGTGTTT ANATGGCGTT 120

35 TTGTCTGTGT TGCTGTTTAG AGTAGATNAA CTGTTTANAT AAAAAAAAAA NAAAATTNAC 180

TNGAGGGGGC NTGNAGGCAT GCNNAAC 207

CLAIMS:

1. A nucleic acid encoding a fusion polypeptide consisting of: a portion of a RAPT1 protein of 20 to 200 amino acid residues, which RAPT1 portion (a) is encoded by a nucleic acid which hybridizes to the complement of one of SEQ ID NO. 1 under stringent conditions including a wash step of 0.2x SSC at 65 °C., and (b) binds an FKBP/rapamycin complex in a rapamycin dependent manner; and one or more additional polypeptide portions heterologous to a native RAPT1 protein, which polypeptide portions are selected from (i) a domain that facilitates purification; (ii) a detectable label; (iii) a matrix-binding domain; (iv) a DNA binding domain; and (v) a transcriptional activation domain.
2. A nucleic acid encoding a fusion polypeptide consisting essentially of: a portion of a RAPT1 protein of 20 to 200 amino acid residues, which RAPT1 portion (a) is encoded by a nucleic acid which hybridizes to the complement of SEQ ID NO. 1 under stringent conditions including a wash step of 0.2x. SSC at 65 °C., and (b) binds an FKBP/rapamycin complex in a rapamycin dependent manner; wherein the RAPT1 portion lacks an endogenous kinase activity; and one or more additional polypeptide portions heterologous to a native RAPT1 protein, which polypeptide portions are selected from (i) a domain that facilitates purification; (ii) a detectable label; (iii) a matrix-binding domain; (iv) a DNA binding domain; and (v) a transcriptional activation domain.
3. The nucleic acid of claim 1 or 2, wherein said fusion polypeptide includes a detectable label for detecting the presence of said fusion polypeptide.
4. The nucleic acid of claim 1 or 2, wherein said fusion polypeptide includes a matrix-binding domain for immobilizing said fusion polypeptide on an insoluble matrix.
5. The nucleic acid of claim 1 or 2, wherein said fusion polypeptide includes a DNA binding domain.

6. The nucleic acid of claim 1 or 2, further comprising a transcriptional regulatory sequence operably linked to said nucleotide sequence so as permit expression of the polypeptide in a cell.
7. The nucleic acid of claim 1 or 2, wherein the RAPT1 portion is a 50 to 133 amino acid fragment of the RAPT1 protein.
8. The nucleic acid of claim 2, wherein the RAPT1 portion comprises an amino acid sequence from a human RAPT1 protein.
9. The nucleic acid of claim 1 or 2, wherein the fusion polypeptide is soluble.
10. The nucleic acid of claim 1 or 2, wherein said fusion polypeptide includes a transcriptional activation domain.
11. An expression vector, capable of replicating in at least one of a prokaryotic cell and a eukaryotic cell, comprising the nucleic acid of claim 1 or 2.
12. The expression vector of claim 11, which is a viral vector.
13. The expression vector of claim 12, wherein the viral vector is a retrovirus, an adenovirus, an adeno-associated virus, a herpes simplex virus, or a vaccinia virus.
14. A host cell transfected with the expression vector of claim 11 and expressing said fusion polypeptide.
15. A method of producing a recombinant fusion polypeptide comprising culturing the cell of claim 14 in a cell culture medium to express said fusion polypeptide and isolating said fusion polypeptide from said cell culture.
16. A method of measuring the nucleic acid level in a cell of a gene, or determining whether said gene is mutated or deleted in a cell, which gene comprises the nucleotide

sequence of SEQ ID NO. 11 and encodes a protein that binds to an FKBP/rapamycin complex, using a probe or primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions, including a wash step of 0.2X SSC at 65 °C, to at least 20 consecutive nucleotides of sense or antisense sequence of SEQ ID NO. 1 which method comprises probing an extract of said cell, hybridization of said probe providing a measure of said nucleic acid level or a determination as to whether or not said gene has mutated or been deleted.

17. The method of claim 16, wherein the probe or primer further comprises a detectable label group attached thereto.

18. The method of claim 17, wherein said label group is a radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

19. The method of claim 16, wherein said probe or primer is part of a diagnostic test kit.

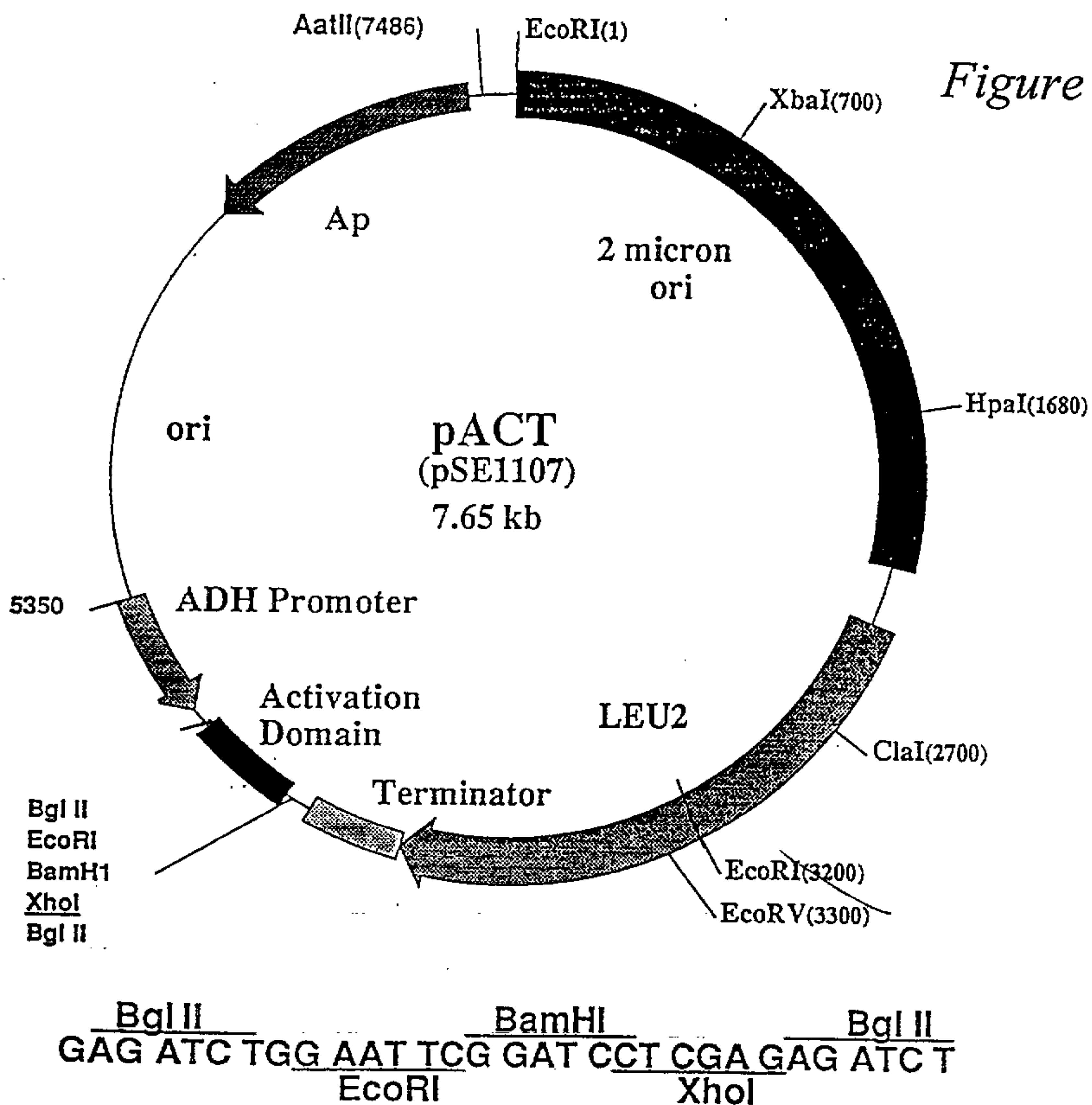
20. The method of claim 19, wherein said diagnostic test kit is for any of the following: identifying transformed cells, measuring mRNA expression levels, determining whether a cell contains a mutated gene, or determining whether said gene is deleted in a cell.

21. The method of claim 16, wherein said oligonucleotide is at least 40 nucleotides in length.

22. A method according to claim 16 for identifying cells transformed with a nucleic acid encoding a portion of a RAPT1 protein.

23. The method of claim 16, wherein said oligonucleotide corresponds to at least 20 consecutive nucleotides of sense or antisense sequence of SEQ ID NO. 1.

24. The method of claim 16, wherein said oligonucleotide corresponds to at least 40 consecutive nucleotides of sense or antisense sequence of SEQ ID NO. 1.

Figure 1

cDNAs are inserted at the XhoI site

2/3

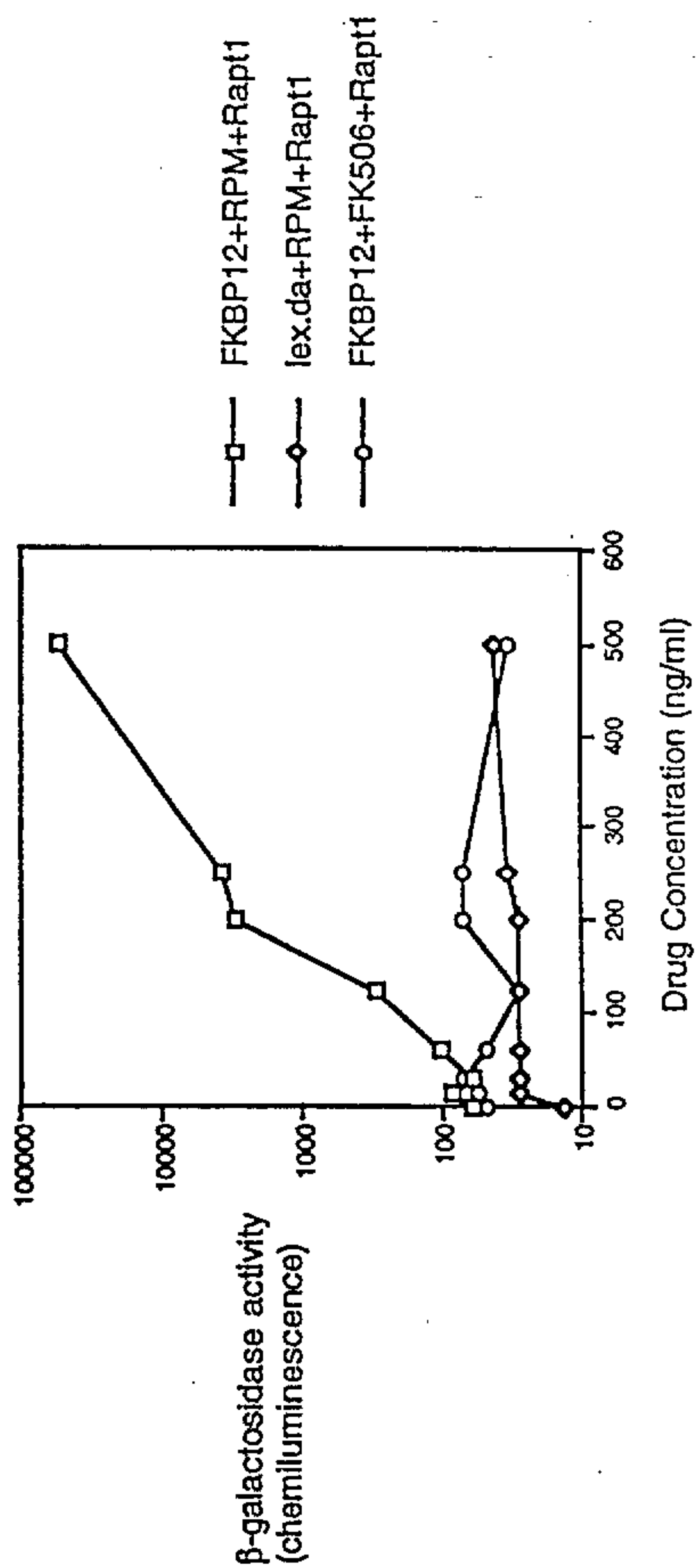


FIGURE 2

3/3

Figure 3

