

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
13 février 2003 (13.02.2003)

PCT

(10) Numéro de publication internationale
WO 03/012102 A2

(51) Classification internationale des brevets⁷ :

C12N 15/12, 9/00, C07K 14/47, C12N 15/52, C12Q
1/68, 1/25, G01N 33/53, A61K 38/17

(21) Numéro de la demande internationale :

PCT/FR02/02782

(22) Date de dépôt international : 1 août 2002 (01.08.2002)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :

01/10364 1 août 2001 (01.08.2001) FR

(71) Déposant (pour tous les États désignés sauf US) : **CY-
TOMICS SYSTEMS** [FR/FR]; Bâtiment 5, 1, avenue de
la Terrasse, F-91198 GIF sur Yvette Cedex (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : **THOMAS,
Dominique** [FR/FR]; 34, allée des Gravières de la
Salmouille, F-91190 GIF sur Yvette (FR). **SARGUEIL,
Bruno** [FR/FR]; 69, rue Oberkampf, F-75011 Paris
(FR). **SCAEROU, Frédéric** [FR/FR]; 5, rue de l'Ecosse,
F-29300 Quimperle (FR).

(74) Mandataires : **CATHERINE, Alain** etc.; Cabinet Harle
et Phelip, 7, rue de Madrid, F-75008 Paris (FR).

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

(84) États désignés (régional) : brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet
eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet
européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), brevet
OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,
MR, NE, SN, TD, TG).

Déclaration en vertu de la règle 4.17 :

— relative à la qualité d'inventeur (règle 4.17.iv)) pour US
seulement

Publiée :

— sans rapport de recherche internationale, sera republiée
dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abrévia-
tions, se référer aux "Notes explicatives relatives aux codes et
abréviations" figurant au début de chaque numéro ordinaire de
la Gazette du PCT.

(54) Title: NUCLEIC ACIDS CODING FOR NOVEL HECT DOMAIN PROTEINS, THEIR DIAGNOSTIC AND THERAPEU-
TIC USES

(54) Titre : ACIDES NUCLEIQUES CODANT DE NOUVELLES PROTEINES A DOMAINE HECT, LEURS UTILISATIONS
EN DIAGNOSTICS ET EN THERAPIE

(57) Abstract: The invention concerns novel HECT domain proteins and novel HECT domain protein substrates. The invention
provides screening methods for identifying novel HECT protein substrates and for identifying small molecules and drugs which
modulate the interaction and/or activity of HECT domain proteins and of their substrates.

(57) Abrégé : La présente invention a pour objet de nouvelles protéines à domaine HECT et de nouveaux substrats des protéines
à domaine HECT. Elle a pour objet les méthodes de criblage mises au point pour identifier les substrats des nouvelles protéines à
domaine HECT et pour identifier les petites molécules et drogues qui modulent l'interaction et/ou l'activité des protéines à domaine
HECT et de leurs substrats.



WO 03/012102 A2

Acides nucléiques codant de nouvelles protéines à domaine HECT, leurs utilisations en diagnostics et en thérapie

L'invention se rapporte au domaine de l'ubiquitination des protéines, et plus particulièrement au domaine du contrôle et du diagnostic de l'ubiquitination des protéines.

5

ART ANTERIEUR

Il est depuis longtemps établi que la durée de vie des protéines cellulaires est extrêmement variable. Cette instabilité constitutive ou régulée de certaines protéines (en général des protéines régulatrices),
10 permet aux cellules d'en contrôler étroitement et d'en modifier rapidement les niveaux.

Les mécanismes de dégradation des protéines sont donc extrêmement spécifiques, et le niveau intra-cellulaire d'une protéine donnée peut être ajusté précisément aux besoins de la cellule, par un
15 équilibre entre ses taux de synthèse et de dégradation.

Ce n'est que très récemment par contre qu'il a été établi que la majeure partie de la protéolyse intra-cellulaire est assurée chez les eucaryotes par une machinerie très sophistiquée, le système Ubiquitine (Ub) /protéasome, qui utilise l'Ub pour marquer les protéines à dégrader
20 et le protéasome comme enzyme protéolytique. Cette voie implique une cascade de réactions qui conduisent, dans un premier temps, au marquage des protéines à détruire par un polypeptide, constitué de 76 acides aminés, appelé ubiquitine. L'ajout de plusieurs molécules d'ubiquitine cible alors la protéine ainsi modifiée vers le protéasome où
25 elle est détruite. Cette voie s'avère être la voie privilégiée de la dégradation de la plupart des protéines à courte durée de vie chez l'ensemble des organismes eucaryotes. Un grand nombre des protéines dont la stabilité est régulée par la voie ubiquitine-protéasome assurent des fonctions régulatrices dans la cellule (des régulateurs du cycle cellulaire, comme les cyclines, par exemple, des facteurs de
30 transcription, des récepteurs etc....). Le système ubiquitine-protéasome est donc un système régulateur important des cellules eucaryotes qui

contrôle la concentration cellulaire de protéines clés par leur dégradation sélective.

La voie ubiquitine-protéasome

5 L'ubiquitination régule un grand nombre de fonctions cellulaires critiques, le plus fréquemment en déterminant la dégradation sélective de certaines protéines clés par le protéasome. La progression du cycle cellulaire (Koepp DM et al., 1999), l'induction de la réponse inflammatoire (Ghosh S, et al., 1998), la présentation antigénique (Rock KL et Goldberg
10 AL., 1999) notamment sont quelques uns des nombreux processus régulés par la voie ubiquitine protéasome. Il n'est pas étonnant de constater qu'une dérégulation de la protéolyse dépendante de l'ubiquitine a été identifiée comme une cause du cancer (Joazeiro CAP et al., 1999; Bignell GR et al., 2000) et de plusieurs maladies héréditaires (Shimura H
15 et al., 2000; Matsuura T et al., 1997). La dégradation n'est pas la seule issue possible pour des protéines liées à l'ubiquitine; l'ubiquitination régule aussi certains processus par des mécanismes qui ne semblent pas nécessiter de protéolyse. Ces processus incluent notamment l'initiation de la réponse inflammatoire (Deng L et al., 2000) et le
20 fonctionnement de certains facteurs de transcription (Kaiser P et al., 2000). Parmi les protéines concernées par l'ubiquitination, on trouve notamment p27, p53, p300, les cyclines, E2F, STAT-1, c-Myc, c-Jun, IKB, NFkB et la β -caténine.

25 La voie ubiquitine conduit à l'attachement covalent d'une chaîne d'ubiquitine aux substrats cibles qui sont ensuite dégradés par un complexe multiprotéique, le protéasome, au sein duquel sont associées de nombreuses activités protéasiques. L'ubiquitination de la protéine cible est un mécanisme en trois étapes qui implique en générale trois
30 classes d'enzymes. L'ubiquitine est d'abord activée dans sa partie C-terminale par la formation d'un lien thioester avec une enzyme d'activation de l'ubiquitine, l'enzyme E1. Ensuite, l'ubiquitine est transférée aux résidus cystéine actifs d'une des nombreuses enzymes de conjugaison de l'ubiquitine (Ubc ou E2). Le transfert final de l'ubiquitine
35 aux résidus lysine de la protéine cible se produit par une réaction qui

requiert ou non une protéine ligase de l'ubiquitine (E3) (Hochstrasser, 1996). Les enzymes E3 ont pour principale fonction de reconnaître le substrat mais ces enzymes catalysent parfois la formation d'un lien peptidique entre l'ubiquitine et un résidu lysine du substrat. La répétition de l'action de ces trois enzymes conduit à la formation d'une chaîne de polyubiquitine. La protéine polyubiquitinée est alors adressée vers le protéasome où elle est dégradée. Alors que le facteur E1 est commun à toutes les voies de dégradation et ne sert qu'à activer l'ubiquitine, la sélectivité du processus de dégradation est assurée par l'ubiquitine ligase E3, qui interagit à la fois avec E2 et avec le substrat (Hershko A. et al., 1983).

Il existe donc un besoin dans l'état de la technique pour l'isolement et la caractérisation de nouvelles protéines à domaine HECT, qui constituent des cibles privilégiées de composés thérapeutiques utilisables pour la prévention ou le traitement des pathologies dans lesquelles ces protéines sont impliquées, notamment les pathologies qui ont été indiquées ci-dessus.

Les ubiquitine-ligases: les enzymes à domaine HECT

Les ubiquitine-ligases ou E3 sont les déterminants centraux de la spécificité de l'ubiquitination. Bien que cette propriété soit clairement établie, il s'est avéré difficile de comprendre la catalyse dépendante des E3 au niveau moléculaire. Dans la plupart des cas, le substrat, l'ubiquitine et l'enzyme qui porte l'ubiquitine (E2) doivent tous être présents au niveau du centre réactionnel. En conséquence, en plus de contribuer au transfert chimique de l'ubiquitine aux résidus lysine, l'enzyme E3 doit également coordonner le positionnement des différents acteurs protéiques au niveau de son site actif.

Les enzymes E3 connues à l'heure actuelle sont des membres de deux familles protéiques: les E3 à domaine HECT et les E3 à RING finger. Les domaines catalytiques de ces deux familles ne sont apparentés ni en séquence ni en structure (néanmoins certaines E2 interagissent bien avec des E3 des deux familles).

Récemment, il est devenu clair que la plupart des E3 ont en commun un élément structural particulier, le RING (Really Interesting New Gene) finger (Lorick KL et al., 1999; Deshaies RJ, 1999). Les 70 acides aminés du RING finger présentent un espacement distinct des résidus cystéine et histidine dont le rôle est d'assurer la fixation de deux ions zinc qui stabilisent la structure globulaire caractéristique (Borden KLB, 2000). Dans certains cas bien documentés, l'E3 à domaine RING est une sous-unité d'un complexe multiprotéique. C'est le cas notamment pour les complexes SCF, APC et VCB (Deshaies RJ, 1999; Tyers M et Jorgensen P, 2000). Les complexes SCF et VCB contiennent une protéine à domaine RING désignée Rbx1, et l'APC contient la protéine APC11. Cependant, dans ces cas particuliers, les enzymes E3 ne semblent pas jouer un rôle dans la reconnaissance du substrat, ce rôle étant assuré par d'autres protéines du complexe multiprotéique, comme les protéines à boîte F dans les complexes SCF (S: Skp1; C: Cdc53 ou culline; F: protéine contenant une boîte F). Certaines protéines à RING finger fonctionnent également de manière autonome comme la protéine c-Cbl impliquée dans la dégradation des récepteurs de l'EGF (epidermal growth factor) ou de PDGF (platelet derived growth factor) (Joazeiro CAP and Weissman AM, 2000) ou la protéine Mdm2 responsable principalement de la dégradation de p53 par le protéasome (Haupt Y et al., 1997).

Les protéines à domaine HECT (pour "Homologous to E6-AP Carboxyl-Terminus") sont caractérisées par un motif d'environ 350 acides aminés contenant un résidu cystéine très conservé qui forme un intermédiaire thiol ester essentiel pendant la catalyse. Au niveau structural, les protéines à domaine HECT sont des protéines de grande taille constituées de nombreux domaines de liaison. Certaines comportent des domaines C2 (Ca²⁺/lipid binding domain) et des domaines WW (répétition de 35-40 acides aminés de deux résidus tryptophane) en plus du domaine HECT. Les domaines WW sont soupçonnés intervenir dans la reconnaissance du substrat (Pirozzi G. et al., 1997). De nombreuses protéines HECT contiennent des domaines WW. Il a été suggéré que in vivo ces protéines peuvent interagir avec

diverses autres protéines par l'intermédiaire de ce domaine (Harvey KF et Kumar S., 1999). Un domaine régulateur de la condensation des chromosomes a également été trouvée dans certaines de ces protéines (RCC).

- 5 La propriété d'un de ces domaines HECT fut d'abord révélée grâce à l'étude de la dégradation conditionnelle du suppresseur de tumeur p53. Quand les cellules sont infectées par des formes oncogéniques des papillomavirus humains (HPV), p53 est sélectivement dégradée par une protéolyse dépendante de la voie ubiquitine protéasome. Des études
10 biochimiques ont révélées que la reconnaissance de p53 pour une dégradation induite par le virus dépendait du produit du gène E6 de HPV et d'une protéine cellulaire de 100 kDa environ, E6-AP (E6-Associated Protein). E6 et E6-AP forment un complexe qui fonctionne comme une E3 spécifique de p53 (Scheffner M., 1993). Le tiers terminal de la
15 séquence de E6-AP est identique à 35-45% à de nombreuses autres protéines dans les banques de données (Huibregtse JM, 1995).

- Dans les bases de données de mammifères notamment, on peut trouver de nombreuses protéines HECT potentielles, mais la grande majorité d'entre elles ne sont pas caractérisées fonctionnellement. A part
20 E6-AP, la protéine E3 à domaine HECT la mieux caractérisée est l'enzyme de levure Rsp5. Rsp5 a été impliquée dans l'ubiquitination de divers types de substrats très différents. Rsp5 reconnaît un ensemble de perméases et récepteurs membranaires dont l'ubiquitination dépend dans certains cas d'un signal produit par la phosphorylation du substrat.
25 Ces substrats sont ensuite endocytés. Rsp5 a aussi des substrats solubles dont Rpb1, la grande sous-unité de l'ARN polymérase II. L'ubiquitination conduit dans ce cas à la protéolyse. Les propriétés de Rsp5 sont conservées chez les eucaryotes supérieurs. La protéine humaine Rpb1 est reconnue par Nedd4, l'orthologue murin/humain de
30 Rsp5 (Beaudenon SL et al., 1999).

La plupart des protéines à domaine HECT sont vraisemblablement des E3 ou des composants de complexes multiprotéiques contenant des activités de type E3. A l'heure actuelle, peu d'informations sont disponibles sur les fonctions possibles des E3 à domaine HECT.

Ces découvertes suggèrent qu'un vaste répertoire de E3 a été établi en rassemblant divers domaines d'interaction protéique avec seulement deux motifs catalytiques. En dépit de ce progrès cependant, et malgré la détermination récente de la structure de deux complexes E2-E3 (Huang L et al., 1999; Zheng N et al., 2000) qui montrent comment les différents intervenants peuvent dans certains cas s'aligner au niveau du site actif de l'E3, de nombreuses questions, mécanistiques notamment, restent posées.

10 **Les pathologies associées aux ubiquitine ligase E3**

Le système ubiquitine est impliqué dans la pathologie de nombreuses maladies génétiques humaines dont les syndromes d'Angelman et de Liddle. Le syndrome d'Angelman est un défaut rare et héréditaire se traduisant par un retard mental associé à une mutation de la protéine à domaine HECT, E3 (Kishino, T et al., 1997; Matsuura, T. et al., 1997). Le syndrome de Liddle impliquant l'ubiquitine ligase à domaine HECT, Nedd4, est une forme rare et héréditaire d'hypertension causée par la délétion de régions du canal sodique épithélial (ENaC) qui le cible pour la dégradation dépendante de l'ubiquitine (Staub O et al., 1996). Cela se traduit par une hyperactivité de ce canal. L'accumulation de produits conjugués à l'ubiquitine est aussi une des caractéristiques de nombreuses maladies neurologiques comme la maladie d'Alzheimer (Checler F et al., 2000). Cette accumulation est en fait une conséquence des tentatives avortées de la cellule pour se débarrasser de protéines abimées ou anormales. Une protéolyse accélérée a également été associée à une atrophie des muscles squelettiques (Wing S. et al., 1995).

Les cancers se développent quand les cellules se multiplient trop rapidement. La prolifération cellulaire dépend d'un équilibre entre les signaux négatifs et les signaux positifs. Quand les signaux positifs sont supérieurs ou que les signaux négatifs sont absents, les cellules se multiplient trop rapidement et le cancer se développe. Normalement, les cellules contrôlent précisément la quantité d'une protéine donnée et éliminent l'excès ou toute protéine non requise. Pour ce faire, la cellule marque spécifiquement ces protéines indésirables avec une longue chaîne d'ubiquitine. Ces molécules sont alors reconnues et détruites par

le protéasome. Cependant, cette machinerie s'enraille dans les tumeurs conduisant à l'accumulation excessive de signaux positifs (protéines oncogéniques), ou à la dégradation anormale de régulateurs négatifs (suppresseurs de tumeurs). En conséquence, en l'absence de
5 suppresseurs de tumeur ou en présence d'un excès d'oncogènes, les cellules se multiplient sans cesse, formant des tumeurs (pour revue, voir Ciechanover, 1998). Ainsi, les dégradations anormales par la voie ubiquitine du suppresseur de tumeur p53 et de l'oncogène putatif β -caténine ont été corrélées au développement de tumeurs, suggérant que
10 certains gènes codant des enzymes d'ubiquitination pourraient être mutés dans les tumeurs.

La dégradation non contrôlée de certains régulateurs du cycle cellulaire a été observée dans certaines tumeurs, il est donc possible qu'une ubiquitine ligase fonctionnant mal soit la cause de ces
15 dégradations altérées de régulateurs du cycle. Ainsi, la surexpression de Mdm2, une ubiquitine ligase, induit une diminution notable de la quantité de son substrat, le suppresseur de tumeur p53.

Vu l'importance des voies de signalisation auxquelles ils participent, les complexes E3 constituent des cibles attractives de
20 criblage pharmacologique. D'ores et déjà, il a été établi que le dysfonctionnement de la voie de dégradation des protéines ubiquitine-protéasome est clairement impliqué dans de nombreuses pathologies de nature extrêmement variée : les cancers, les maladies génétiques, la maladie de parkinson, la maladie d'Alzheimer, les syndromes
25 inflammatoires et les infections virales.

La caractérisation d'agents chimiques capables de moduler l'activité de reconnaissance et d'ubiquitination des ubiquitine-ligases E3 est donc susceptible de permettre la mise au point de nouvelles
30 stratégies thérapeutiques humaines dans de nombreuses aires thérapeutiques comme le cancer, l'inflammation, les infections virales, bactériennes et fongiques, les syndromes inflammatoires, les pathologies cardio-vasculaires et neuro-dégénératives.

SOMMAIRE DE L'INVENTION

La présente invention a pour objet de nouvelles protéines à domaine HECT et de nouveaux substrats des protéines à domaine HECT. Elle a pour objet les méthodes de criblage mises au point pour identifier les substrats des nouvelles protéines à domaine HECT et pour identifier les petites molécules et drogues qui modulent l'interaction et/ou l'activité des protéines à domaine HECT et de leurs substrats. Les méthodes de criblage de cette invention peuvent être utilisées pour identifier des agents thérapeutiques qui pourraient être utilisés dans des protocoles et pour le traitement de diverses pathologies, incluant cancer, inflammations, pathologies cardio-vasculaires, maladies neuro-dégénératives, infections virales, bactériennes et fongiques. Cette invention englobe l'utilisation des nucléotides codant ces nouvelles protéines à domaine HECT, les protéines à domaine HECT et les peptides dérivants de ces protéines, les vecteurs permettant l'expression de ces protéines chez des bactéries, champignons, insectes, plantes et organismes mammifères ainsi que les animaux chez lesquels les gènes codant ces protéines à domaine HECT auront été inactivés.

Un premier objet de l'invention consiste en un acide nucléique codant pour un polypeptide à domaine HECT choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11, ou pour un fragment ou un variant du polypeptide choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11.

Selon l'invention, le fragment du polypeptide à domaine HECT comprend ou consiste en le peptide définissant le domaine HECT stricto sensu d'un polypeptide choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11.

Selon l'invention, un fragment du polypeptide à domaine HECT comprend ou consiste en un polypeptide choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11 dans lequel a été déléetée la séquence en acides aminés du domaine HECT stricto sensu.

Un acide nucléique tel que défini ci-dessus peut en outre être caractérisé en ce que le polypeptide à domaine HECT, un fragment ou

un variant de ce polypeptide est fusionné avec un polypeptide hétérologue, par exemple un marqueur détectable ou la protéine GAL4.

L'invention est également relative à des vecteurs recombinants, de clonage ou d'expression, contenant un acide nucléique selon l'invention,
5 le cas échéant sous le contrôle d'une séquence régulatrice permettant son expression dans une cellule hôte choisie.

Elle concerne aussi des cellules hôtes recombinantes dans lesquelles a été artificiellement inséré un acide nucléique ou un vecteur recombinant tel que défini ci-dessus.

10 L'invention a également pour objet un polypeptide à domaine HECT comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID N°1 à SEQ ID N°11, un fragment ou un variant de ce polypeptide, ainsi que des anticorps dirigés contre ce polypeptide ou encore le fragment ou le variant de celui-ci..

15 Elle est également relative à des procédés de criblages acellulaires ou cellulaires de composés candidats interagissant avec un polypeptide à domaine HECT tel que défini ci-dessus.

Elle concerne aussi des procédés de criblages de composés candidats modulant la force de l'interaction entre un polypeptide à
20 domaine HECT selon l'invention et une protéine substrat, ainsi que des procédés de criblage de composés candidats capables d'inhiber ou d'activer l'activité ubiquitine ligase de la protéine à domaine HECT spécifiée par le polypeptide considéré.

Elle a encore pour objet une composition pharmaceutique
25 comprenant une quantité thérapeutiquement efficace d'un acide nucléique ou d'un vecteur recombinant tel que défini ci-dessus, en association avec un ou plusieurs excipients physiologiquement compatibles..

Elle concerne aussi une composition pharmaceutique comprenant
30 une quantité thérapeutiquement efficace d'un polypeptide selon l'invention, en association avec un ou plusieurs excipients physiologiquement compatibles.

DEFINITIONS GENERALES

Selon l'invention, toute technique classique de biologie moléculaire, de microbiologie et d'ADN recombinant connue de l'homme du métier peut être utilisée. De telles techniques sont décrites par
5 exemple par SAMBROOK et al. (1989), GLOVER (1985), GAIT (1984), HAMES et HIGGINS (1984) et AUSUBEL et al. (1994) .

De manière préférée, tout acide nucléique et tout polypeptide selon l'invention se présentent sous une forme isolée ou purifiée.

Le terme " isolé " au sens de la présente invention désigne un
10 matériel biologique qui a été soustrait à son environnement originel (l'environnement dans lequel il est localisé naturellement). Par exemple, un polynucléotide présent à l'état naturel dans une plante n'est pas isolé. Le même polynucléotide séparé des acides nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de la plante est
15 isolé. Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeurer néanmoins à l'état isolé du fait que le vecteur ou la composition ne constitue pas son environnement naturel.

Le terme " purifié " ne nécessite pas que le matériel soit présent
20 sous une forme de pureté absolue, exclusive de la présence d'autres composés. Il s'agit plutôt d'une définition relative.

Un polynucléotide ou un polypeptide est à l'état purifié après purification du matériel de départ ou du matériel naturel d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement quatre ou
25 cinq ordres de grandeur.

Aux fins de la présente description, l'expression " séquence nucléotidique " peut être employée pour désigner indifféremment un polynucléotide ou un acide nucléique. L'expression " séquence nucléotidique " englobe le matériel génétique lui-même et n'est donc pas
30 restreinte à l'information concernant sa séquence.

Les termes " acide nucléique ", " polynucléotide ", " oligonucléotide " ou encore " séquence nucléotidique " englobent des séquences d'ARN, d'ADN, d'ADNc ou encore des séquences hybrides ARN/ADN de plus d'un nucléotide, indifféremment sous la forme simple
35 brin ou sous la forme de duplex.

Le terme "nucléotide" désigne à la fois les nucléotides naturels (A, T, G, C) ainsi que des nucléotides modifiés qui comprennent au moins une modification telle que (i) un analogue d'une purine, (ii) un analogue d'une pyrimidine, ou (iii) un sucre analogue, de tels nucléotides modifiés étant décrits par exemple dans la demande PCT N°WO 95/04064.

Aux fins de la présente invention, un premier polynucléotide est considéré comme étant "complémentaire" d'un second polynucléotide lorsque chaque base du premier nucléotide est appariée à la base complémentaire du second polynucléotide dont l'orientation est inversée. Les bases complémentaires sont A et T (ou A et U), et C et G.

Selon l'invention, un premier acide nucléique ayant au moins 95% d'identité avec un second acide nucléique de référence, possédera au moins 95%, de préférence au moins 96%, 97%, 98%, 98,5%, 99%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%, 99,7%, 99,8% ou 99,9% d'identité en nucléotides avec ce second polynucléotide de référence, le pourcentage d'identité entre deux séquences étant déterminé comme décrit ci-dessous.

Le "pourcentage d'identité" entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptidique dans la fenêtre de comparaison peut ainsi comprendre des additions ou des délétions (par exemple des "gaps") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

Le pourcentage est calculé en déterminant le nombre de positions auxquelles une base nucléique ou un résidu d'acide aminé identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auxquelles il y a identité entre les deux bases ou résidus d'acides aminés comparés, par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat par cent afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus.

De manière préférée, le pourcentage d'identité de séquences est déterminé à l'aide du logiciel BLAST (version BLAST 2.06 de Septembre 5 1998), en utilisant exclusivement les paramètres par défaut.

Un acide nucléique possédant au moins 95% d'identité en nucléotides avec un acide nucléique selon l'invention englobe les " variants " d'un acide nucléique selon l'invention.

Par " variant " d'un acide nucléique selon l'invention, on entend un 10 acide nucléique qui diffère de l'acide nucléique de référence par une ou plusieurs substitutions, additions ou délétions d'un nucléotide, par rapport à l'acide nucléique de référence. Un variant d'un acide nucléique selon l'invention peut être d'origine naturelle, tel qu'un variant allélique qui existe naturellement. Un tel acide nucléique variant peut être également 15 un acide nucléique non naturel obtenu, par exemple, par des techniques de mutagenèse.

En général, les différences entre l'acide nucléique de référence et l'acide nucléique " variant " sont réduites de telle sorte que l'acide nucléique de référence et l'acide nucléique variant ont des séquences 20 nucléotidiques très similaires et, dans de nombreuses régions, identiques. Les modifications nucléotidiques présentes dans un acide nucléique variant peuvent être silencieuses, ce qui signifie qu'elles n'affectent pas la séquence d'acides aminés qui peut être codée par cet acide nucléique variant.

25 Les modifications de nucléotides dans l'acide nucléique variant peuvent aussi résulter en des substitutions, additions ou délétions d'un ou plusieurs acides aminés dans la séquence du polypeptide qui peut être codé par cet acide nucléique variant.

De manière tout à fait préférée, un acide nucléique variant selon 30 l'invention comportant une phase de lecture ouverte, code pour un polypeptide qui conserve la même fonction ou la même activité biologique que le polypeptide codé par l'acide nucléique de référence.

De manière tout à fait préférée, un acide nucléique variant selon l'invention et qui comporte une phase de lecture ouverte, code pour un

polypeptide qui conserve la capacité d'être reconnu par des anticorps dirigés contre le polypeptide codé par l'acide nucléique de référence.

Par "fragment" d'un acide nucléique selon l'invention, on entend une séquence nucléotidique d'une longueur réduite par rapport à l'acide nucléique de référence, le fragment d'acide nucléique possédant une séquence nucléotidique identique à la séquence nucléotidique de l'acide nucléique de référence sur la partie commune. De tels fragments d'un acide nucléique selon l'invention possèdent au moins 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 60, 100, 150, 200, 300, 400, 500, 1000, 2000 ou 3000 nucléotides consécutifs de l'acide nucléique de référence, la longueur maximale en nucléotides d'un fragment d'un acide nucléique selon l'invention étant bien entendue limitée par la longueur maximale en nucléotides de l'acide nucléique de référence.

Par « fragment » d'un polypeptide à domaine HECT selon l'invention, on entend un fragment polypeptidique d'une longueur réduite par rapport au polypeptide de référence, le fragment polypeptidique possédant une séquence en acides aminés identique à la séquence en acides aminés du polypeptide de référence sur la partie commune. De tels fragments d'un polypeptide à domaine HECT selon l'invention possèdent au moins 5, 10, 15, 20, 25, 30, 40, 50, 60, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, ou 390 acides aminés consécutifs d'un polypeptide à domaine HECT de référence.

Par conditions d'hybridation de forte stringence, au sens de l'invention, on entend les conditions d'hybridation suivantes:

Préhybridation:

mêmes conditions que pour l'hybridation
durée: 1 nuit.

30

Hybridation:

5 x SSPE (0.9 M NaCl, 50 mM phosphate de sodium pH 7.7, 5 mM EDTA)
5 x Denhardt's (0.2% PVP, 0.2% Ficoll, 0.2% SAB)

100 µg/ml ADN de sperme de saumon

0.1% SDS

durée: 1 nuit.

Lavages:

- 5 2 x SSC, 0.1% SDS 10 min 65°C
 1 x SSC, 0.1 % SDS 10 min 65°C
 0.5 x SSC, 0.1 % SDS 10 min 65°C
 0.1 x SSC, 0.1% SDS 10 min 65°C.

- 10 Les paramètres définissant les conditions de stringence dépendent de la température à laquelle 50% des brins appariés se séparent (T_m).

Pour les séquences comprenant plus de 360 bases, T_m est définie par la relation:

- $T_m = 81,5 + 0,41 (\%G+C) + 16,6 \log(\text{concentration en cations}) -$
15 $0,63 (\% \text{ formamide}) - (600/\text{nombre de bases})$ (SAMBROOK et al., (1989), pages 9.54-9.62).

Pour les séquences de longueur inférieure à 30 bases, T_m est définie par la relation: $T_m = 4(G+C) + 2(A+T)$.

- Dans des conditions de stringence appropriées, dans lesquelles
20 les séquences aspécifiques n'hybrident pas, la température d'hybridation est approximativement de 5 à 30°C, de préférence de 5 à 10°C en-dessous de T_m .

- Les conditions d'hybridation ci-dessus décrites sont mises en œuvre pour l'hybridation d'un acide nucléique de 20 bases de longueur et
25 peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon les techniques connues de l'homme du métier.

- Les conditions convenables d'hybridation peuvent par exemple être adaptées selon l'enseignement contenu dans l'ouvrage de HAMES
30 et HIGGINS (1985) ou encore dans l'ouvrage de AUSUBEL et al. (1989).

En particulier, il faut prendre en compte que le niveau et la spécificité d'hybridation dépendent de différents paramètres, tels que :

- a) la pureté de la préparation de l'acide nucléique sur lequel la sonde ou l'amorce doit s'hybrider ;
- 5 b) la composition en base de la sonde ou de l'amorce, les paires de bases G-C possédant une plus grande stabilité thermique que les paires de bases A-T ou A-U ;
- c) la longueur de la séquence de bases homologue entre la sonde ou l'amorce et l'acide nucléique ;
- 10 d) la force ionique : le taux d'hybridation augmente avec l'accroissement de la force ionique et la durée du temps d'incubation ;
- e) la température d'incubation ;
- f) la concentration de l'acide nucléique sur lequel la sonde ou l'amorce doit s'hybrider ;
- 15 g) la présence d'agents dénaturants, tels que des agents favorisant la rupture de liaisons hydrogène , comme le formamide ou l'urée, qui accroissent la stringence de l'hybridation ;
- h) le temps d'incubation, le taux d'incubation augmentant avec la durée de l'incubation ;
- 20 i) la présence d'agents d'exclusion de volume , tels que le dextran ou le sulfate de dextran, qui augmentent le taux d'hybridation du fait qu'ils accroissent les concentrations effectives de la sonde ou de l'amorce et de l'acide nucléique qui doit s'hybrider, au sein de la préparation.

25 **BREVE DESCRIPTION DES FIGURES**

La Figure 1 illustre l'alignement des séquences d'acides aminés correspondant au domaine HECT des différentes protéines selon l'invention.

La figure 2 est une représentation schématique des protéines à
30 domaine HECT. Les domaines putatifs d'interaction protéine-protéine sont représentés. La double barre oblique indique que les ADNs complémentaires correspondant sont incomplets en 5' et/ou en 3'. Dans la légende, « HECT » signifie Homologous to E6-AP Carboxyl-Terminus, « WW » signifie domaine avec deux résidus Trp conservés, « C2 »
35 signifie motif de liaison de Ca^{2+} ou de certains types de lipides, «RCC1 »

signifie Regulator of Chromosome Condensation, « IQ » signifie petit motif de liaison de la calmoduline avec des résidus Ile et Gln et «RING » signifie un motif de type doigt de zinc.

5 **DESCRIPTION DETAILLEE DE L'INVENTION**

Il est fourni selon l'invention de nouveaux polypeptides à domaine HECT impliqués dans l'ubiquitination des protéines cellulaires, étape importante dans la voie de dégradation des protéines.

La caractérisation de ces nouveaux polypeptides à domaine HECT
10 selon l'invention rend possible la mise au point de nouveaux outils diagnostiques pour la détection d'altérations dans l'activité biologique des ubiquitine ligase à domaine HECT et de nouveaux outils thérapeutiques destinés à prévenir ou corriger ces altérations dans l'activité biologique des ubiquitine ligases à domaine HECT.

15 Les nouveaux polypeptides à domaine HECT selon l'invention sont les polypeptides suivants :

- le polypeptide HECT9 de séquence en acides aminés SEQ ID N°1 ;
- le polypeptide HECT17 de séquence en acides aminés SEQ ID N°2 ;
- le polypeptide HECT18 de séquence en acides aminés SEQ ID N°3 ;
- 20 - le polypeptide HECT19 de séquence en acides aminés SEQ ID N°4 ;
- le polypeptide HECT20 de séquence en acides aminés SEQ ID N°5 ;
- le polypeptide HECT21 de séquence en acides aminés SEQ ID N°6 ;
- le polypeptide HECT22 de séquence en acides aminés SEQ ID N°7 ;
- le polypeptide HECT23 de séquence en acides aminés SEQ ID N°8 ;
- 25 - le polypeptide HECT24 de séquence en acides aminés SEQ ID N°9 ;
- le polypeptide HECT25 de séquence en acides aminés SEQ ID N°10 ;
- le polypeptide HECT26 de séquence en acides aminés SEQ ID N°11 ;

Le polypeptide HECT9 possède une homologie de séquence avec
30 la protéine Pub1 de *S. pombe*.

Les polypeptides HECT17, HECT18, HECT19, HECT20, HECT21, HECT22, HECT23, HECT24, HECT25, HECT26 possèdent des homologies avec des protéines référencées dans la base de données SMART.

Un premier objet de l'invention consiste en un acide nucléique codant pour un polypeptide à domaine HECT choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11, ou pour un fragment ou un variant du polypeptide choisi parmi les séquences d'acides aminés comprenant les séquences SEQ ID N°1 à SEQ ID N°11.

Dans un mode de réalisation préféré de l'invention, le polypeptide à domaine HECT codé par un acide nucléique tel que défini ci-dessus consiste en une séquence d'acides aminés choisie parmi les séquences SEQ ID N°1 à SEQ ID N°11, ou pour un variant ou un fragment de ce polypeptide.

De préférence, un acide nucléique codant pour un polypeptide à domaine HECT selon l'invention comprend un polynucléotide choisi parmi les séquences nucléotidiques SEQ ID N°12 à SEQ ID N°22, ainsi que les séquences nucléotidiques qui en sont dérivées et qui comprennent le cadre de lecture ouvert contenu dans les séquences SEQ ID N°12 à SEQ ID N°22.

Ainsi, un autre objet de l'invention consiste en un acide nucléique comprenant une séquence nucléotidique choisie parmi les séquences suivantes :

- la séquence nucléotidique allant du nucléotide en position 21 jusqu'au nucléotide en position 4775 de la séquence SEQ ID N°12 (*hect9*);
- la séquence nucléotidique allant du nucléotide en position 1 jusqu'au nucléotide en position 4891 de la séquence SEQ ID N°13 (*hect17*);
- la séquence nucléotidique allant du nucléotide en position 2066 jusqu'au nucléotide en position 3754 de la séquence SEQ ID N°14 (*hect18*);
- la séquence nucléotidique allant du nucléotide en position 1 jusqu'au nucléotide en position 4860 de la séquence SEQ ID N°15 (*hect19*);
- la séquence nucléotidique allant du nucléotide en position 304 jusqu'au nucléotide en position 3552 de la séquence SEQ ID N°16 (*hect20*);
- la séquence nucléotidique allant du nucléotide en position 472 jusqu'au nucléotide en position 2940 de la séquence SEQ ID N°17 (*hect21*);

- la séquence nucléotidique allant du nucléotide en position 172 jusqu'au nucléotide en position 801 de la séquence SEQ ID N°18 (*hect22*);
- la séquence nucléotidique allant du nucléotide en position 201 jusqu'au nucléotide en position 1337 de la séquence SEQ ID N°19 (*hect23*);
- 5 - la séquence nucléotidique allant du nucléotide en position 496 jusqu'au nucléotide en position 3339 de la séquence SEQ ID N°20 (*hect24*);
- la séquence nucléotidique allant du nucléotide en position 184 jusqu'au nucléotide en position 4899 de la séquence SEQ ID N°21 (*hect25*);
- 10 - la séquence nucléotidique allant du nucléotide en position 120 jusqu'au nucléotide en position 2237 de la séquence SEQ ID N°22 (*hect26*);

Font également partie de l'invention les acides nucléiques suivants :

a) un acide nucléique codant pour un polypeptide ayant une
15 séquence en acides aminés choisie dans le groupe des séquences SEQ ID N°1 à SEQ ID N°11 ou un fragment ou un variant de ce polypeptide, éventuellement fusionné à un polypeptide hétérologue ;

b) un acide nucléique comprenant un polynucléotide choisi parmi les séquences SEQ ID N°12 à SEQ ID N°22, ou un fragment ou un
20 variant de ce dernier ;

c) un acide nucléique ayant au moins 95% d'identité en nucléotides avec un acide nucléique choisi dans le groupe constitué des séquences SEQ ID N°12 à SEQ ID N°22 ou un fragment ou un variant de ce dernier ;

25 d) un acide nucléique hybridant, dans des conditions d'hybridation de forte stringence, avec un acide nucléique de séquences SEQ ID N°12 à SEQ ID N°22, ou un fragment ou un variant de ce dernier.

L'invention concerne aussi un polypeptide comprenant le domaine
30 HECT d'un polypeptide à domaine HECT tel que défini ci-dessus. Elle est également relative à un polypeptide consistant en le domaine HECT d'un polypeptide selon l'invention.

Le domaine HECT d'un polypeptide selon l'invention est considéré comme un « fragment » d'un tel polypeptide aux fins de la présente
35 description.

Ainsi, fait également partie de l'invention un polypeptide comprenant ou consistant en une séquence d'acides aminés choisie parmi les séquences suivantes :

- a) la séquence allant de l'acide aminé en position 1248 jusqu'à l'acide aminé en position 1585 de la séquence SEQ ID N°1 ;
- b) la séquence allant de l'acide aminé en position 1259 jusqu'à l'acide aminé en position 1624 de la séquence SEQ ID N°2 ;
- c) la séquence allant de l'acide aminé en position 230 jusqu'à l'acide aminé en position 567 de la séquence SEQ ID N°3 ;
- d) la séquence allant de l'acide aminé en position 1153 jusqu'à l'acide aminé en position 1620 de la séquence SEQ ID N°4 ;
- e) la séquence allant de l'acide aminé en position 742 jusqu'à l'acide aminé en position 1083 de la séquence SEQ ID N°5 ;
- f) la séquence allant de l'acide aminé en position 481 jusqu'à l'acide aminé en position 823 de la séquence SEQ ID N°6 ;
- g) la séquence allant de l'acide aminé en position 1 jusqu'à l'acide aminé en position 206 de la séquence SEQ ID N°7 ;
- h) la séquence allant de l'acide aminé en position 48 jusqu'à l'acide aminé en position 372 de la séquence SEQ ID N°8 ;
- i) la séquence allant de l'acide aminé en position 618 jusqu'à l'acide aminé en position 947 de la séquence SEQ ID N°9 ;
- j) la séquence allant de l'acide aminé en position 1235 jusqu'à l'acide aminé en position 1572 de la séquence SEQ ID N°10 ;
- k) la séquence allant de l'acide aminé en position 373 jusqu'à l'acide aminé en position 694 de la séquence SEQ ID N°11 ;

Pour obtenir une délétion totale ou partielle de la boîte HECT, l'homme du métier peut, par exemple, mettre en œuvre une digestion enzymatique de l'acide nucléique codant le polypeptide considéré en utilisant les sites de restriction flanquants, par exemple selon la technique décrite par MARGOTTIN et al. (Margottin F. et al., 1998, Molecular cell, 1 : 565-574).

Les acides nucléiques codant un polypeptide à domaine HECT tel que défini ci-dessus peuvent être facilement obtenus par l'homme du métier, par exemple à partir d'un échantillon d'ADN humain ou encore à

partir d'une banque d'ADN humain, à l'aide d'amorces spécifiques de l'une des séquences SEQ ID N°12 à SEQ ID N°22, par exemple par amplification PCR.

Comme cela sera exposé plus loin dans la description, l'expression d'un variant d'un polypeptide à domaine HECT selon l'invention, dans lequel le domaine HECT a été complètement ou partiellement délétée, ou muté peut être recherchée dans le cas où l'on désire obtenir une diminution ou un blocage de la dégradation du substrat spécifique. Un tel variant d'un polypeptide à domaine HECT selon l'invention peut conserver sa capacité à interagir spécifiquement avec le substrat, mais n'est plus capable d'ubiquitiner ce substrat.

Selon un autre aspect, l'invention est donc aussi relative à un acide nucléique codant pour un variant d'un polypeptide à domaine HECT tel que défini ci-dessus, dans la séquence duquel tout ou partie de la séquence du domaine HECT a été délétée.

Par « délétion partielle » du domaine HECT, on entend la suppression d'au moins 2, de préférence d'au moins 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 30, 40, 50, 60, 70, 80, 90, 100 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340 ou 350 acides aminés consécutifs de la séquence du domaine HECT d'un polypeptide selon l'invention, la longueur de la délétion étant limitée par la longueur en acides aminés d'un domaine HECT dans chacun des polypeptides à domaine HECT selon l'invention.

Dans un mode de réalisation préféré, l'acide nucléique codant pour un variant d'un polypeptide à domaine HECT dans la séquence duquel la séquence du domaine HECT a été complètement ou partiellement délétée est choisi parmi les acides nucléiques suivants :

a) un polypeptide comprenant la séquence SEQ ID N°12 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 1248 jusqu'à l'acide aminé en position 1585 de la séquence SEQ ID N°12 ;

b) un polypeptide comprenant la séquence SEQ ID N°13 dans lequel a été totalement ou partiellement délétée la séquence allant de

l'acide aminé en position 1259 jusqu'à l'acide aminé en position 1624 de la séquence SEQ ID N°13 ;

c) un polypeptide comprenant la séquence SEQ ID N°14 dans lequel a été totalement ou partiellement délétée la séquence allant de
5 l'acide aminé en position 230 jusqu'à l'acide aminé en position 567 de la séquence SEQ ID N°14 ;

d) un polypeptide comprenant la séquence SEQ ID N°15 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 1153 jusqu'à l'acide aminé en position 1620 de
10 la séquence SEQ ID N°15 ;

e) un polypeptide comprenant la séquence SEQ ID N°16 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 742 jusqu'à l'acide aminé en position 1083 de la séquence SEQ ID N°16 ;

f) un polypeptide comprenant la séquence SEQ ID N°17 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 481 jusqu'à l'acide aminé en position 823 de la séquence SEQ ID N°17 ;
15

g) un polypeptide comprenant la séquence SEQ ID N°18 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 1 jusqu'à l'acide aminé en position 206 de la séquence SEQ ID N°18 ;
20

h) un polypeptide comprenant la séquence SEQ ID N°19 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 48 jusqu'à l'acide aminé en position 372 de la séquence SEQ ID N°19 ;
25

i) un polypeptide comprenant la séquence SEQ ID N°20 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 618 jusqu'à l'acide aminé en position 947 de la séquence SEQ ID N°20 ;
30

j) un polypeptide comprenant la séquence SEQ ID N°21 dans lequel a été totalement ou partiellement délétée la séquence allant de l'acide aminé en position 1235 jusqu'à l'acide aminé en position 1572 de la séquence SEQ ID N°21 ;

k) un polypeptide comprenant la séquence SEQ ID N°22 dans lequel a été totalement ou partiellement déléetée la séquence allant de l'acide aminé en position 373 jusqu'à l'acide aminé en position 694 de la séquence SEQ ID N°22 ;

5 Selon un autre mode de réalisation d'un acide nucléique selon l'invention, cet acide nucléique est caractérisé en ce qu'il code pour un polypeptide à domaine HECT, ou un variant ou un fragment de ce polypeptide, qui est fusionné avec un polypeptide hétérologue.

10 Par « polypeptide hétérologue », on entend une séquence d'acides aminés qui n'est pas exprimée naturellement par la cellule sous une forme chimiquement liée à un polypeptide à domaine HECT tel que défini ci-dessus.

15 Selon un premier mode de réalisation, le polypeptide hétérologue consiste en un polypeptide capable de se fixer sélectivement sur un support, tel que la glutathione S transférase ou GST, bine connue de l'homme du métier.

20 Selon un second mode de réalisation, le polypeptide hétérologue consiste en un polypeptide détectable, par exemple un polypeptide fluorescent comme les protéines GFP et YFP, bien connues de l'homme du métier.

25 Selon un troisième mode de réalisation, le polypeptide hétérologue consiste en un polypeptide accepteur ou donneur de fluorescence utilisable dans des systèmes de détection d'interactions par transfert d'énergie de fluorescence (ou FRET pour « Fluorescence Resonance Energy Transfer »).

30 Selon un quatrième mode de réalisation, le polypeptide hétérologue consiste en un premier partenaire protéique lequel, lorsqu'il interagit spécifiquement avec un second partenaire protéique, a la capacité de se fixer sur un motif d'acide nucléique régulant la transcription d'un gène rapporteur, pour une utilisation dans des systèmes de détection d'interactions du type « double-hybride » bien connus de l'homme du métier. Un tel polypeptide peut être, par exemple, la protéine GAL4 ou la protéine LexA.

Selon encore un autre aspect d'un acide nucléique selon l'invention, cet acide nucléique comprend un polynucléotide régulateur contrôlant l'expression du polypeptide à domaine HECT ou de son fragment ou variant, éventuellement fusionné à un polypeptide
5 hétérologue, dans une cellule hôte. Des exemples illustratifs de polynucléotides régulateurs préférés selon l'invention sont listés ci-dessous dans la partie descriptive des vecteurs recombinants préférés.

Fait aussi partie de l'invention tout acide nucléique dont la séquence est complémentaire à l'un des acides nucléiques tels que
10 définis dans la présente description.

Vecteurs recombinants

Tout acide nucléique selon l'invention peut être inséré dans un vecteur à des fins de clonage ou à des fins d'expression.

15 L'invention est également relative à un vecteur recombinant comprenant un acide nucléique codant pur un polypeptide à domaine HECT, un fragment ou un variant de ce polypeptide, éventuellement fusionné avec un polypeptide hétérologue et pour lequel ledit acide nucléique a été artificiellement inséré dans le vecteur.

20 Avantageusement, un tel vecteur recombinant comprendra un acide nucléique choisi parmi les acides nucléiques suivants :

a) un acide nucléique codant pour un polypeptide ayant une séquence en acides aminés choisie dans le groupe des séquences SEQ ID N°1 à SEQ ID N°11 ou un fragment ou un variant de ce polypeptide,
25 éventuellement fusionné à un polypeptide hétérologue ;

b) un acide nucléique comprenant un polynucléotide choisi parmi les séquences SEQ ID N°12 à SEQ ID N°22, ou un fragment ou un variant de ce dernier

c) un acide nucléique ayant au moins 95% d'identité en
30 nucléotides avec un acide nucléique choisi dans le groupe constitué des séquences SEQ ID N°12 à SEQ ID N°22 ou un fragment ou un variant de ce dernier ;

d) un acide nucléique hybridant, dans des conditions d'hybridation de forte stringence, avec un acide nucléique de séquences SEQ ID N°12 à SEQ ID N°22, ou un fragment ou un variant de ce dernier.

Par "vecteur" au sens de la présente invention on entendra une
5 molécule d'ADN ou d'ARN circulaire ou linéaire qui est indifféremment sous forme de simple brin ou double brin.

Selon un premier mode de réalisation, un vecteur recombinant selon l'invention est utilisé afin d'amplifier l'acide nucléique qui y est inséré après transformation ou transfection de l'hôte cellulaire désiré.

10 Selon un second mode de réalisation, il s'agit de vecteurs d'expression comprenant, outre un acide nucléique conforme à l'invention, des séquences régulatrices permettant d'en diriger la transcription et/ou la traduction.

Selon un mode de réalisation avantageux, un vecteur recombinant
15 selon l'invention comprendra notamment les éléments suivants :

(1) des éléments de régulation de l'expression de l'acide nucléique à insérer, tels que des promoteurs et des enhanceurs ;

(2) la séquence codante comprise dans l'acide nucléique conforme à l'invention à insérer dans un tel vecteur, ladite séquence codante étant
20 placée en phase avec les signaux de régulation décrits aux (1) ; et

(3) des séquences d'initiation et d'arrêt de la transcription appropriées.

En outre, les vecteurs recombinants selon l'invention pourront
25 inclure une ou plusieurs origines de réplication chez les hôtes cellulaires dans lesquels leur amplification ou leur expression est recherchée, des marqueurs ou des marqueurs de sélection.

A titre d'exemples, les promoteurs bactériens pourront être les promoteurs LacI, LacZ, les promoteurs de l'ARN polymérase du
30 bactériophage T3 ou T7, les promoteurs PR, ou PL du phage lambda.

Les promoteurs pour cellules eucaryotes comprendront le promoteur de la thymidine kinase du virus HSV ou encore le promoteur de la métallothionéine-L de souris.

De manière générale, pour le choix d'un promoteur adapté,
35 l'homme du métier pourra avantageusement se référer à l'ouvrage de

SAMBROOK et al. (1989) précité ou encore aux techniques décrites par FULLER et al. (1996).

Les vecteurs bactériens préférés selon l'invention sont par exemple les vecteurs pBR322(ATCC37017) ou encore des vecteurs tels que pAA223-3 (Pharmacia, Uppsala, Suède), et pGEM1 (Promega Biotech, Madison, WI, ETATS-UNIS).

On peut encore citer d'autres vecteurs commercialisés tels que les vecteurs pQE70, pQE60, pQE9 (Qiagen), psiX174, pBluescript SA, pNH8A, pNH16A, pNH18A, pNH46A, pWLNEO, pSV2CAT, pOG44, pXTI, pSG(Stratagene).

Il peut s'agir également de vecteurs de type *baculovirus* tel que le vecteur pVL1392/1393 (Pharmingen) utilisé pour transfecter les cellules de la lignée Sf9 (ATCC N°CRL 1711) dérivées de *Spodoptera frugiperda*.

Il peut encore s'agir de vecteurs adénoviraux tels que l'adénovirus humain de type 2 ou 5.

Un vecteur recombinant selon l'invention peut aussi être un vecteur rétroviral ou encore un vecteur adéno-associé (AAV). De tels vecteurs adéno-associés sont par exemple décrits par FLOTTE et al. (1992), SAMULSKI et al. (1989), ou encore McLAUGHLIN BA et al. (1996).

Cellules hôtes recombinantes

L'invention concerne aussi une cellule hôte recombinante qui a été artificiellement transfectée ou transformée par un acide nucléique ou par un vecteur recombinant tels que définis ci-dessus.

De préférence, une cellule hôte recombinante selon l'invention est une cellule eucaryote, et de manière tout à fait préférée une cellule hôte recombinante humaine.

Les cellules hôtes préférées selon l'invention sont par exemple les suivantes :

a) cellules hôtes procaryotes: souches d'*Escherichia coli* (souche DH5- α), de *Bacillus subtilis*, de *Salmonella typhimurium*, ou encore des souches d'espèces telles que *Pseudomonas*, *Streptomyces* et *Staphylococcus* ;

b) cellules hôtes eucaryotes: cellules HeLa (ATCC N°CCL2), cellules Cv 1 (ATCC N°CCL70), cellules COS (ATCC N°CRL 1650), cellules Sf-9 (ATCC N°CRL 1711), cellules CHO (ATCC N°CCL-61) ou encore cellules 3T3 (ATCC N°CRL-6361).

5

Polypeptides de l'invention

Selon un autre aspect, l'invention concerne un polypeptide comprenant une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°1 à SEQ ID N°11, ou un
10 fragment ou un variant de ce polypeptide, éventuellement fusionné avec un polypeptide hétérologue..

L'invention est relative à un polypeptide à domaine HECT, un fragment ou un variant de ce polypeptide, éventuellement fusionné avec un polypeptide hétérologue, caractérisé en ce qu'il est codé par un acide
15 nucléique tel que défini dans la présente description.

L'invention concerne aussi un polypeptide comprenant au moins 15 acides aminés consécutifs d'une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°1 à SEQ ID N°11, ou un fragment ou un variant de ce polypeptide, éventuellement
20 fusionné avec un polypeptide hétérologue.

L'invention est également relative à un polypeptide comprenant une séquence en acides aminés ayant au moins 95% d'identité en acides aminés avec une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°1 à SEQ ID N°11, ou un
25 fragment ou un variant de ce polypeptide.

De manière générale, les polypeptides selon la présente invention se présentent sous une forme isolée ou purifiée.

L'invention concerne aussi un polypeptide comprenant des modifications d'acides aminés de 1, 2, 3, 4, 5, 10 à 20 substitutions, additions ou délétions d'un acide aminé par rapport à la séquence en
30 acides aminés d'un polypeptide de séquences SEQ ID N°1 à SEQ ID N°11, ou encore d'un fragment ou d'un variant de ce dernier.

L'invention est également relative à un procédé pour la production de l'un des polypeptides de séquences SEQ ID N°1 à SEQ ID N°11 ou

d'un fragment peptidique ou d'un variant de ce dernier, ladite méthode comprenant les étapes de :

- a) insérer un acide nucléique codant pour ledit polypeptide dans un vecteur approprié ;
- 5 b) cultiver, dans un milieu de culture approprié, une cellule hôte préalablement transformée ou transfecter avec le vecteur recombinant de l'étape a) ;
- c) récupérer le milieu de culture conditionné ou lyser la cellule hôte, par exemple par sonication ou par choc osmotique ;
- 10 d) séparer et purifier à partir dudit milieu de culture ou encore à partir des lysats cellulaires obtenus à l'étape c), ledit polypeptide ;
- e) le cas échéant, caractériser le polypeptide recombinant produit.

Les peptides selon l'invention peuvent être caractérisés par fixation sur une colonne de chromatographie d'immunoaffinité sur laquelle les anticorps dirigés contre ce polypeptide ou contre un fragment
15 ou un variant de ce dernier ont été préalablement immobilisés.

Selon un autre aspect, un polypeptide recombinant selon l'invention peut être purifié par passage sur une série appropriée de colonnes de chromatographie, selon les méthodes connues de l'homme
20 de l'art.

Un polypeptide selon l'invention peut être également préparé par les techniques classiques de synthèse chimique indifféremment en solution homogène ou phase solide.

A titre illustratif, un polypeptide selon l'invention pourra être
25 préparé par la technique ou en solution homogène décrite par Houben Weyl (1974) ou encore la technique de synthèse en phase solide décrite par Merrifield (1965a; 1965b).

Font également partie de l'invention des polypeptides dits "homologues" à l'un quelconque des polypeptides de séquences d'acides aminés SEQ ID N°1 à SEQ ID N°11, ou de leurs fragments ou
30 variants.

De tels polypeptides homologues ont des séquences d'acides aminés possédant une ou plusieurs substitutions d'un acide aminé par un acide aminé équivalent, par rapport aux polypeptides de référence.

On entendra par acide aminé équivalent selon la présente invention, par exemple remplacement d'un résidu sous la forme L par un résidu sous la forme D ou encore le remplacement d'un acide glutamique (E) par un acide pyro-glutamique selon des techniques bien connues de l'homme du métier. A titre illustratif, la synthèse de peptide contenant au moins un résidu sous la forme D est décrite par KOCH (1977).

Selon un autre aspect, sont également considérés comme des acides aminés équivalents deux acides aminés appartenant à la même classe, c'est-à-dire deux acides aminés acide, basique, non polaire ou encore polaire non chargé.

Font également partie de l'invention des polypeptides comprenant au moins une liaison non peptidique telle qu'une liaison rétro-inverso (NHCO), une liaison carba (CH_2CH_2) ou encore une liaison cétométhylène ($\text{CO}-\text{CH}_2$).

De préférence, les polypeptides selon l'invention comprenant une ou plusieurs additions, délétions, substitutions d'au moins un acide aminé conserveront leur capacité à être reconnus par des anticorps dirigés contre les polypeptides non modifiés. Ces polypeptides conserveront également leur capacité à reconnaître leur substrat spécifique et/ou partenaires potentiels avec lesquels ils se lient.

Anticorps

Les polypeptides selon l'invention, en particulier les polypeptides de séquences en acides aminés SEQ ID N°1 à SEQ ID N°11 ou les fragments et les variants de ces derniers ainsi que les peptides homologues peuvent être utilisés pour la préparation d'anticorps.

Par " anticorps " au sens de la présente invention, on entendra notamment des anticorps polyclonaux ou monoclonaux ou des fragments (par exemple les fragments $\text{F}(\text{ab})'_2$, Fab) ou encore tout polypeptide comprenant un domaine de l'anticorps initial reconnaissant le polypeptide ou le fragment de polypeptide cible selon l'invention .

Des anticorps monoclonaux peuvent être préparés à partir d'hybridomes selon la technique décrite par KOHLER et MILSTEIN (1975).

La présente invention concerne également des anticorps dirigés contre un polypeptide tel que décrit ci-dessus ou un fragment ou un variant de ce dernier, tels que produits dans la technique du trioma ou encore la technique d'hybridome décrite par KOZBOR et al. (1983).

5

L'invention a également trait à des fragments d'anticorps simple chaîne Fv (ScFv) tels que décrits dans le brevet US N° 4,946,778 ou encore par MARTINEAU et al. (1998).

Les anticorps selon l'invention comprennent également des fragments d'anticorps obtenus à l'aide de banques de phages RIDDER et al., (1995) ou encore des anticorps humanisés REIMANN et al. (1997); LEGER et al., (1997).

Les préparations d'anticorps selon l'invention sont utiles dans des tests de détection immunologiques destinés à l'identification de la présence et/ou de la quantité d'antigènes présents dans un échantillon.

Un anticorps selon l'invention pourra comprendre en outre un marqueur détectable isotopique ou non-isotopique, par exemple fluorescent ou encore être couplé à une molécule telle que la biotine, selon des techniques bien connues de l'homme du métier.

Ainsi, la mention a en outre pour objet un procédé pour détecter la présence d'un polypeptide conforme à l'invention dans un échantillon, ledit procédé comprenant les étapes de :

- a) mettre en contact l'échantillon à tester avec un anticorps tel que décrit ci-dessus ;
- b) détecter le complexe antigène/anticorps formé.

25

L'invention est également relative à un nécessaire ou kit de diagnostic ou pour la détection de la présence d'un polypeptide conforme à l'invention dans un échantillon, ledit nécessaire comprenant :

- a) un anticorps tel que défini ci-dessus ;
- b) un réactif permettant la détection des complexes antigène/anticorps formés.

30

Sondes et amorces nucléotidiques

Les fragments d'acides nucléiques dérivés de l'une quelconque des séquences nucléotidiques SEQ ID N°12 à SEQ ID N°22 sont utiles pour la détection de la présence d'au moins une copie d'une séquence
5 nucléotidique choisie parmi les séquences SEQ ID N°12 à SEQ ID N°22 ou encore d'un fragment ou d'un variant de cette dernière dans un échantillon.

Les sondes ou les amorces nucléotidiques selon l'invention comprennent au moins huit nucléotides consécutifs d'un acide nucléique
10 choisi dans le groupe constitué des séquences SEQ ID N°12 à SEQ ID N°22, ou d'un acide nucléique de séquence complémentaire.

De préférence, des sondes ou amorces nucléotidiques selon l'invention auront une longueur de 10, 12, 15, 18 ou 20 à 25, 35, 40, 50, 70, 80, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide
15 nucléique selon l'invention, en particulier un acide nucléique de séquence nucléotidique choisie parmi les séquences SEQ ID N°12 à SEQ ID N°22 ou d'un acide nucléique de séquence complémentaire.

Alternativement, une sonde ou une amorce nucléotidique selon l'invention consistera et/ou comprendra les fragments d'une longueur de
20 12, 15, 18, 20, 25, 35, 40, 50, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide nucléique selon l'invention, plus particulièrement d'un acide nucléique choisi parmi les séquences SEQ ID N°12 à SEQ ID N°22, ou d'un acide nucléique de séquence complémentaire.

La définition d'une sonde et d'une amorce nucléotidique selon l'invention englobe donc des oligonucléotides qui hybrident, dans les
25 conditions d'hybridation de forte stringence définies ci-avant, avec un acide nucléique choisi parmi les séquences SEQ ID N°12 à SEQ ID N°22 ou avec une séquence complémentaire de ces derniers.

Une amorce ou une sonde nucléotidique selon l'invention peut être
30 préparée par toute méthode adaptée bien connue de l'homme du métier, y compris par clonage et action d'enzymes de restriction ou encore par synthèse chimique directe selon des techniques telles que la méthode au phosphodiester de NARANG et al. (1979) ou de BROWN et al. (1979), la méthode aux diéthylphosphoramidites de BEAUCAGE et al. (1980) ou

encore la technique sur support solide décrite dans le brevet EU N°EP 0 707 592.

Chacun des acides nucléiques selon l'invention, y compris les sondes et amorces oligonucléotidiques décrites ci-dessus, peuvent être
5 marqués, si désiré, en incorporant un marqueur détectable par des moyens spectroscopiques, photochimiques, biochimiques, immunochimiques ou encore chimiques.

Par exemple, de tels marqueurs peuvent consister en des isotopes radioactifs (^{32}P , ^{33}P , ^3H , ^{35}S ,), des molécules fluorescentes (5-
10 bromodeoxyuridine, fluorescéine, acétylaminofluorène, digoxigénine) ou encore des ligands tels que la biotine.

Le marquage des sondes est fait de préférence par incorporation de molécules marquées au sein des polynucléotides par extension d'amorces, ou bien par rajout sur les extrémités 5' ou 3'

15 Les sondes oligonucléotides selon l'invention peuvent être utilisées notamment dans des hybridations de type Southern à l'ADN génomique ou encore dans des hybridations à l'ARN messager correspondant lorsque l'expression du transcrit correspondant est recherchée dans un échantillon.

20 Les sondes selon l'invention peuvent aussi être utilisées pour la détection de produits d'amplification PCR ou encore pour la détection de mésappariements.

Des sondes ou amorces nucléotidiques selon l'invention peuvent être immobilisées sur un support solide. De tels supports solides sont
25 bien connus de l'homme du métier et comprennent des surfaces des puits de plaques de microtitration, des lits de polystyrène, des lits magnétiques, des bandes de nitrocellulose, ou encore des microparticules telles que des particules de latex.

En conséquence, la présente invention concerne également un
30 procédé de détection de la présence d'un acide nucléique tel que décrit ci-avant dans un échantillon, ladite méthode comprenant les étapes de :

1) mettre en contact une ou plusieurs sondes nucléotidiques selon l'invention avec l'échantillon à tester ;

2) détecter le complexe éventuellement formé entre la ou les
35 sondes et l'acide nucléique présent dans l'échantillon.

Selon un mode de réalisation particulier du procédé de détection selon l'invention, la ou les sondes oligonucléotidiques sont immobilisées sur un support.

5 Selon un autre aspect, les sondes oligonucléotidiques comprennent un marqueur détectable.

L'invention concerne en outre un nécessaire ou kit pour la détection de la présence d'un acide nucléique selon l'invention dans un échantillon, ledit nécessaire comprenant :

10 a) une ou plusieurs sondes nucléotidiques telles que décrites ci-dessus ;

b) le cas échéant, les réactifs nécessaires à la réaction d'hybridation.

 Selon un premier aspect, le nécessaire ou kit de détection est
15 caractérisé en ce que la ou les sondes sont immobilisées sur un support.

 Selon un second aspect, le nécessaire ou kit de détection est caractérisé en ce que les sondes oligonucléotidiques comprennent un marqueur détectable.

 Selon un mode de réalisation particulier du kit de détection décrit
20 ci-dessus, un tel kit comprendra une pluralité de sondes oligonucléotidiques conformes à l'invention qui pourront être utilisées pour détecter des séquences cibles d'intérêt ou alternativement détecter des mutations dans les régions codantes ou les régions non codantes des acides nucléiques selon l'invention, plus particulièrement des acides
25 nucléiques de séquences SEQ ID N°12 à SEQ ID N°22 ou les acides nucléiques de séquence complémentaire.

 Ainsi, les sondes selon l'invention immobilisées sur un support peuvent être ordonnées en matrices telles que les " puces à ADN ". De telles matrices ordonnées ont été en particulier décrites dans le brevet
30 US N° 5,143,854, dans les demandes PCT N° WO 90/150 70 et 92/10092.

 Des matrices supports sur lesquelles des sondes oligonucléotidiques ont été immobilisées à une haute densité sont par exemple décrites dans les brevets US N°5,412,087 et dans la demande
35 PCT N°WO 95/11995.

Les amorces nucléotidiques selon l'invention peuvent être utilisées pour amplifier l'un quelconque des acides nucléiques selon l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences SEQ ID N°12 à SEQ ID N°22, ou encore un variant de celui-ci.

5 Un autre objet de l'invention concerne un procédé pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement un acide nucléique de séquences SEQ ID N°12 à SEQ ID N°22 ou un fragment ou un variant de celui-ci contenu dans un échantillon, ledit procédé comprenant les étapes de :

10 a) mettre en contact l'échantillon dans lequel la présence de l'acide nucléique cible est suspectée avec une paire d'amorces nucléotidiques dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de la région de l'acide nucléique cible dont l'amplification est recherchée, en présence des réactifs nécessaires à la
15 réaction d'amplification ; et

b) détection des acides nucléiques amplifiés.

Pour mettre en oeuvre le procédé d'amplification tel que défini ci-dessus, on aura avantageusement recours à l'une quelconque des
20 amorces nucléotidiques décrites ci-avant.

L'invention a en outre pour objet un nécessaire ou kit pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences SEQ ID N°12 à SEQ ID N°22, ledit nécessaire ou kit comprenant :

25 a) un couple d'amorces nucléotidiques conformes à l'invention, dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de l'acide nucléique cible dont l'amplification est recherchée ;

b) le cas échéant, les réactifs nécessaires à la réaction d'amplification.

30

Un tel nécessaire ou kit d'amplification comprendra avantageusement au moins une paire d'amorces nucléotidiques telles que décrites ci-dessus.

35

Procédés de criblages

Identification des partenaires ou des cibles des protéines à domaine HECT

5 L'invention concerne aussi des procédés de criblage de composés candidats interagissant avec un polypeptide à domaine HECT tel que défini dans la présente description.

La première étape vers l'utilisation des protéines à domaine HECT comme cible thérapeutique est l'identification des substrats ciblés pour la
10 dégradation par ces composants. L'identification et la caractérisation de ces cibles peuvent notamment apporter des informations essentielles sur les mécanismes biologiques dans lesquels elles interviennent.

Les méthodes pour identifier les polypeptides qui interagissent avec les protéines à domaine HECT de la présente invention sont variées
15 et comprennent les expériences de double hybride chez la levure ainsi que les méthodes de phage display; ces méthodes présentent l'avantage que la séquence nucléique codant les peptides interagissant est identifiée simultanément, au contraire d'autres méthodes plus biochimiques qui nécessitent un séquençage protéique.

20 Le double hybride chez la levure peut être utilisé pour cribler une banque d'expression d'ADNc de mammifères (humaines), où un ADNc est fusionné au domaine de liaison à l'ADN ou au domaine activateur de GAL4, et l'acide nucléique codant le peptide d'intérêt, comme la protéine cible, est fusionné également soit au domaine activateur soit au domaine
25 de liaison à l'ADN de GAL4 respectivement (Gyuris et al., 1993). Le double hybride a été développé afin de détecter des interactions protéine-protéine spécifiques. Un des aspects de la présente invention est d'utiliser une protéine pour détecter des interactants. Cette protéine, qui peut être une protéine à domaine HECT (tronquée ou non), est
30 utilisée comme appât pour identifier un ou des substrats cible potentiels. Alternativement, des substrats cibles d'intérêt peuvent être utilisés pour identifier l'ubiquitine ligase ou le complexe ubiquitine ligase responsable de sa dégradation.

Le ou les polypeptides cibles peuvent également être identifiés par
35 la technologie de "phage display". Une banque de "phage display" est

une banque d'expression protéique construite dans un vecteur de type phagique qui exprime une collection de séquences protéiques clonées en fusion avec une protéine de la paroi du phage. Cette construction particulière permet l'expression de la protéine de fusion à l'extérieur de la particule phagique. Cette disposition favorise donc l'interaction entre la protéine recombinante exposée au niveau de la paroi et une protéine ligand immobilisée. La protéine à domaine HECT est par exemple immobilisée sur une matrice solide et une solution contenant les phages est passée sur ce support. On analyse alors les phages retenus et la nature de la protéine qu'ils expriment.

Les partenaires cible d'une protéine à domaine HECT pourront également être identifiés de la manière suivante. Un peptide correspondant à une protéine à domaine HECT concernée par la présente invention ayant subi ou non des modifications post-traductionnelles est immobilisé sur une résine, par exemple une résine activée NHS-sepharose® (pharmacia). Un lysat cellulaire, par exemple de cellules HELA ou d'autres cellules appropriées est fractionné sur la colonne d'affinité préparée selon une méthode standard. Après lavage, les protéines retenues sur la colonne sont éluées en augmentant la force ionique des lavages ou par élution compétitive avec des peptides représentant le domaine HECT en partie ou dans son intégralité. Les protéines cellulaires sont ensuite identifiées par une ou l'autre des combinaisons de techniques mentionnées ci-dessous : fractionnement sur gel SDS PAGE, suivi ou non d'un western-blot, une fois purifiées du gel et éventuellement hydrolysées partiellement les protéines pourront être caractérisées plus avant en utilisant le micro-séquencage peptidique ou l'analyse par MALDI-TOF.

Selon un autre aspect, l'invention concerne aussi un procédé pour cribler des composés candidats se liant à un polypeptide à domaine HECT, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mettre en contact un polypeptide à domaine HECT tel que défini ci-dessus avec au moins un composé candidat à tester ;
- b) détecter les complexes éventuellement formés entre ledit polypeptide et le ou les composé(s) candidat(s) .

Le procédé de criblage ci-dessus peut en outre être caractérisé en ce qu'il comprend l'étape additionnelle suivante :

c) caractériser le ou les composé(s) candidat(s) ayant formé un complexe avec le polypeptide à domaine HECT.

5 Selon un premier mode de réalisation particulier du procédé ci-dessus, le polypeptide à domaine HECT est préalablement immobilisé sur un support.

Selon un second mode de réalisation particulier, le ou les composé(s) candidat(s) sont initialement contenus dans le lysat cellulaire d'une culture de cellules eucaryotes, de préférence des cellules humaines.

Selon un troisième mode de réalisation, le composé candidat est le produit d'expression d'un phage constitutif d'une collection combinatoire de clones de phages (« phage display »).

15 L'invention est également relative à un procédé pour cribler des composés candidats se liant à un polypeptide à domaine HECT, caractérisé en ce qu'il comprend les étapes suivantes :

a) cultiver une cellule hôte recombinante telle que définie ci-dessus dans un milieu de culture approprié, de telle manière à ce que ladite cellule hôte produise un polypeptide à domaine HECT selon l'invention;

20 b) détecter les complexes éventuellement formés entre le polypeptide à domaine HECT et d'autres composés produits par la cellule hôte.

Selon un mode particulier de réalisation du procédé de criblage ci-dessus, ledit procédé est caractérisé en ce qu'il s'agit d'un système double-hybride dans lequel la cellule hôte transfectée à l'étape a) est transfectée avec un acide nucléique codant pour un polypeptide à domaine HECT, un fragment ou un variant de ce polypeptide qui est fusionné avec un polypeptide hétérologue, ledit polypeptide hétérologue consistant en un premier partenaire protéique lequel, lorsqu'il interagit spécifiquement avec un second partenaire protéique, a la capacité de se fixer sur un motif d'acide nucléique régulant la transcription d'un gène rapporteur, pour une utilisation dans des systèmes de détection d'interactions du type « double-hybride » bien connus de l'homme du

métier. Un tel polypeptide peut être, par exemple, la protéine GAL4 ou la protéine LexA.

Selon ce même mode de réalisation particulier du procédé de criblage ci-dessus, la cellule hôte recombinante est en outre co-transfectée avec un acide nucléique codant le composé candidat sous une forme fusionnée avec le second partenaire protéique mentionné ci-dessus. De cette manière, l'interaction entre le polypeptide à domaine HECT, ou son fragment ou son variant, et le composé candidat est détecté par visualisation de l'activation ou au contraire de l'inhibition du gène rapporteur également contenu dans la cellule hôte recombinante co-transfectée.

L'invention concerne aussi un composé se liant à un polypeptide à domaine HECT tel que décrit ci-dessus, caractérisé en ce qu'il est susceptible d'être obtenu par l'un des procédés de criblages ci-dessus.

Identification de composés qui modulent l'interaction des protéines à domaine HECT avec d'autres protéines

L'invention est également relative à des procédés de criblage de composés modulant l'interaction entre un polypeptide à domaine HECT et une protéine cible spécifique, ou alternativement entre un polypeptide à domaine HECT et un ou plusieurs partenaires protéiques éventuels.

Une fois qu'un substrat d'une protéine à domaine HECT ou qu'une protéine capable d'interagir avec cette protéine à domaine HECT est identifiée, la recherche de modulateurs de l'interaction peut alors être entreprise. La présente invention présente des méthodes pour identifier des drogues qui sont soit des agonistes soit des antagonistes de la fonction cellulaire normale des protéines à domaine HECT, ou du rôle des protéines à domaine HECT dans la pathogenèse de la prolifération et/ou différenciation cellulaire normale ou anormale provoquée par l'ubiquitination d'une protéine par un processus dépendant d'une protéine à domaine HECT. Les tests évaluent l'habilité d'un composé à moduler l'association et/ou l'ubiquitination d'une protéine cible (cellulaire ou virale) par une protéine à domaine HECT. De tels modulateurs peuvent être utilisés par exemple dans le traitement de désordres prolifératifs ou

différenciatifs, pour moduler l'apoptose et dans le traitement d'infections virales.

Une variété de méthodologies est applicable à ces études. Les méthodologies qui permettent l'ubiquitination d'une protéine cible par une
5 voie dépendante des protéines à domaine HECT incluent les systèmes acellulaires, utilisant par exemple des protéines purifiées ou des lysats, ainsi que des systèmes cellulaires utilisant des cellules intactes (cellules de mammifères, humaines, d'insecte ou de levure). Des expériences de simple liaison peuvent aussi être utilisées pour détecter les agents qui
10 empêchent l'ubiquitination en se liant par exemple au substrat ou à la protéine à domaine HECT. Les agents testés pour être des inhibiteurs peuvent être produits par des bactéries, des levures ou d'autres organismes (produits naturels) ou produits chimiquement (petites molécules nucléiques ou chimiques).

15 Dans de nombreux programmes de criblage de drogues qui testent des banques de composés, des criblages à haut débit (high throughput assays) sont préférables afin de maximiser le nombre de composés testés dans une même expérience.

20 **Tests dans des systèmes acellulaires**

Ces méthodes peuvent être utilisées pour identifier des composés capables d'interagir avec une protéine à domaine HECT ou un de ses partenaires pour modifier l'activité de la protéine à domaine HECT ou de son partenaire. Un tel composé peut par exemple modifier la structure
25 d'une protéine à domaine HECT ou de son partenaire et donc en affecter l'activité. Ces tests peuvent également être utilisés pour identifier des composés qui modulent l'interaction de la protéine à domaine HECT et d'un partenaire de cette protéine. Un tel système a été utilisé pour reconstituer, par exemple, la polyubiquitination de la protéine p53 par E6-
30 AP (Yabuki, N. et al., 1999).

Ces tests consistent essentiellement en un mélange réactionnel comprenant la protéine à domaine HECT, le composé à tester ou une banque de composés en présence ou non du partenaire de liaison. Les composés testés peuvent être un dérivé du partenaire de la protéine à
35 domaine HECT, par exemple un peptide cible inactif, ou une petite

molécule chimique ou nucléique (aptamères). Un exemple de criblage de cette invention pourrait consister à détecter la formation d'un complexe entre une protéine à domaine HECT de l'invention et un de ses partenaires en présence d'un composé ou d'une banque de composés.

5 Pour la détection, la molécule peut être marquée avec un marqueur spécifique, et le ou les composés(s) par un marqueur différent. L'interaction du composé avec la protéine à domaine HECT ou le partenaire de la protéine à domaine HECT peut être détecté en déterminant le niveau des deux marquages après une étape d'incubation et de lavage. Si les deux marquages sont présents cela signifie qu'il y a eu interaction.

10

Une interaction entre deux molécules peut également être déterminée en utilisant l'analyse d'interaction biomoléculaire (BIA) en temps réel qui détecte la résonance plasmonique de surface, un phénomène optique. La détection dépend des changements dans la concentration de masse des macromolécules sur l'interface, et ne requiert aucun marquage préalable. Une banque de composés à tester peut être immobilisée sur une surface sensible. Une solution contenant la protéine à domaine HECT ou un partenaire de cette protéine est alors versée en continue sur cette surface. Un changement dans l'angle de résonance détectable grâce à un enregistrement des signaux indique qu'une interaction s'est produite. Cette technique est décrite dans le livre BIA technology de Pharmacia.

15

20

Un autre exemple d'expérience de criblage de la présente invention inclut les étapes de: (a) réaliser un mélange réactionnel comprenant, une protéine à domaine HECT, un partenaire de cette protéine, et un composé à tester; (b) détecter l'interaction entre la protéine à domaine HECT et son partenaire. Cette protéine à domaine HECT et son partenaire peuvent être produites par une technologie recombinante, être purifiées ou synthétisées chimiquement. Un changement statistique significatif (activation ou inhibition) dans l'interaction de la protéine à domaine HECT et de son partenaire en présence d'un composé indique un agoniste potentiel (mimétique ou activateur) ou antagoniste (inhibiteur). L'efficacité du composé peut être

25

30

testée en effectuant des études de dose-réponse en utilisant différentes concentrations du composé.

La formation d'un complexe entre une protéine à domaine HECT et son partenaire peut être détectée par une variété de techniques. La modulation de la formation des complexes peut être quantifiée en utilisant par exemple des protéines marquées détectables grâce à un radiomarquage, un marquage fluorescent ou enzymatique.

Typiquement, il est souhaitable d'immobiliser la protéine à domaine HECT ou son partenaire pour faciliter la séparation des complexes des formes non complexées et faciliter l'automatisation de l'expérience. La liaison peut être réalisée dans des plaques de microtitration, des tubes à essais, des microtubes, toute vaisselle permettant de contenir les différents réactifs. Une protéine de fusion peut être produite qui permet la liaison de la, protéine à une matrice. Par exemple, des protéines de fusion à domaine HECT/GST (glutathione S transferase) peuvent être adsorbées sur des billes de sépharose qui sont ensuite mis en présence du partenaire marqué au ^{35}S et du composé testé. Ce mélange est incubé dans des conditions favorables à la formation du complexe, des conditions physiologiques en terme de sel et de pH par exemple. Après cette incubation, les billes sont lavées pour retirer tout matériel non fixé, la matrice immobilisée et la radioactivité déterminée directement (les billes sont placées dans du liquide à scintillation par exemple), ou dans le surnageant après dissociation des complexes. Les complexes peuvent également être dissociés de la matrice, séparés par une électrophorèse de type SDS-PAGE, et le niveau de protéine à domaine HECT ou de partenaire trouvé dans la fraction de bille quantifié sur gel en utilisant les techniques d'électrophorèse standard.

L'interaction entre deux molécules peut également être déterminée en utilisant la technique de FRET (Fluorescence resonance energy transfer). Le principe est basé sur le fait que le transfert d'énergie qui se produit entre deux fluorophores quand ils sont proches ($<10\text{nm}$), et que le spectre d'émission du premier fluorophore (le donneur) recouvre le spectre d'excitation du second (l'accepteur). Une association étroite entre deux fluorophores adaptés et l'excitation du donneur, se traduit donc par une augmentation de la fluorescence de l'accepteur et/ou une diminution

de la fluorescence du donneur par FRET (Herman, 1989). Cette technique peut notamment être appliquée à l'étude de l'interaction entre des protéines à domaine HECT et leurs substrats et être utilisée pour identifier des inhibiteurs ou des activateurs de l'interaction. Comme le

5 FRET est une méthode spectroscopique non destructive pour mesurer les interactions moléculaires, elle peut se faire sur des cellules vivantes.

L'invention est donc aussi relative à un procédé pour cribler des composés modulant l'interaction entre un polypeptide à boîte à domaine

10 HECT et un partenaire, comprenant les étapes suivantes :

- a) mettre en contact un polypeptide à domaine HECT tel que défini ci-dessus, une protéine partenaire et le composé candidat à tester ;
- b) détecter les complexes éventuellement formés entre le polypeptide à domaine HECT et la protéine partenaire.

15

Test dans des systèmes cellulaires

L'invention fournit également des méthodes expérimentales cellulaires pour identifier des agents agonistes ou antagonistes modulant l'activité des protéines à domaine HECT. L'effet d'un composé sur

20 l'expression d'un gène codant une protéine à domaine HECT peut être déterminé par des expériences de transfection en utilisant un gène rapporteur. Ce gène rapporteur peut être n'importe quel gène codant une protéine quantifiable, par exemple la luciférase ou le gène CAT. Le niveau d'expression de ce gène rapporteur est alors déterminé en

25 présence ou en l'absence du composé. Des tests cellulaires utilisables pour détecter des inhibiteurs des activités ubiquitine ligase ont déjà été décrits qui utilisent des protéines fusionnées à la GFP comme substrats (Dantuma, N., 2000). Une approche similaire a également été utilisée pour étudier l'ubiquitination de c-jun en utilisant une protéine de fusion c-

30 jun- β -galactosidase (Treier, M. et al., 1994). Le même principe pourra être utilisé pour étudier la protéolyse dépendante de l'ubiquitine de n'importe quelle protéine cellulaire en construisant la fusion appropriée.

Une autre application des méthodes cellulaires est le double hybride chez la levure (Gyuris et al., 1993), adapté à l'isolement de

35 peptides naturels (provenant par exemple d'une banque d'expression

d'ADNc), ou synthétiques(d'une banque d'expression de cadres de lecture au hasard), interagissant avec un polypeptide de l'invention. Ce polypeptide interagissant peut ensuite être utilisé pour détecter une augmentation ou une diminution de son interaction avec la protéine à
5 domaine HECT, permettant donc l'identification d'agonistes ou d'antagonistes.

Le système de conjugaison peut être effectué dans des cellules entières, tirant avantage des techniques de culture cellulaire. Par exemple, le système de conjugaison peut être constitué dans un système
10 de cellules eucaryotes en culture, y compris les cellules de mammifères et de levure. Les avantages d'effectuer l'expérimentation dans une cellule intacte incluent la possibilité de détecter des inhibiteurs qui sont fonctionnels dans un environnement plus proche de celui que l'utilisation thérapeutique de l'inhibiteur nécessitera, incluant la capacité de l'agent
15 de rentrer dans la cellule. De plus, certaines des expériences in vivo sont compatibles avec une analyse à grande échelle, comme cela est indiqué plus bas.

Les composants du système de conjugaison de l'ubiquitine peuvent être endogènes à la cellule sélectionnée pour effectuer
20 l'expérience. Autrement, certains ou tous les composants peuvent provenir de sources exogènes. Par exemple, des protéines de fusion peuvent être introduites dans la cellule par des techniques recombinantes (comme l'utilisation d'un vecteur d'expression), ou en microinjectant dans la cellule la protéine de fusion elle-même ou l'ARNm
25 codant cette protéine.

Dans tous les cas, la cellule n'est manipulée qu'après incubation avec un inhibiteur candidat afin de faciliter la détection de l'ubiquitination ou la dégradation de la cible. Comme décrit plus haut, pour des expériences réalisées avec des mélanges reconstitués de protéines ou
30 des lysats, l'efficacité du candidat inhibiteur est évaluée en mesurant des caractéristiques directes de la protéine cible, tel que le changement de poids moléculaire par des méthodes électrophorétiques, ou par détection dans un test de liaison. Dans ce cas, la cellule doit être lysée à la fin de l'incubation avec l'agent candidat, et le lysat manipulé dans une étape de
35 détection de la même manière que précédemment.

Dans ces expériences utilisant des cellules entières, l'utilisation d'un gène rapporteur peut également s'avérer judicieuse. Un gène rapporteur comprend tout gène qui exprime un produit détectable, sous forme d'ARN ou de protéine. Les gènes rapporteurs privilégiés sont ceux
5 qui sont détectables directement. Ce gène rapporteur peut être inclus dans une construction sous la forme d'un gène de fusion avec la protéine cible d'intérêt.

Parmi les exemples variés de gènes rapporteurs, il y a le gène CAT (chloramphenicol acetyl transferase) (Alton et Vapnek, 1979), la
10 luciférase, et d'autres systèmes de détection enzymatiques telles que la beta-galactosidase, la luciférase bactérienne, la phosphatase alcaline, etc. Le produit du gène rapporteur est détecté par une activité intrinsèque associée avec ce produit. par exemple, le gène rapporteur peut coder le produit d'un gène avec une activité détectable par la couleur, la
15 fluorescence, ou la luminescence. Le niveau d'expression du gène rapporteur est alors comparé au niveau d'expression dans la même cellule en l'absence du composé testé. La moindre différence statistique ou d'un autre ordre indique que le composé a d'une manière ou d'une autre altéré l'activité de la ligase.

20 L'invention a également pour objet un procédé pour cribler des composés modulant l'interaction entre un polypeptide à domaine HECT et une protéine partenaire, comprenant les étapes suivantes :

a) co-transfecter ou co-transformer une cellule hôte recombinante selon l'une des revendications 11 ou 12 avec un vecteur codant une
25 protéine partenaire, dans un système double-hybride ;

b) incuber la cellule hôte obtenue à l'étape a) avec le composé candidat à tester ;

c) détecter les modifications dans l'interaction entre le polypeptide à domaine HECT et la protéine partenaire.

30

L'invention est également relative à un composé modulant l'interaction entre un polypeptide à domaine HECT tel que défini ci-dessus et une protéine partenaire, caractérisé en ce qu'il est susceptible d'être obtenu selon le procédé de criblage ci-dessus.

L'invention fournit donc des moyens pour stabiliser la cible d'une ubiquitine ligase in vivo en empêchant son ubiquitination et donc sa dégradation. En d'autres termes, ces inhibiteurs empêchent l'interaction de l'ubiquitine ligase avec un polypeptide cible. L'inhibiteur est une petite
5 molécule. Le mécanisme d'inhibition peut être compétitif ou non compétitif, et le polypeptide cible peut être une protéine cellulaire, une protéine codée par un organisme pathogène.

Des inhibiteurs compétitifs efficaces peuvent par exemple empêcher l'interaction entre l'ubiquitine ligase et la protéine cible.

10 Des inhibiteurs non compétitifs peuvent comprendre par exemple des composés qui se lient à une région du polypeptide cible autre que la région d'interaction avec l'ubiquitine ligase et altèrent la structure de la région d'interaction. De tels non-compétiteurs allostériques présentent certains avantages, ils peuvent notamment fournir un très haut degré de
15 spécificité de l'inhibition car ils ne seront pas capables de se lier à d'autres cibles de l'ubiquitine ligase.

Les inhibiteurs compétitifs des ubiquitine ligase peuvent mimer la structure polypeptidique de la région d'interaction de l'ubiquitine ligase ou de la cible. De tels inhibiteurs empêchent le recrutement du polypeptide
20 cible par l'ubiquitine ligase. En fonction de la structure de l'inhibiteur, l'inhibition du recrutement peut être très spécifique de la cible ou affecter le recrutement d'autres polypeptides qui sont la cible de la même ubiquitine ligase. Par exemple, des inhibiteurs spécifiques peuvent mimer la structure WW de la séquence polypeptidique qui interagit avec la
25 protéine cible. Ces inhibiteurs compétitifs des répétitions WW se lient à la protéine cible, empêchant ainsi son interaction avec les répétitions WW d'une protéine à domaine HECT.

D'une manière générale, les molécules recherchées seront des molécules capables d'inhiber ou d'activer l'activité ubiquitine ligase vis à
30 vis de ses substrats grâce notamment à des essais fonctionnels permettant de visualiser in vivo ou in vitro cette activité ubiquitine ligase.

Compositions pharmaceutiques

Utilisation thérapeutique des protéines à domaine HECT, de leurs dérivés et de leurs modulateurs

5 L'utilisation thérapeutique d'une protéine à domaine HECT ou d'un de ses dérivés peptidiques peut également être envisagée. Ainsi, les composés ou les méthodes qui augmentent l'activité d'une protéine à domaine HECT peuvent être utilisés dans le traitement de désordres prolifératifs. Les symptômes d'un cancer pourraient ainsi être atténués
10 par l'utilisation d'un composé qui stimule l'activité d'une protéine à domaine HECT. Ainsi, une baisse de l'activité ou d'expression d'une protéine à domaine HECT dont le substrat est un régulateur positif du cycle cellulaire, comme un membre de la famille des cyclines, se traduira par une augmentation de la prolifération cellulaire. Dans ce cas, une
15 augmentation du niveau d'expression ou d'activité de la protéine à domaine HECT pourrait permettre l'amélioration des symptômes

De la même façon, des composés qui diminuent ou inactivent une protéine à domaine HECT, peuvent être utilisés en thérapie pour améliorer les symptômes d'une maladie proliférative. Ainsi, si une
20 protéine à domaine HECT responsable normalement de la dégradation d'un inhibiteur du cycle cellulaire devient trop active suite à une mutation, l'utilisation d'une drogue qui en limite l'activité pourrait s'avérer utile pour limiter certains symptômes de la maladie et notamment la prolifération cellulaire. Dans les cas où une diminution de la dégradation du substrat est recherchée, l'expression d'un mutant dit « Δ HECT » de la protéine à
25 domaine HECT est également une stratégie de choix pour bloquer la dégradation du substrat. Un mutant « Δ HECT » est un dérivé de la protéine à domaine HECT ne comprenant plus le motif HECT qui aura été délété. Un tel dérivé est capable d'interagir avec le substrat mais
30 n'est plus capable de l'ubiquitiner. Dans ce cas, la surproduction du mutant « Δ HECT » empêche l'ubiquitination et la dégradation concomitante de ce substrat. On parle de stabilisation par protection (ou par "trapping").

Les agents thérapeutiques modulant l'activité des protéines à
35 domaine HECT pourront être des acides nucléiques. Ils pourront agir soit

d'eux-mêmes comme par exemple des aptamères modulant une interaction, soit par l'intermédiaire de la protéine qu'ils codent comme par exemple des séquences codant pour un peptide servant d'inhibiteur compétitif. Il pourra également s'agir d'acides nucléiques non naturels
5 comprenant des nucléotides modifiés. Dans tous les cas, ils devront être apportés dans la cellule par un vecteur approprié dont la nature sera fonction des progrès des techniques de vectorisation. Il pourra s'agir par exemple de vecteurs d'expression et de pénétration viraux ou bactériens, de système de dispersion colloïdal ou de liposomes. Dans le cas d'acides
10 nucléiques devant être exprimés, ils devront comprendre outre la séquence codante, toutes les séquences permettant l'expression dans les cellules hôtes visées.

Les vecteurs viraux qui peuvent être utilisés pour introduire une séquence d'acide nucléique dans les cellules cibles d'un patient
15 comprennent, sans y être limité, des dérivés des virus de la vaccine, de l'herpes ou de rétrovirus ou de l'adénovirus. Dans tous les cas, les vecteurs utilisés seront défectifs, c'est à dire non pathogène.

Les autres moyens d'introduire les acides nucléiques dans les cellules du patient sont les systèmes de dispersion colloïdal. Ceux ci
20 incluent des complexes macromoléculaires, des nanocapsules, des microsphères, des billes, des systèmes mettant en jeux des émulsions de lipides dans l'eau, des micelles, et des liposomes. Dans le cas des liposomes par exemple, les acides nucléiques sont encapsulés dans une vésicule de membrane artificielle. Introduits dans l'organisme les
25 liposomes fusionnent avec les cibles du patient et délivrent ainsi les acides nucléiques. Préférentiellement ces liposomes fusionneront préférentiellement avec des cellules spécifiques.

L'invention couvre les plantes ou animaux transgéniques contenant une ou plusieurs des protéines à domaine HECT présentées
30 ici. Ces organismes transgéniques peuvent être obtenus en utilisant la séquence d'acide nucléique de l'invention. Ces organismes transgéniques peuvent être utilisés pour exprimer une protéine à domaine HECT entière ou tronquée, ou pour inactiver une protéine à domaine HECT (knock-out). Ces derniers organismes (souris
35 transgéniques par exemple), mutants pour une protéine à domaine HECT,

peuvent notamment être utilisés pour la recherche de drogues qui atténuent ou suppriment les effets de l'absence de la protéine.

Les méthodes pour obtenir de tels organismes transgéniques sont bien connues. Les souris knock-out par exemple sont obtenues par
5 intégration homologue d'une construction dans une cellule souche embryonnaire de souris (cellule ES), construction qui code le gène à inactiver. La séquence de la construction est modifiée et idéalement permet l'intégration conjointe d'un marqueur de sélection positif (un gène de résistance à la néomycine par exemple) au locus du gène à inactiver.
10 Ces changements peuvent alors être introduits dans la lignée germinale de la souris pour générer des individus chimères. (pour détails voir, Doetschman et al., 1985)

L'invention a en outre pour objet une composition pharmaceutique comprenant une quantité thérapeutiquement efficace d'un acide
15 nucléique selon l'invention ou d'un vecteur recombinant tel que défini ci-dessus, en association avec un ou plusieurs excipients physiologiquement compatibles.

L'invention est également relative à une composition pharmaceutique comprenant une quantité efficace d'un polypeptide à
20 domaine HECT, un fragment ou un variant de ce polypeptide, en association avec un ou plusieurs excipients physiologiquement compatibles.

Utilisation diagnostique des protéines à domaine HECT

25 Les régulateurs du cycle cellulaire sont souvent les produits des oncogènes ou des suppresseurs de tumeur. Les protéines à domaine HECT, en tant qu'ubiquitine ligases, pourraient en conséquence être les produits des oncogènes ou de gènes suppresseurs de tumeurs, en fonction des protéines régulatrices du cycle cellulaire dont elles régulent
30 l'abondance. Les protéines à domaine HECT, ses analogues, dérivés et par voie de conséquence les séquences nucléiques des protéines à domaine HECT, ou les anticorps anti-protéines à domaine HECT peuvent donc avoir leur utilité dans le diagnostique. Les séquences nucléiques des nouvelles protéines à domaine HECT peuvent être utilisées pour
35 détecter, pronostiquer ou diagnostiquer des désordres tels que la

tumorigénèse, les carcinomes, les adénomes, etc., par la recherche d'éventuelles mutations par exemple (délétions, changements dans la séquence,...). L'activité des ces ubiquitine ligases peut également être testée dans différentes pathologies afin de détecter un éventuel
5 changement dans leur niveau d'expression et leur possible implication dans la pathologie considérée.

Les molécules de l'invention peuvent ainsi être utilisées dans des tests immunologiques utilisant des techniques tels que les immunoblots, l'immunohistochimie par radioisotope, les test ELISA (enzyme linked
10 immunosorbent assay), l'immunoprécipitation, les test d'immunodiffusion, les tests d'agglutination pour en nommer certains.

Les gènes des protéines à domaine HECT, leur séquence nucléique directe ou complémentaire peuvent aussi être utilisés dans des tests d'hybridation dans le but de détecter, de diagnostiquer des
15 désordres ou maladies associés avec des changements dans l'expression ou l'activité des protéines à domaine HECT. Ces tests d'hybridation sont effectués en mettant en présence un échantillon contenant des acides nucléiques (comme une coupe de tissu) avec une sonde nucléique capable de s'hybrider à de l'ARN ou de l'ADN de
20 protéine à domaine HECT dans des conditions favorables à l'hybridation.

Des maladies ou des désordres induisant la surprolifération cellulaire peuvent être diagnostiqués ou détectés (ou leur prédisposition) en recherchant des baisses d'expressions d'une protéine à domaine HECT ou de son ARN, ou de son activité (par exemple, l'activité
25 ubiquitine ligase), ou en détectant des mutations dans l'ARN, l'ADN ou la séquence protéique d'une protéine à domaine HECT (par exemple, délétions, translocations) qui causent une diminution de l'expression ou de l'activité. Le niveau de protéine à domaine HECT peut être détecté par immunochimie, le niveau d'ARN par hybridation (northern blot,
30 hybridation in situ). L'activité de la protéine à domaine HECT, peut être détectée en mesurant l'activité ubiquitine ligase E3. Les translocations, délétions et mutations ponctuelles dans les séquences nucléiques des protéines à domaine HECT peuvent être détectées par Southern blot, FISH, analyse RFLP, SSCP, PCR, séquençage, etc.

REFERENCES

- AUSUBEL et T al., 1989. Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.
- 5 AUSUBEL FM, BRENT R, KINGSTON RE, MOORE DD, SEIDMAN JG, SMITH JA, STRUHL K Editors, *Current protocols in Molecular Biology*, Wiley Interscience. (1994).
- Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862
- BERBAL, 1984 (A CHERCHER S.V.P.).
- 10 Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;**68**:109-151
- Flotte et al., 1992, *Am. J. Respir. Cell Mol. Biol.*, **7** : 349-356.
- Fuller S.A. et al., 1996, *Immunology in Current Protocols in Molecular Biology*, Ausubel et al.
- 15 GAIT (ed.), (1984). Nucleic Acid Hybridization.
- GLOVER (ed.), 1985. *DNA Cloning: A Practical Approach*, Volumes I and II Oligonucleotide Synthesis, MRL Press, Ltd., Oxford, U.K.
- HAMES and HIGGINS, 1985. Nucleic Acid Hybridization: a practical approach, Hames & Higgins Ed. IRL Press, Oxford.
- 20 Houben Weyl, 1974, in *Meuthode der Organischen Chemie*, E. Wunsch Ed., Volume 15-I et 15-II,
- Koch Y., 1977, *Biochem. Biophys. Res. Commun.*, **74**:488-491
- Kohler G. and Milstein C., 1975, *Nature*, **256** : 495.
- Kozbor et al., 1983, *Hybridoma*, **2**(1):7-16.
- 25 Leger OJ, et al., 1997, *Hum Antibodies*, **8**(1): 3-16
- Martineau P, Jones P, Winter G, 1998, *J Mol Biol*, **280**(1):117-127•
- McLaughlin BA et al., 1996, *Am. J. Hum. Genet.*, **59** : 561-569.

- Merrifield RB, 1965a, *Nature*, **207**(996): 522-523.
- Merrifield RB., 1965b, *Science*, **150**(693): 178-185.
- Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;**68**:90-98•.
- Reimann KA, et al., 1997, *AIDS Res Hum Retroviruses*. **13**(11): 933-943
- 5 Ridder R, Schmitz R, Legay F, Gram H, 1995, *Biotechnology (N Y)*,
13(3):255-260
- SAMBROOK, J. FRITSCH, E. F., and T. Maniatis, 1989. *Molecular cloning: a laboratory manual*. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York.
- 10 Samulski et al., 1989, *J. Virol.*, **63** : 3822-3828.

REVENDICATIONS

1. Acide nucléique codant pour un polypeptide à domaine HECT
choisi parmi les séquences d'acides aminés comprenant les séquences
5 SEQ ID N°1 à SEQ ID N°11, ou pour un fragment ou un variant du
polypeptide choisi parmi les séquences d'acides aminés comprenant les
séquences SEQ ID N°1 à SEQ ID N°11.

2. Acide nucléique selon la revendication 1, caractérisé en ce qu'il
10 code pour un polypeptide choisi parmi les séquences d'acides aminés
SEQ ID N°1 à SEQ ID N°11 ou pour un fragment ou un variant de ce
polypeptide.

3. Acide nucléique selon l'une des revendications 1 ou 2,
15 caractérisé en ce qu'il comprend un polynucléotide choisi parmi les
séquences nucléotidiques SEQ ID N° 12 à SEQ ID N°22, ainsi que les
séquences nucléotidiques qui en sont dérivées et qui comprennent le
cadre de lecture ouvert contenu dans les séquences SEQ ID N°12 à SEQ
ID N°22.

20

4. Acide nucléique selon l'une des revendications 1 ou 2,
caractérisé en ce que le fragment du polypeptide à domaine HECT est
choisi parmi les séquences en acides aminés suivantes :

a) la séquence allant de l'acide aminé en position 1248 jusqu'à
25 l'acide aminé en position 1585 de la séquence SEQ ID N°1 ;

b) la séquence allant de l'acide aminé en position 1259 jusqu'à
l'acide aminé en position 1624 de la séquence SEQ ID N°2 ;

c) la séquence allant de l'acide aminé en position 230 jusqu'à
l'acide aminé en position 567 de la séquence SEQ ID N°3 ;

30 d) la séquence allant de l'acide aminé en position 1153 jusqu'à
l'acide aminé en position 1620 de la séquence SEQ ID N°4 ;

e) la séquence allant de l'acide aminé en position 742 jusqu'à l'acide aminé en position 1083 de la séquence SEQ ID N°5 ;

f) la séquence allant de l'acide aminé en position 481 jusqu'à l'acide aminé en position 823 de la séquence SEQ ID N°6 ;

5 g) la séquence allant de l'acide aminé en position 1 jusqu'à l'acide aminé en position 206 de la séquence SEQ ID N°7 ;

h) la séquence allant de l'acide aminé en position 48 jusqu'à l'acide aminé en position 372 de la séquence SEQ ID N°8 ;

10 i) la séquence allant de l'acide aminé en position 618 jusqu'à l'acide aminé en position 947 de la séquence SEQ ID N°9 ;

j) la séquence allant de l'acide aminé en position 1235 jusqu'à l'acide aminé en position 1572 de la séquence SEQ ID N°10 ;

k) la séquence allant de l'acide aminé en position 373 jusqu'à l'acide aminé en position 694 de la séquence SEQ ID N°11 ;

15

5. Acide nucléique selon l'une des revendications 1 ou 2, caractérisé en ce que le variant du polypeptide à domaine HECT est un polypeptide à domaine HECT comprenant une séquence d'acides aminés choisie parmi les séquences SEQ ID N°1 à SEQ ID N°11 et dans
20 la séquence duquel la séquence du domaine HECT est totalement ou partiellement déléetée.

6. Acide nucléique selon l'une des revendications 1 à 5, caractérisé en ce que le polypeptide à domaine HECT ou le fragment ou
25 le variant du polypeptide à domaine HECT est fusionné avec un polypeptide hétérologue.

7. Acide nucléique selon la revendication 6, caractérisé en ce que le polypeptide hétérologue consiste en un marqueur détectable.

30

8. Acide nucléique selon la revendication 6, caractérisé en ce que le polypeptide hétérologue consiste en un premier partenaire protéique lequel, lorsqu'il interagit spécifiquement avec un second partenaire

protéique, a la capacité de se fixer sur un motif d'acide nucléique régulant la transcription d'un gène rapporteur, pour une utilisation dans des systèmes de détection d'interactions du type « double-hybride ».

5 9. Acide nucléique selon l'une des revendications 1 à 8, caractérisé en ce qu'il comprend un polynucléotide régulateur contrôlant l'expression du polypeptide à domaine HECT ou de son fragment ou variant, éventuellement fusionné à un polypeptide hétérologue, dans une cellule hôte.

10

10. Vecteur recombinant dans lequel a été artificiellement inséré un acide nucléique selon l'une des revendications 1 à 9.

11. Cellule hôte recombinante qui a été artificiellement transfectée
15 ou transformée par un acide nucléique selon l'une des revendications 1 à 9 ou par un vecteur recombinant selon la revendication 10.

12. Cellule hôte selon la revendication 11, caractérisée en ce qu'elle est une cellule eucaryote.

20

13. Polypeptide à domaine HECT, ou un fragment ou un variant de ce polypeptide, le cas échéant fusionné avec un polypeptide hétérologue, caractérisé en ce qu'il est codé par un acide nucléique selon l'une des revendications 1 à 9.

25

14. Polypeptide selon la revendication 13, caractérisé en ce qu'il comprend une séquence d'acides aminés choisie parmi les séquences SEQ ID N°1 à SEQ ID N°11.

15. Procédé pour cribler des composés candidats se liant à un polypeptide à domaine HECT, caractérisé en ce qu'il comprend les étapes suivantes :

- 5 a) mettre en contact un polypeptide selon l'une des revendications 13 ou 14 avec au moins un composé candidat à tester ;
b) détecter les complexes éventuellement formés entre ledit polypeptide et le ou les composé(s) candidat(s) .

10 16. Procédé selon la revendication 15, caractérisé en ce qu'il comprend en outre l'étape suivante :

- c) caractériser le ou les composé(s) candidat(s) ayant formé un complexe avec le polypeptide à domaine HECT.

15 17. Procédé selon l'une des revendications 15 ou 16, caractérisé en ce que le polypeptide à domaine HECT est immobilisé sur un support.

20 18. Procédé selon l'une des revendications 15 à 17, caractérisé en ce que le ou les composé(s) candidat(s) sont contenus dans le lysat cellulaire d'une culture de cellules eucaryotes, de préférence de cellules humaines.

25 19. Procédé selon l'une des revendications 15 à 17, caractérisé en ce que le composé candidat est le produit d'expression d'un phage constitutif d'une collection combinatoire de clones de phages.

20. Procédé pour cribler des composés candidats se liant à un polypeptide à domaine HECT, caractérisé en ce qu'il comprend les étapes suivantes :

- 30 a) cultiver une cellule hôte recombinante selon l'une des revendications 11 ou 12 dans un milieu de culture approprié, de telle manière à ce que ladite cellule hôte produise un polypeptide à domaine HECT selon l'une des revendications 13 ou 14 ;

b) détecter les complexes éventuellement formés entre le polypeptide à domaine HECT et d'autres composés produits par la cellule hôte.

21. Procédé selon la revendication 20, caractérisé en ce qu'il s'agit d'un système double-hybride dans lequel :

5 (i) la cellule hôte cultivée à l'étape a) est transfectée avec un acide nucléique codant pour un polypeptide à domaine HECT, un fragment ou un variant de ce polypeptide qui est fusionné avec un polypeptide hétérologue, ledit polypeptide hétérologue consistant en un premier partenaire protéique lequel, lorsqu'il interagit spécifiquement avec un
10 second partenaire protéique, a la capacité de se fixer sur un motif d'acide nucléique régulant la transcription d'un gène rapporteur, et

(ii) la cellule hôte recombinante est en outre co-transfectée avec un acide nucléique codant le composé candidat sous une forme fusionnée avec le second partenaire protéique mentionné en (i).
15

22. Procédé pour cribler des composés modulant l'interaction entre un polypeptide à domaine HECT et une protéine cible ou un partenaire, comprenant les étapes suivantes :

a) mettre en contact un polypeptide à domaine HECT selon l'une
20 des revendications 13 ou 14, une protéine cible ou un partenaire et le composé candidat à tester ;

b) détecter les complexes éventuellement formés entre le polypeptide à domaine HECT et la protéine cible ou le partenaire.

23. Procédé pour cribler des composés modulant l'interaction entre un polypeptide à domaine HECT et une protéine cible ou un partenaire, comprenant les étapes suivantes :

a) co-transfecter ou co-transformer une cellule hôte recombinante selon l'une des revendications 11 ou 12 avec un vecteur codant une
30 protéine cible ou un partenaire, dans un système double-hybride ;

b) incuber la cellule hôte obtenue à l'étape a) avec le composé candidat à tester ;

c) détecter les modifications dans l'interaction entre le polypeptide à domaine HECT et la protéine cible ou un partenaire.
35

24. Anticorps dirigé contre un polypeptide selon l'une des revendications 13 ou 14.

25. Amorce nucléotidique hybridant avec un acide nucléique selon
5 l'une des revendications 1 à 9.

26. Amorce nucléotidique selon la revendication 27, caractérisée en ce qu'elle est choisie parmi les séquences SEQ ID N°12 à SEQ ID N°22.

10

27. Sonde nucléotidique hybridant avec un acide nucléique selon l'une des revendications 1 à 9.

28. Composition pharmaceutique comprenant une quantité
15 thérapeutiquement efficace d'un acide nucléique selon l'une des revendications 1 à 9 ou d'un vecteur recombinant selon l'une des revendications 11 ou 12, en association avec un ou plusieurs excipients physiologiquement compatibles.

20 29. Composition pharmaceutique comprenant une quantité thérapeutiquement efficace d'un polypeptide selon l'une des revendications 13 ou 14, en association avec un ou plusieurs excipients physiologiquement compatibles.

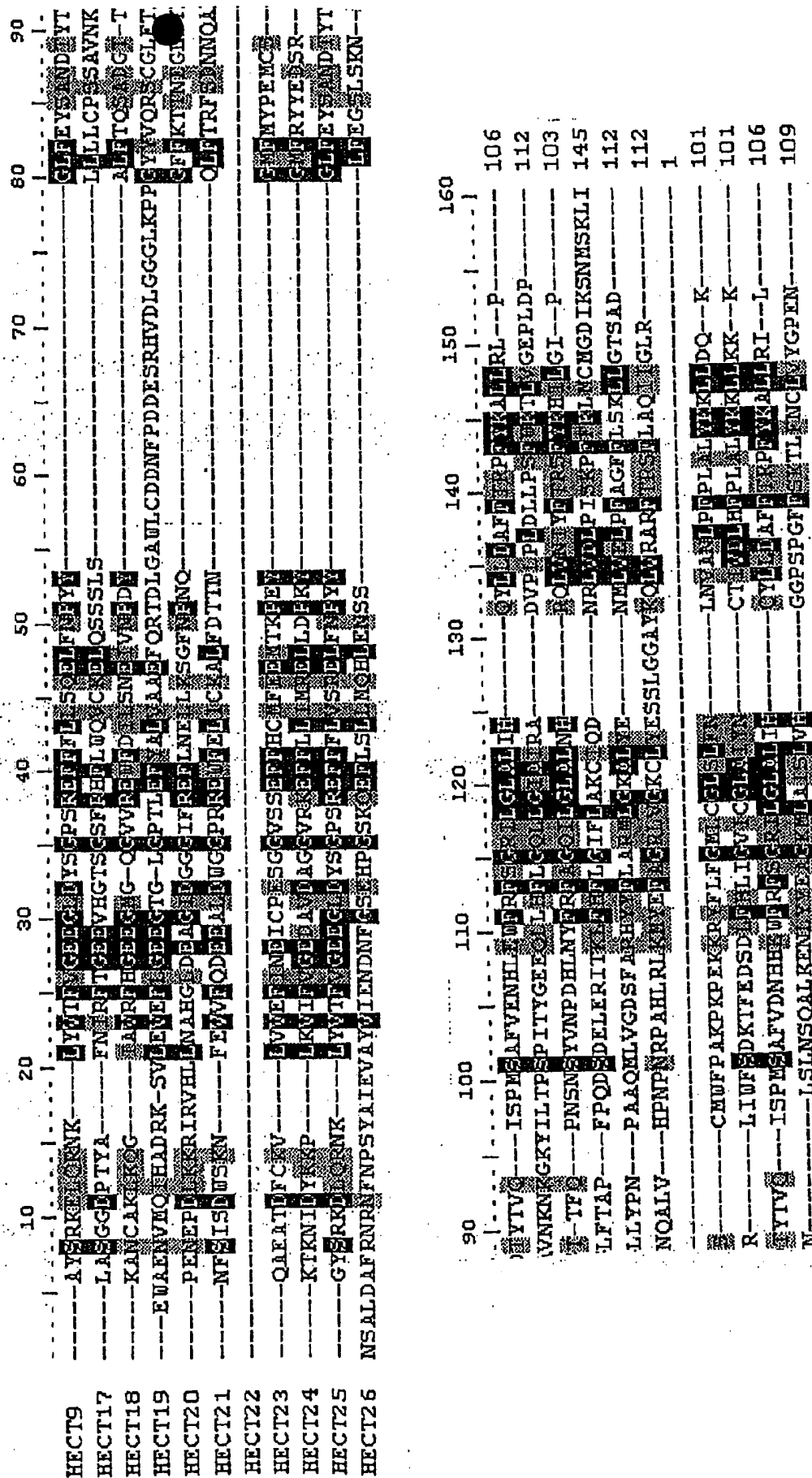


FIGURE 1

[illegible]

FIGURE 1 (Suite 1)

HECT9
HECT17
HECT18
HECT19
HECT20
HECT21
HECT22
HECT23
HECT24
HECT25
HECT26

330 340 350 360 370 380 390 400 410

GANTQVTEAKKEV RRVVMEVER GVVCTEDLVGF VVWDSLS ELA EEL ELA S EHDGHLWT
GRHEVADENNDI AAAARSLS RRELONV CTVVANGAGS FLQLLI LLSPL EEL RTCC LPYIM LFL LMAE LTMVVG LMETDQHL
EGSHVATONNNAE VVCL TILNTE A AOPK MAF L GFAMH PSL QO FFEY EEL EEL SSG EPEID SD IKN IEV TSG EERD EFWI
SEDEK TMDNAEEVVMH FBUCHT G GCKL E DFRDEN VFFEMKLS SFSHEEVOV E CQNCPSD LAD INI TEPKLG NTRD SFGF
GKDIPTSAKKIA HLMADVENEK OIRBHCLAF RGLANW SLEU R FFOE SGA QVPI S EPLKSF INVS GC NSADHEWI
GKQTPV NASHIT FNV L A VPL AS QVKE E EHF L G NNE W PENI L L FENELEL LCC FGDH SSI KAAV VVGCS HFRKEM
GAGTNGYGNESR ION QARKE ESKA AAEK AELL KMPQA L DLTTC EELKKVCC DPEV VV LALH L LRTD EPE DSR
GISLPDD TITPRV VSKYD WIFNV SVKAVYEF EFCF N CEKE LRHFYP EELMTA LIG NTYVHT QV KNS VTRQG WOK HETI
GADTANKN OCE VVAY VVIFNK SVASLF DATHA CFFH VCCGK LLL FOPNELOAN LIG NINYP EUL E KNT EYCE WTAEHETI
GANIPV TEKNTKEE EPRV VV F I E GVVATE SLV RFP FWDAD LVS VFFARELEL E L VAS EAEIH SDNNNTEY RCG YHDHETI
CLRE IHLSDNYMLNEDH ECGHRI I RHTPFES FVCLKIT GVLEKQ APEAFCS ECHK PESI HX ILS E LFTVHTLPDVKALGFU

410 420 430 440 450 460 470 480

LMTIUFMAVER FNNV ERL LLOFVTC EESPYE EALRGSGNG RRFCL EKKGKIT S EERA 303
OEH FVGLDEM FTCEELCSF AFACNOER PFTCPCKDGGPD TEEV PYPK IAPPDGTAGSPDSRYIRV 337
EWTQMPHETWED TIOEEVVL LLOFVTC ESR PHGC EANNNGSGSGH ONFTIAAVPYTPNL EPTS 302
EGFLRFVRM CG EESSEERKAE LQF ITGC EEL EPPGC LANNI H ERLT VWRKVDATDAS YFSV 434
PWIKVFURVWEG FTDEEER L LKVTWSC EERPELLC EKELYPAFC ENG GSDLER LP 307
KWRKRFUWVSS TCEELAE LLOFTT EES EPPGC EALCESTFO IAAPTHST LP 307
SRVATFUEHNN FTNEDES ERLVTC ESR P ERYIYPD KLG YTTDA LPER 169
FTI L FVGFHK FLDEKK EEL FTERDE F HARG IOKEIVFRCPETTSRD EPTS 292
ETHIFUOFHE PLEAKOELLE TCEDR EILC NKSKLVIOSTGG G EY LPVS 295
IIVIRFVMAVER FNNV ERL LLOFVTC EESPYE EALRGSGNGP RRF CV EKKUGKIT ALPRA 303
GFONS ELQWEDGKSTTTMED L I L A TGCES E P A S FKETPS ECLHVB FVVG 291

FIGURE 1 (Suite 2)

HECT9	HTCFNRRLDLP	PPSPSYSLYEKLLT	ETST	EGLE	338
HECT17	ETCHFFKLP	YSSLEHLEKLRCL	HYR		366
HECT18	STCHNRLKLP	PPSKELKQNLV	HCGSYG	TTIA	338
HECT19	NTCHYLKLP	YSSSEETRRERLLA	TMKQ	FHN	468
HECT20	STCHNRLKLP	YDETLSKLLYAI	CAAG	FEIS	342
HECT21	HTCFNQLCLF	TYDSYEEHHRMLQL	ISEGCEG	FGML	343
HECT22	STCSSILFLF	HTHETARQCFERLRY	AYNCVAIDTMS		206
HECT23	ITCHNALSLF	YSSMEREEALQV	INNMRG	FV	325
HECT24	HTCFNRLDLF	YVEKEITLSKL	QTHHNEG	FSLI	330
HECT25	HTCFNRLDLF	PPSPFSLYEKLIT	ETST	EGLE	338
HECT26	NECNCLAF	PIITNVEEFEN	DFTRNTRLR		322

FIGURE 1 (Suite 3)

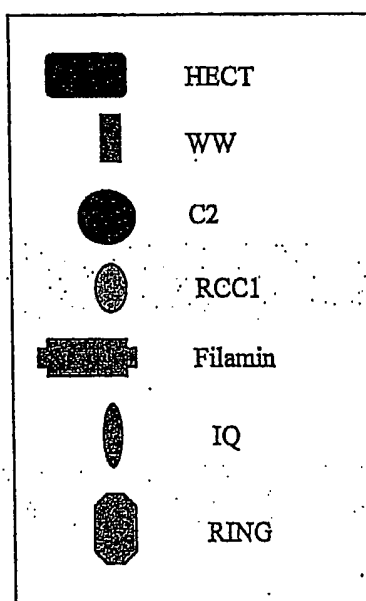
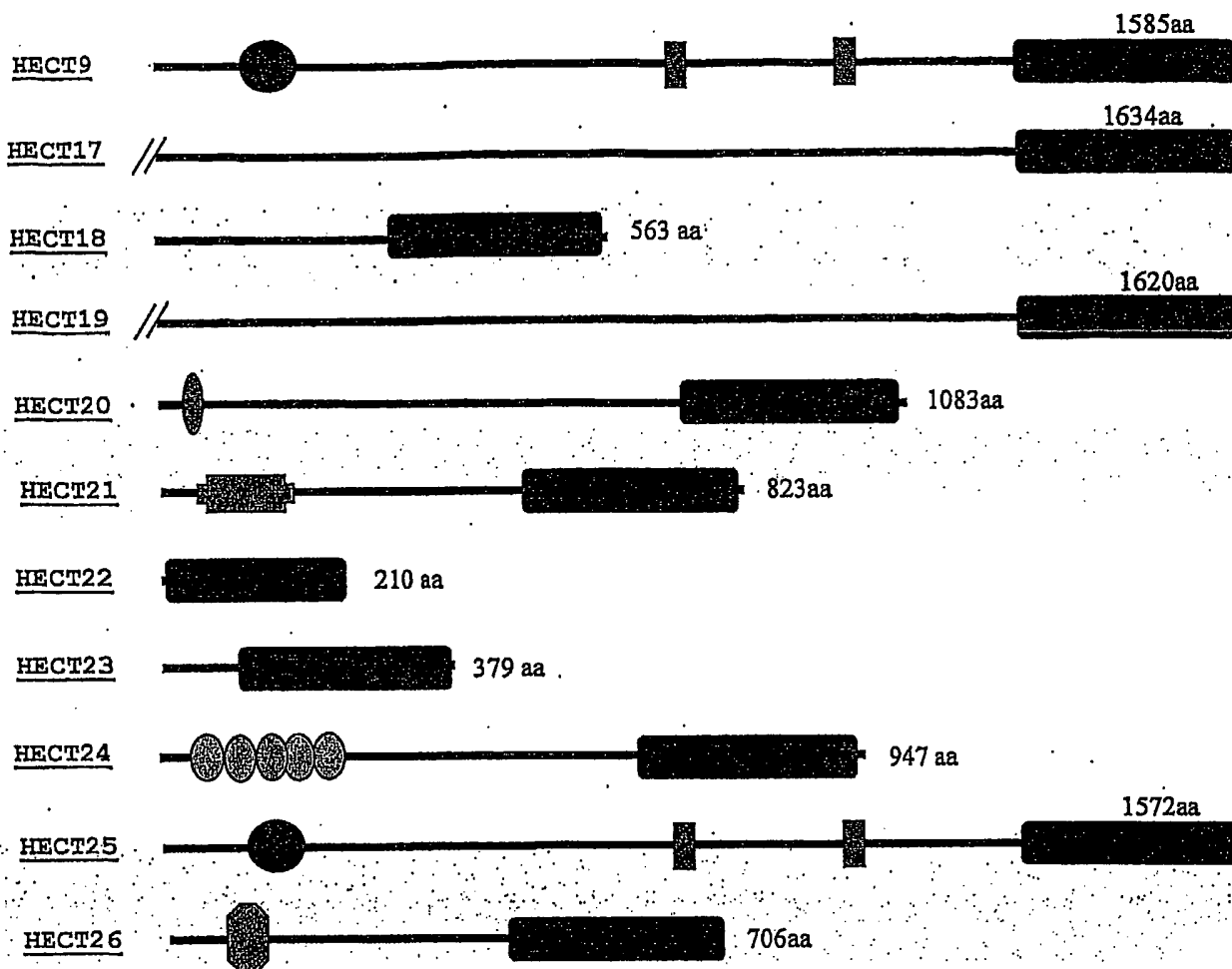


FIGURE 2

LISTE DE SEQUENCES

<110> CYTOMICS SYSTEMS

<120> Acides nucléiques codant de nouvelles protéines à
domaine HECT, leurs utilisations en diagnostic et
thérapie

<130> Cytomics - Domaine HECT

<140>

<141>

<160> 22

<170> PatentIn Ver. 2.1

<210> 1

<211> 1585

<212> PRT

<213> Homo sapiens

<400> 1

```

Met Ala Ser Pro Ser Arg Asn Ser Gln Ser Arg Arg Arg Cys Lys Glu
  1              5              10              15

Pro Leu Arg Tyr Ser Tyr Asn Pro Asp Gln Phe His Asn Met Asp Leu
      20              25              30

Arg Gly Gly Pro His Asp Gly Val Thr Ile Pro Arg Ser Thr Ser Asp
      35              40              45

Thr Asp Leu Val Thr Ser Asp Ser Arg Ser Thr Leu Met Val Ser Ser
      50              55              60

Ser Tyr Tyr Ser Ile Gly His Ser Gln Asp Leu Val Ile His Trp Asp
      65              70              75              80

Ile Lys Glu Glu Val Asp Ala Gly Asp Trp Ile Gly Met Tyr Leu Ile
      85              90              95

Asp Glu Val Leu Ser Glu Asn Phe Leu Asp Tyr Lys Asn Arg Gly Val
      100             105             110

Asn Gly Ser His Arg Gly Gln Ile Ile Trp Lys Ile Asp Ala Ser Ser
      115             120             125

Tyr Phe Val Glu Pro Glu Thr Lys Ile Cys Phe Lys Tyr Tyr His Gly
      130             135             140

Val Ser Gly Ala Leu Arg Ala Thr Thr Pro Ser Val Thr Val Lys Asn
      145             150             155             160

Ser Ala Ala Pro Ile Phe Lys Ser Ile Gly Ala Asp Glu Thr Val Gln
      165             170             175

Gly Gln Gly Ser Arg Arg Leu Ile Ser Phe Ser Leu Ser Asp Phe Gln
      180             185             190

Ala Met Gly Leu Lys Lys Gly Met Phe Phe Asn Pro Asp Pro Tyr Leu
      195             200             205

```

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

Cys Gly Asp Pro Glu Thr Pro Arg Thr His Tyr Ile Arg Ile His Thr
 530 535 540
 Leu Leu His Ser Met Pro Ser Ala Gln Gly Gly Ser Ala Ala Glu Glu
 545 550 555 560
 Glu Asp Gly Ala Glu Glu Glu Ser Thr Leu Lys Asp Ser Ser Glu Lys
 565 570 575
 Asp Gly Leu Ser Glu Val Asp Thr Val Ala Ala Asp Pro Ser Ala Leu
 580 585 590
 Glu Glu Asp Arg Glu Glu Pro Glu Gly Ala Thr Pro Gly Thr Ala His
 595 600 605
 Pro Gly His Ser Gly Gly His Phe Pro Ser Leu Ala Asn Gly Ala Ala
 610 615 620
 Gln Asp Gly Asp Thr His Pro Ser Thr Gly Ser Glu Ser Asp Ser Ser
 625 630 635 640
 Pro Arg Gln Gly Gly Asp His Ser Cys Glu Gly Cys Asp Ala Ser Cys
 645 650 655
 Cys Ser Pro Ser Cys Tyr Ser Ser Ser Cys Tyr Ser Thr Ser Cys Tyr
 660 665 670
 Ser Ser Ser Cys Tyr Ser Ala Ser Cys Tyr Ser Pro Ser Cys Tyr Asn
 675 680 685
 Gly Asn Arg Phe Ala Ser His Thr Arg Phe Ser Ser Val Asp Ser Ala
 690 695 700
 Lys Ile Ser Glu Ser Thr Val Phe Ser Ser Gln Asp Asp Glu Glu Glu
 705 710 715 720
 Glu Asn Ser Ala Phe Glu Ser Val Pro Asp Ser Met Gln Ser Pro Glu
 725 730 735
 Leu Asp Pro Glu Ser Thr Asn Gly Ala Gly Pro Trp Gln Asp Glu Leu
 740 745 750
 Ala Ala Pro Ser Gly His Val Glu Arg Ser Pro Glu Gly Leu Glu Ser
 755 760 765
 Pro Val Ala Gly Pro Ser Asn Arg Arg Glu Gly Glu Cys Pro Ile Leu
 770 775 780
 His Asn Ser Gln Pro Val Ser Gln Leu Pro Ser Leu Arg Pro Glu His
 785 790 795 800
 His His Tyr Pro Thr Ile Asp Glu Pro Leu Pro Pro Asn Trp Glu Ala
 805 810 815
 Arg Ile Asp Ser His Gly Arg Val Phe Tyr Val Asp His Val Asn Arg
 820 825 830
 Thr Thr Thr Trp Gln Arg Pro Thr Ala Ala Ala Thr Pro Asp Gly Met
 835 840 845
 Arg Arg Ser Gly Ser Ile Gln Gln Met Glu Gln Leu Asn Arg Arg Tyr

850	855	860
Gln Asn Ile Gln Arg Thr Ile Ala Thr Glu Arg Ser Glu Glu Asp Ser		
865	870	875 880
Gly Ser Gln Ser Cys Glu Gln Ala Pro Ala Gly Gly Gly Gly Gly Gly		
	885	890 895
Gly Ser Asp Ser Glu Ala Glu Ser Ser Gln Ser Ser Leu Asp Leu Arg		
	900	905 910
Arg Glu Gly Ser Leu Ser Pro Val Asn Ser Gln Lys Ile Thr Leu Leu		
	915	920 925
Leu Gln Ser Pro Ala Val Lys Phe Ile Thr Asn Pro Glu Phe Phe Thr		
	930	935 940
Val Leu His Ala Asn Tyr Ser Ala Tyr Arg Val Phe Thr Ser Ser Thr		
	945	950 955 960
Cys Leu Lys His Met Ile Leu Lys Val Arg Arg Asp Ala Arg Asn Phe		
	965	970 975
Glu Arg Tyr Gln His Asn Arg Asp Leu Val Asn Phe Ile Asn Met Phe		
	980	985 990
Ala Asp Thr Arg Leu Glu Leu Pro Arg Gly Trp Glu Ile Lys Thr Asp		
	995	1000 1005
Gln Gln Gly Lys Ser Phe Phe Val Asp His Asn Ser Arg Ala Thr Thr		
	1010	1015 1020
Phe Ile Asp Pro Arg Ile Pro Leu Gln Asn Gly Arg Leu Pro Asn His		
	1025	1030 1035 1040
Leu Thr His Arg Gln His Leu Gln Arg Leu Arg Ser Tyr Ser Ala Gly		
	1045	1050 1055
Glu Ala Ser Glu Val Ser Arg Asn Arg Gly Ala Ser Leu Leu Ala Arg		
	1060	1065 1070
Pro Gly His Ser Leu Val Ala Ala Ile Arg Ser Gln His Gln His Glu		
	1075	1080 1085
Ser Leu Pro Leu Ala Tyr Asn Asp Lys Ile Val Ala Phe Leu Arg Gln		
	1090	1095 1100
Pro Asn Ile Phe Glu Met Leu Gln Glu Arg Gln Pro Ser Leu Ala Arg		
	1105	1110 1115 1120
Asn His Thr Leu Arg Glu Lys Ile His Tyr Ile Arg Thr Glu Gly Asn		
	1125	1130 1135
His Gly Leu Glu Lys Leu Ser Cys Asp Ala Asp Leu Val Ile Leu Leu		
	1140	1145 1150
Ser Leu Phe Glu Glu Glu Ile Met Ser Tyr Val Pro Leu Gln Ala Ala		
	1155	1160 1165
Phe His Pro Gly Tyr Ser Phe Ser Pro Arg Cys Ser Pro Cys Ser Ser		
	1170	1175 1180

Pro Gln Asn Ser Pro Gly Leu Gln Arg Ala Ser Ala Arg Ala Pro Ser
 1185 1190 1195 1200
 Pro Tyr Arg Arg Asp Phe Glu Ala Lys Leu Arg Asn Phe Tyr Arg Lys
 1205 1210 1215
 Leu Glu Ala Lys Gly Phe Gly Gln Gly Pro Gly Lys Ile Lys Leu Ile
 1220 1225 1230
 Ile Arg Arg Asp His Leu Leu Glu Gly Thr Phe Asn Gln Val Met Ala
 1235 1240 1245
 Tyr Ser Arg Lys Glu Leu Gln Arg Asn Lys Leu Tyr Val Thr Phe Val
 1250 1255 1260
 Gly Glu Glu Gly Leu Asp Tyr Ser Gly Pro Ser Arg Glu Phe Phe Phe
 1265 1270 1275 1280
 Leu Leu Ser Gln Glu Leu Phe Asn Pro Tyr Tyr Gly Leu Phe Glu Tyr
 1285 1290 1295
 Ser Ala Asn Asp Thr Tyr Thr Val Gln Ile Ser Pro Met Ser Ala Phe
 1300 1305 1310
 Val Glu Asn His Leu Glu Trp Phe Arg Phe Ser Gly Arg Ile Leu Gly
 1315 1320 1325
 Leu Ala Leu Ile His Gln Tyr Leu Leu Asp Ala Phe Phe Thr Arg Pro
 1330 1335 1340
 Phe Tyr Lys Ala Leu Leu Arg Leu Pro Cys Asp Leu Ser Asp Leu Glu
 1345 1350 1355 1360
 Tyr Leu Asp Glu Glu Phe His Gln Ser Leu Gln Trp Met Lys Asp Asn
 1365 1370 1375
 Asn Ile Thr Asp Ile Leu Asp Leu Thr Phe Thr Val Asn Glu Glu Val
 1380 1385 1390
 Phe Gly Gln Val Thr Glu Arg Glu Leu Lys Ser Gly Gly Ala Asn Thr
 1395 1400 1405
 Gln Val Thr Glu Lys Asn Lys Lys Glu Tyr Ile Glu Arg Met Val Lys
 1410 1415 1420
 Trp Arg Val Glu Arg Gly Val Val Gln Gln Thr Glu Ala Leu Val Arg
 1425 1430 1435 1440
 Gly Phe Tyr Glu Val Val Asp Ser Arg Leu Val Ser Val Phe Asp Ala
 1445 1450 1455
 Arg Glu Leu Glu Leu Val Ile Ala Gly Thr Ala Glu Ile Asp Leu Asn
 1460 1465 1470
 Asp Trp Arg Asn Asn Thr Glu Tyr Arg Gly Gly Tyr His Asp Gly His
 1475 1480 1485
 Leu Val Ile Arg Trp Phe Trp Ala Ala Val Glu Arg Phe Asn Asn Glu
 1490 1495 1500

Gln Arg Leu Arg Leu Leu Gln Phe Val Thr Gly Thr Ser Ser Val Pro
1505 1510 1515 1520

Tyr Glu Gly Phe Ala Ala Leu Arg Gly Ser Asn Gly Leu Arg Arg Phe
1525 1530 1535

Cys Ile Glu Lys Trp Gly Lys Ile Thr Ser Leu Pro Arg Ala His Thr
1540 1545 1550

Cys Phe Asn Arg Leu Asp Leu Pro Pro Tyr Pro Ser Tyr Ser Met Leu
1555 1560 1565

Tyr Glu Lys Leu Leu Thr Ala Val Glu Glu Thr Ser Thr Phe Gly Leu
1570 1575 1580

Glu
1585

<210> 2
<211> 1630
<212> PRT
<213> Homo sapiens

<400> 2
Lys Thr Phe Gly Leu His Gly Val Val Leu Asp Val Asp Ser Val Asn
1 5 10 15

Glu Leu Val Gln Val Glu Thr Tyr Leu Arg Ser Glu Gly Val Leu Val
20 25 30

Arg Tyr Trp Tyr Pro Ile Asp Met Leu Glu Arg Pro Pro Ala Gly Tyr
35 40 45

Arg Arg Thr Ala Thr Asn Gly Leu Val Thr Leu Asp Asn Thr Asn Leu
50 55 60

Gln Ile His Arg Glu Leu Leu Arg Cys Glu Ala Ala Leu Ala Arg Leu
65 70 75 80

Tyr Cys Arg Met Ala Leu Leu Asn Ile Phe Ala Pro Lys Leu Pro His
85 90 95

Leu Phe Thr Arg Leu Phe His Ile Pro Ala Ile Arg Asp Ile Thr Leu
100 105 110

Glu His Leu Gln Leu Leu Ser Asn Gln Leu Leu Ala Pro Pro Leu Pro
115 120 125

Asp Gly Thr Ile Ser Ser Ser Ser Ile Leu Leu Ala Gln Ser Leu Gln
130 135 140

His Cys Ile His Ser Gln Asn Cys Ser Ala Thr Asp Leu Phe Tyr Gln
145 150 155 160

Gly Asn Ser Gln Thr Val Arg Glu Trp Leu Asn Val Ala Ile Thr Arg
165 170 175

Thr Leu His Gln Gly Glu Glu Ser Leu Leu Glu Leu Thr Lys Gln Ile
180 185 190

Cys Ser Phe Leu Gln Thr Ala Pro Glu Gln Phe Pro Ser Glu Glu Phe
 195 200 205
 Pro Ile Ser Glu Ser Lys Val Asn Met Asp Val Asn Phe Pro Gly Ala
 210 215 220
 Ala Phe Val Val Val Ser Cys Lys Glu Ser Gln Ser Gly Phe Arg Lys
 225 230 235 240
 Asp Ser Ser Leu Tyr Lys Ala Pro Trp Ala Arg Val Leu Val Tyr Gly
 245 250 255
 Leu Gly His Lys Val Lys Arg Asn Gly Gln Leu Asn Leu Ile Glu Ala
 260 265 270
 Ala Cys Tyr Pro Arg Asp Ala Ser Pro Ala Asn Thr Gly Leu Ala Pro
 275 280 285
 Pro Pro Thr Ala Asp Gln Tyr Pro Ser Val Val Leu Ser Thr Asp Arg
 290 295 300
 Val His Ile Lys Leu Gly Val Ser Pro Pro Pro Gly Ala Val Leu Val
 305 310 315 320
 Leu His Ser Leu Pro Leu Glu Phe Pro Leu Ala Met Ala Phe Ala Glu
 325 330 335
 Gln Leu Leu Ser Trp Lys Ser Glu Asp Ser Glu Gly Lys Ser Glu Asp
 340 345 350
 Glu Pro Asp Thr Ile Pro Thr Ser Val Leu Leu Gln Val Val Glu Leu
 355 360 365
 Leu Gly Asn Phe Leu Trp Thr Thr Asp Met Ala Ala Cys Val Lys Glu
 370 375 380
 Leu Val Phe His Leu Leu Ala Glu Leu Leu Arg Thr Val His Thr Leu
 385 390 395 400
 Glu Gln Arg Arg His Pro Ala Gly Leu Ser Ser Ser Ile Ala Leu Gln
 405 410 415
 Leu Asn Pro Cys Leu Ala Met Leu Met Ala Leu Gln Ser Glu Leu His
 420 425 430
 Lys Leu Tyr Asp Glu Glu Thr Gln Asn Trp Val Ser Gly Gly Ala Cys
 435 440 445
 Gly Gly Ser Gly Gly Ala Ala Ala Gly Asp Gln Gly Arg Phe Ser Thr
 450 455 460
 Tyr Phe His Ala Leu Met Glu Gly Cys Leu Ala Val Ala Glu Val Thr
 465 470 475 480
 Leu Pro Thr Asn Met Ser Val Thr Ala Ser Gly Val Thr Ser Ala Thr
 485 490 495
 Ala Pro Asn Leu Ser Asp Ser Ser Ser Ser Ser Ser Ser Pro Gly
 500 505 510

Gln Thr Pro Gln Ser Pro Ser Leu Leu Ser Lys Arg Lys Lys Val Lys
 515 520 525
 Met Lys Arg Glu Lys Ala Ser Ser Ser Gly Lys Arg Gln Ser Ser Arg
 530 535 540
 Thr Val Asp Ser Asp Pro Thr Val Leu Ser Ile Gly Gly Ser Lys Pro
 545 550 555 560
 Glu Asp Met Leu Trp Phe His Arg Ala Leu Thr Leu Leu Ile Ile Leu
 565 570 575
 Arg His Leu Thr Arg Lys Asp Pro Gln Gly Leu Gly Val Thr Ser Asp
 580 585 590
 Ala Ile Ala Asp Ala Cys Gln Ala Leu Val Gly Pro Thr Ala His Ser
 595 600 605
 Arg Leu Leu Val Ile Ser Gly Ile Pro Thr His Leu Asp Glu Gly Val
 610 615 620
 Val Arg Gly Ala Ile Arg Lys Ala Cys Asn Ala His Gly Gly Val Phe
 625 630 635 640
 Lys Asp Glu Ile Tyr Ile Pro Leu Gln Glu Glu Asp Thr Lys Lys Pro
 645 650 655
 Lys Asp Lys Ala Glu Gly Gly Asp Gly Lys Val Glu Pro Glu Lys Thr
 660 665 670
 Leu Ala Phe Pro Gly Thr Asp Ser Met Glu Val Ser Thr Ser Ser Ser
 675 680 685
 Leu Thr Pro Ala Met Ser Ile Ser Ala Ser Ala Ser Thr Ser Gln Ala
 690 695 700
 Ser Ile Cys Ser Ser Gln Gly Ile Ser Gln Thr Val Ser Asp Leu Ser
 705 710 715 720
 Val Asp Pro Leu Pro Ala Gly Leu Glu Leu Pro Ile Pro Pro Gly Leu
 725 730 735
 Leu Glu Pro His Ala Val Ser Ser Gln Glu Ser Leu Asp Ile Ser Leu
 740 745 750
 Cys Ser Thr Gly Ser Leu Gly Ser Leu Gly Ser Leu Gly Glu Pro Leu
 755 760 765
 Asp Asn Ala Glu Thr Ala Ser Val Ser Asp Met Gly Ser Met Tyr Thr
 770 775 780
 Val Thr Ser Leu Asp Asn Gln Pro Leu Ala Ala Arg Pro Ile Lys Gly
 785 790 795 800
 Phe Ala Val Val Glu Ile Arg Ser Arg Ala Lys Ile Glu Lys Ile Arg
 805 810 815
 Ala Ser Leu Phe Asn Asn Asn Asp Leu Ile Gly Leu Ser Ser Leu Asp
 820 825 830
 Gly Glu Asp Glu Leu Met Glu Met Ser Thr Glu Glu Ile Leu Thr Val

835	840	845
Ser Val Val Asn Gln Ser Leu Phe Asp Thr Gln Gly Ser Pro Gly Leu		
850	855	860
Glu Asp Tyr Phe Asn Asp Lys Ser Ile Lys Gly Glu Lys Leu Val Pro		
865	870	875 880
Gly Ala Arg Glu Val Leu Thr Glu Ile Phe Lys Ser Cys Ala His Ser		
	885	890 895
Glu Gln Thr Leu Ser Leu Thr Pro Ala Lys Pro Ile Arg Val Ser Asp		
	900	905 910
Ile Tyr Leu Ser Lys Glu Gln Ile Asn Ser Gln Thr Pro Gly Asn Leu		
	915	920 925
Leu His Leu Phe Phe Thr Asn Val Arg Pro Pro Lys Lys Val Leu Glu		
	930	935 940
Asp Gln Leu Thr Gln Ile Leu Arg Lys Tyr Gly Val Pro Lys Pro Lys		
	945	950 955 960
Phe Asp Lys Ser Lys Tyr Ser Lys Ala Gly Lys Glu Gln His Pro Val		
	965	970 975
Lys Val Val Ser Thr Lys Arg Pro Ile Thr Lys Pro Pro Ala Lys Asp		
	980	985 990
Lys Ala Val Leu Asn Ser Val Ser Arg Thr Ala Leu Ser Glu Lys Lys		
	995	1000 1005
Pro Thr Val Lys Pro Lys Ser Pro Glu Lys Ser Lys Pro Asp Glu Lys		
	1010	1015 1020
Asp Pro Glu Lys Ser Pro Thr Lys Lys Gln Glu Val Pro Glu Glu Lys		
	1025	1030 1035 1040
Tyr Leu Thr Leu Glu Gly Phe His Lys Phe Val Ile Asp Arg Ala Arg		
	1045	1050 1055
Gln Asp Ile Arg Ser Val Trp Arg Ala Ile Leu Ser Cys Gly Tyr Asp		
	1060	1065 1070
Leu His Phe Glu Arg Cys Ala Cys Ile Asp Val Arg His Ala Gln Lys		
	1075	1080 1085
Ala Ser Arg Lys Trp Thr Leu Glu Met Asp Val Ala Leu Val Gln Tyr		
	1090	1095 1100
Ile Asn Gln Leu Cys Arg His Leu Ala Ile Thr Pro Ala Arg Leu His		
	1105	1110 1115 1120
Pro His Glu Val Tyr Leu Asp Pro Ala Asp Ala Ala Asp Pro Arg Val		
	1125	1130 1135
Ala Cys Leu Leu Asn Val Pro Ile Glu Ser Leu Arg Leu Arg Phe Ala		
	1140	1145 1150
Leu Leu Gln Ser Leu Asn Thr Thr Leu Glu Thr Phe Phe Leu Pro Leu		
	1155	1160 1165

Val Glu Leu Arg Gln Thr Pro Met Tyr Thr His Ser Ile Ala Ala Leu
 1170 1175 1180

Leu Lys Glu Ala Lys Gly Leu Ile Phe Tyr Asp Thr Lys Val Thr Val
 1185 1190 1195 1200

Met Asn Arg Val Leu Asn Ala Thr Val Gln Arg Thr Ala Asp His Ala
 1205 1210 1215

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

Arg Ala Ser Glu Asn Ser Tyr Phe Cys Gln Ala Ala Arg Gln Leu Ala
 1235 1240 1245

Ser Val Pro Ser Ser Gln Leu Cys Val Lys Leu Ala Ser Gly Gly Asp
 1250 1255 1260

Pro Thr Tyr Ala Phe Asn Ile Arg Phe Thr Gly Glu Glu Val His Gly
 1265 1270 1275 1280

Thr Ser Gly Ser Phe Arg His Phe Leu Trp Gln Val Cys Lys Glu Leu
 1285 1290 1295

Gln Ser Ser Ser Leu Ser Leu Leu Leu Leu Cys Pro Ser Ser Ala Val
 1300 1305 1310

Asn Lys Asn Lys Gly Lys Tyr Ile Leu Thr Pro Ser Pro Ile Thr Tyr
 1315 1320 1325

Gly Glu Glu Gln Leu Leu His Phe Leu Gly Gln Leu Leu Gly Ile Ala
 1330 1335 1340

Ile Arg Ala Asp Val Pro Leu Pro Leu Asp Leu Leu Pro Ser Phe Trp
 1345 1350 1355 1360

Lys Thr Leu Val Gly Glu Pro Leu Asp Pro Glu Gln Asp Leu Gln Glu
 1365 1370 1375

Ala Asp Ile Leu Thr Tyr Asn Tyr Val Lys Lys Phe Glu Ser Ile Asn
 1380 1385 1390

Asp Glu Thr Glu Leu Glu Ala Leu Cys Ala Glu Ile Ala Ser Gln His
 1395 1400 1405

Leu Ala Thr Glu Ser Pro Asp Ser Pro Asn Lys Pro Cys Cys Arg Phe
 1410 1415 1420

Thr Tyr Leu Thr Met Thr Gly Glu Glu Val Glu Leu Cys Ser Arg Gly
 1425 1430 1435 1440

Arg His Ile Leu Val Ala Trp Glu Asn Lys Asp Ile Tyr Ala Ala Ala
 1445 1450 1455

Ile Arg Ser Leu Arg Leu Arg Glu Leu Gln Asn Val Glu Cys Val Thr
 1460 1465 1470

Ala Val Arg Ala Gly Leu Gly Ser Ile Ile Pro Leu Gln Leu Leu Thr
 1475 1480 1485

Met Leu Ser Pro Leu Glu Met Glu Leu Arg Thr Cys Gly Leu Pro Tyr
 1490 1495 1500

Ile Asn Leu Glu Phe Leu Lys Ala His Thr Met Tyr Gln Val Gly Leu
 1505 1510 1515 1520

Met Glu Thr Asp Gln His Ile Glu Phe Phe Trp Gly Ala Leu Glu Met
 1525 1530 1535

Phe Thr Gln Glu Glu Leu Cys Lys Phe Ile Lys Phe Ala Cys Asn Gln
 1540 1545 1550

Glu Arg Ile Pro Phe Thr Cys Pro Cys Lys Asp Gly Gly Pro Asp Thr
 1555 1560 1565

Ala His Val Pro Pro Tyr Pro Met Lys Ile Ala Pro Pro Asp Gly Thr
 1570 1575 1580

Ala Gly Ser Pro Asp Ser Arg Tyr Ile Arg Val Glu Thr Cys Met Phe
 1585 1590 1595 1600

Met Ile Lys Leu Pro Gln Tyr Ser Ser Leu Glu Ile Met Leu Glu Lys
 1605 1610 1615

Leu Arg Cys Ala Ile His Tyr Arg Glu Asp Pro Leu Ser Gly
 1620 1625 1630

<210> 3

<211> 562

<212> PRT

<213> Homo sapiens

<400> 3

Met Met Phe Lys Lys His Phe Cys Phe Ser Gln Pro Leu Glu Leu Leu
 1 5 10 15

Trp His Ser Leu Asp Glu Trp Leu Val Leu Ile Ala Thr Glu Leu Met
 20 25 30

Lys Asn Lys Arg Asp Ser Thr Glu Ile Thr Ser Ile Leu Leu Lys Gln
 35 40 45

Lys Gly Gln Asp Gln Asp Ala Ala Ser Ile Pro Pro Phe Glu Pro Pro
 50 55 60

Gly Pro Gly Ser Tyr Glu Asn Leu Ser Thr Gly Thr Arg Glu Ser Lys
 65 70 75 80

Pro Asp Ala Leu Ala Gly Arg Gln Glu Ala Ser Ala Asp Cys Gln Asp
 85 90 95

Val Ile Ser Met Thr Ala Asn Arg Leu Ser Ala Val Ile Gln Ala Phe
 100 105 110

Tyr Met Cys Cys Ser Cys Gln Met Pro Pro Gly Met Thr Ser Pro Arg
 115 120 125

Phe Ile Glu Phe Val Cys Lys His Asp Glu Val Leu Lys Cys Phe Val
 130 135 140

Asn	Arg	Asn	Pro	Lys	Ile	Ile	Phe	Asp	His	Phe	His	Phe	Leu	Leu	Glu	
145					150					155					160	
Cys	Pro	Glu	Leu	Met	Ser	Arg	Phe	Met	His	Ile	Ile	Lys	Ala	Gln	Pro	
				165					170					175		
Phe	Lys	Asp	Arg	Cys	Glu	Trp	Phe	Tyr	Glu	His	Leu	His	Ser	Gly	Gln	
			180					185					190			
Pro	Asp	Ser	Asp	Met	Val	His	Arg	Pro	Val	Asn	Glu	Asn	Asp	Ile	Leu	
		195					200					205				
Leu	Val	His	Arg	Asp	Ser	Ile	Phe	Arg	Ser	Ser	Cys	Glu	Val	Val	Ser	
	210					215					220					
Lys	Ala	Asn	Cys	Ala	Lys	Leu	Lys	Gln	Gly	Ile	Ala	Val	Arg	Phe	His	
225					230					235					240	
Gly	Glu	Glu	Gly	Met	Gly	Gln	Gly	Val	Val	Arg	Glu	Trp	Phe	Asp	Ile	
				245					250					255		
Leu	Ser	Asn	Glu	Ile	Val	Asn	Pro	Asp	Tyr	Ala	Leu	Phe	Thr	Gln	Ser	
			260					265						270		
Ala	Asp	Gly	Thr	Thr	Phe	Gln	Pro	Asn	Ser	Asn	Ser	Tyr	Val	Asn	Pro	
		275					280					285				
Asp	His	Leu	Asn	Tyr	Phe	Arg	Phe	Ala	Gly	Gln	Ile	Leu	Gly	Leu	Ala	
	290					295					300					
Leu	Asn	His	Arg	Gln	Leu	Val	Asn	Ile	Tyr	Phe	Thr	Arg	Ser	Phe	Tyr	
305					310					315					320	
Lys	His	Ile	Leu	Gly	Ile	Pro	Val	Asn	Tyr	Gln	Asp	Val	Ala	Ser	Ile	
				325					330					335		
Asp	Pro	Glu	Tyr	Ala	Lys	Asn	Leu	Gln	Trp	Ile	Leu	Asp	Asn	Asp	Ile	
			340					345					350			
Ser	Asp	Leu	Gly	Leu	Glu	Leu	Thr	Phe	Ser	Val	Glu	Thr	Asp	Val	Phe	
		355					360					365				
Gly	Ala	Met	Glu	Glu	Val	Pro	Leu	Lys	Pro	Gly	Gly	Gly	Ser	Ile	Leu	
		370				375					380					
Val	Thr	Gln	Asn	Asn	Lys	Ala	Glu	Tyr	Val	Gln	Leu	Val	Thr	Glu	Leu	
385					390					395					400	
Arg	Met	Thr	Arg	Ala	Ile	Gln	Pro	Gln	Ile	Asn	Ala	Phe	Leu	Gln	Gly	
				405					410					415		
Phe	His	Met	Phe	Ile	Pro	Pro	Ser	Leu	Ile	Gln	Leu	Phe	Asp	Glu	Tyr	
			420					425					430			
Glu	Leu	Glu	Leu	Leu	Leu	Ser	Gly	Met	Pro	Glu	Ile	Asp	Val	Ser	Asp	
		435					440					445				
Trp	Ile	Lys	Asn	Thr	Glu	Tyr	Thr	Ser	Gly	Tyr	Glu	Arg	Glu	Asp	Pro	
						455					460					

Val Ile Gln Trp Phe Trp Glu Val Val Glu Asp Ile Thr Gln Glu Glu
465 470 475 480

Arg Val Leu Leu Leu Gln Phe Val Thr Gly Ser Ser Arg Val Pro His
485 490 495

Gly Gly Phe Ala Asn Ile Met Gly Gly Ser Gly Leu Gln Asn Phe Thr
500 505 510

Ile Ala Ala Val Pro Tyr Thr Pro Asn Leu Leu Pro Thr Ser Ser Thr
515 520 525

Cys Ile Asn Met Leu Lys Leu Pro Glu Tyr Pro Ser Lys Glu Ile Leu
530 535 540

Lys Asp Arg Leu Leu Val Ala Leu His Cys Gly Ser Tyr Gly Tyr Thr
545 550 555 560

Met Ala

<210> 4

<211> 1620

<212> PRT

<213> Homo sapiens

<400> 4

Asn Leu Gln Ile Leu Thr Arg Arg Leu Arg Phe Arg Leu Glu Arg Ala
1 5 10 15

Pro Gly Glu Thr Ala Leu Ile Asp Arg Thr Gly Arg Met Leu Lys Met
20 25 30

Glu Pro Leu Ala Thr Val Glu Ser Leu Glu Gln Tyr Leu Leu Lys Met
35 40 45

Val Ala Lys Gln Trp Tyr Asp Phe Asp Arg Ser Ser Phe Val Phe Val
50 55 60

Arg Lys Leu Arg Glu Gly Gln Asn Phe Ile Phe Arg His Gln His Asp
65 70 75 80

Phe Asp Glu Asn Gly Ile Ile Tyr Trp Ile Gly Thr Asn Ala Lys Thr
85 90 95

Ala Tyr Glu Trp Val Asn Pro Ala Ala Tyr Gly Leu Val Val Val Thr
100 105 110

Ser Ser Glu Gly Arg Asn Leu Pro Tyr Gly Arg Leu Glu Asp Ile Leu
115 120 125

Ser Arg Asp Asn Ser Ala Leu Asn Cys His Ser Asn Asp Asp Lys Asn
130 135 140

Ala Trp Phe Ala Ile Asp Leu Gly Leu Trp Val Ile Pro Ser Ala Tyr
145 150 155 160

Thr Leu Arg His Ala Arg Gly Tyr Gly Arg Ser Ala Leu Arg Asn Trp
165 170 175

Val Phe Gln Val Ser Lys Asp Gly Gln Asn Trp Thr Ser Leu Tyr Thr
 180 185 190
 His Val Asp Asp Cys Ser Leu Asn Glu Pro Gly Ser Thr Ala Thr Trp
 195 200 205
 Pro Leu Asp Pro Pro Lys Asp Glu Lys Gln Gly Trp Arg His Val Arg
 210 215 220
 Ile Lys Gln Met Gly Lys Asn Ala Ser Gly Gln Thr His Tyr Leu Ser
 225 230 235 240
 Leu Ser Gly Phe Glu Leu Tyr Gly Thr Val Asn Gly Val Cys Glu Asp
 245 250 255
 Gln Leu Gly Lys Ala Ala Lys Glu Ala Glu Ala Asn Leu Arg Arg Gln
 260 265 270
 Arg Arg Leu Val Arg Ser Gln Val Leu Lys Tyr Met Val Pro Gly Ala
 275 280 285
 Arg Val Ile Arg Gly Leu Asp Trp Lys Trp Arg Asp Gln Asp Gly Ser
 290 295 300
 Pro Gln Gly Glu Gly Thr Val Thr Gly Glu Leu His Asn Gly Trp Ile
 305 310 315 320
 Asp Val Thr Trp Asp Ala Gly Gly Ser Asn Ser Tyr Arg Met Gly Ala
 325 330 335
 Glu Gly Lys Phe Asp Leu Lys Leu Ala Pro Gly Tyr Asp Pro Asp Thr
 340 345 350
 Val Ala Ser Pro Lys Pro Val Ser Ser Thr Val Ser Gly Thr Thr Gln
 355 360 365
 Ser Trp Ser Ser Leu Val Lys Asn Asn Cys Pro Asp Lys Thr Ser Ala
 370 375 380
 Ala Ala Gly Ser Ser Ser Arg Lys Gly Ser Ser Ser Ser Val Cys Ser
 385 390 395 400
 Val Ala Ser Ser Ser Asp Ile Ser Leu Gly Ser Thr Lys Thr Glu Arg
 405 410 415
 Arg Ser Glu Ile Val Met Glu His Ser Ile Val Ser Gly Ala Asp Val
 420 425 430
 His Glu Pro Ile Val Val Leu Ser Ser Ala Glu Asn Val Pro Gln Thr
 435 440 445
 Glu Val Gly Ser Ser Ser Ser Ala Ser Thr Ser Thr Leu Thr Ala Glu
 450 455 460
 Thr Gly Ser Glu Asn Ala Glu Arg Lys Leu Gly Pro Asp Ser Ser Val
 465 470 475 480
 Arg Thr Pro Gly Glu Ser Ser Ala Ile Ser Met Gly Ile Val Ser Val
 485 490 495

Ser Ser Pro Asp Val Ser Ser Val Ser Glu Leu Thr Asn Lys Glu Ala
 500 505 510
 Ala Ser Gln Arg Pro Leu Ser Ser Ser Ala Ser Asn Arg Leu Ser Val
 515 520 525
 Ser Ser Leu Leu Ala Ala Gly Ala Pro Met Ser Ser Ser Ala Ser Val
 530 535 540
 Pro Asn Leu Ser Ser Arg Glu Thr Ser Ser Leu Glu Ser Phe Val Arg
 545 550 555 560
 Arg Val Ala Asn Ile Ala Arg Thr Asn Ala Thr Asn Asn Met Asn Leu
 565 570 575
 Ser Arg Ser Ser Ser Asp Asn Asn Thr Asn Thr Leu Gly Arg Asn Val
 580 585 590
 Met Ser Thr Ala Thr Ser Pro Leu Met Gly Ala Gln Ser Phe Pro Asn
 595 600 605
 Leu Thr Thr Pro Gly Thr Thr Ser Thr Val Thr Met Ser Thr Ser Ser
 610 615 620
 Val Thr Ser Ser Ser Asn Val Ala Thr Ala Thr Thr Val Leu Ser Val
 625 630 635 640
 Gly Gln Ser Leu Ser Asn Thr Leu Thr Thr Ser Leu Thr Ser Thr Ser
 645 650 655
 Ser Glu Ser Asp Thr Gly Gln Glu Ala Glu Tyr Ser Leu Tyr Asp Phe
 660 665 670
 Leu Asp Ser Cys Arg Ala Ser Thr Leu Leu Ala Glu Leu Asp Asp Asp
 675 680 685
 Glu Asp Leu Pro Glu Pro Asp Glu Glu Asp Asp Glu Asn Glu Asp Asp
 690 695 700
 Asn Gln Glu Asp Gln Glu Tyr Glu Glu Val Met Ile Leu Arg Arg Pro
 705 710 715 720
 Ser Leu Gln Arg Arg Ala Gly Ser Arg Ser Asp Val Thr His His Ala
 725 730 735
 Val Thr Ser Gln Leu Pro Gln Val Pro Ala Gly Ala Gly Ser Arg Pro
 740 745 750
 Ile Gly Glu Gln Glu Glu Glu Glu Tyr Glu Thr Lys Gly Gly Arg Arg
 755 760 765
 Arg Thr Trp Asp Asp Asp Tyr Val Leu Lys Arg Gln Phe Ser Ala Leu
 770 775 780
 Val Pro Ala Phe Asp Pro Arg Pro Gly Arg Thr Asn Val Gln Gln Thr
 785 790 795 800
 Thr Asp Leu Glu Ile Pro Pro Pro Gly Thr Pro His Ser Glu Leu Leu
 805 810 815
 Glu Glu Val Glu Cys Thr Pro Ser Pro Arg Leu Ala Leu Thr Leu Lys

820										825										830																									
Val	Thr	Gly	Leu	Gly	Thr	Thr	Arg	Glu	Val	Glu	Leu	Pro	Leu	Thr	Asn																														
835										840										845																									
Phe	Arg	Ser	Thr	Ile	Phe	Tyr	Tyr	Val	Gln	Lys	Leu	Leu	Gln	Leu	Ser																														
850										855										860																									
Cys	Asn	Gly	Asn	Val	Lys	Ser	Asp	Lys	Leu	Arg	Arg	Ile	Trp	Glu	Pro																														
865										870										875																									
Thr	Tyr	Thr	Ile	Met	Tyr	Arg	Glu	Met	Lys	Asp	Ser	Asp	Lys	Glu	Lys																														
885										890										895																									
Glu	Asn	Gly	Lys	Met	Gly	Cys	Trp	Ser	Ile	Glu	His	Val	Glu	Gln	Tyr																														
900										905										910																									
Leu	Gly	Thr	Asp	Glu	Leu	Pro	Lys	Asn	Asp	Leu	Ile	Thr	Tyr	Leu	Gln																														
915										920										925																									
Lys	Asn	Ala	Asp	Ala	Ala	Phe	Leu	Arg	His	Trp	Lys	Leu	Thr	Gly	Thr																														
930										935										940																									
Asn	Lys	Ser	Ile	Arg	Lys	Asn	Arg	Asn	Cys	Ser	Gln	Leu	Ile	Ala	Ala																														
945										950										955																									
Tyr	Lys	Asp	Phe	Cys	Glu	His	Gly	Thr	Lys	Ser	Gly	Leu	Asn	Gln	Gly																														
965										970										975																									
Ala	Ile	Ser	Thr	Leu	Gln	Ser	Ser	Asp	Ile	Leu	Asn	Leu	Thr	Lys	Glu																														
980										985										990																									
Gln	Pro	Gln	Ala	Lys	Ala	Gly	Asn	Gly	Gln	Asn	Ser	Cys	Gly	Val	Glu																														
995										1000										1005																									
Asp	Val	Leu	Gln	Leu	Leu	Arg	Ile	Leu	Tyr	Ile	Val	Ala	Ser	Asp	Pro																														
1010										1015										1020																									
Tyr	Ser	Arg	Ile	Ser	Gln	Glu	Asp	Gly	Asp	Glu	Gln	Pro	Gln	Phe	Thr																														
1025										1030										1035																									
Phe	Pro	Pro	Asp	Glu	Phe	Thr	Ser	Lys	Lys	Ile	Thr	Thr	Lys	Ile	Leu																														
1045										1050										1055																									
Gln	Gln	Ile	Glu	Glu	Pro	Leu	Ala	Leu	Ala	Ser	Gly	Ala	Leu	Pro	Asp																														
1060										1065										1070																									
Trp	Cys	Glu	Gln	Leu	Thr	Ser	Lys	Cys	Pro	Phe	Leu	Ile	Pro	Phe	Glu																														
1075										1080										1085																									
Thr	Arg	Gln	Leu	Tyr	Phe	Thr	Cys	Thr	Ala	Phe	Gly	Ala	Ser	Arg	Ala																														
1090										1095										1100																									
Ile	Val	Trp	Leu	Gln	Asn	Arg	Arg	Glu	Ala	Thr	Val	Glu	Arg	Thr	Arg																														
1105										1110										1115																									
Thr	Thr	Ser	Ser	Val	Arg	Arg	Asp	Asp	Pro	Gly	Glu	Phe	Arg	Val	Gly																														
1125										1130										1135																									
Arg	Leu	Lys	His	Glu	Arg	Val	Lys	Val	Pro	Arg	Gly	Glu	Ser	Leu	Met																														
1140										1145										1150																									

Glu Trp Ala Glu Asn Val Met Gln Ile His Ala Asp Arg Lys Ser Val
 1155 1160 1165
 Leu Glu Val Glu Phe Leu Gly Glu Glu Gly Thr Gly Leu Gly Pro Thr
 1170 1175 1180
 Leu Glu Phe Tyr Ala Leu Val Ala Ala Glu Phe Gln Arg Thr Asp Leu
 1185 1190 1195 1200
 Gly Ala Trp Leu Cys Asp Asp Asn Phe Pro Asp Asp Glu Ser Arg His
 1205 1210 1215
 Val Asp Leu Gly Gly Gly Leu Lys Pro Pro Gly Tyr Tyr Val Gln Arg
 1220 1225 1230
 Ser Cys Gly Leu Phe Thr Ala Pro Phe Pro Gln Asp Ser Asp Glu Leu
 1235 1240 1245
 Glu Arg Ile Thr Lys Leu Phe His Phe Leu Gly Ile Phe Leu Ala Lys
 1250 1255 1260
 Cys Ile Gln Asp Asn Arg Leu Val Asp Leu Pro Ile Ser Lys Pro Phe
 1265 1270 1275 1280
 Phe Lys Leu Met Cys Met Gly Asp Ile Lys Ser Asn Met Ser Lys Leu
 1285 1290 1295
 Ile Tyr Glu Ser Arg Gly Asp Arg Asp Leu His Cys Thr Glu Ser Gln
 1300 1305 1310
 Ser Glu Ala Ser Thr Glu Glu Gly His Asp Ser Leu Ser Val Gly Ser
 1315 1320 1325
 Phe Glu Glu Asp Ser Lys Ser Glu Phe Ile Leu Asp Pro Pro Lys Pro
 1330 1335 1340
 Lys Pro Pro Ala Trp Phe Asn Gly Ile Leu Thr Trp Glu Asp Phe Glu
 1345 1350 1355 1360
 Leu Val Asn Pro His Arg Ala Arg Phe Leu Lys Glu Ile Lys Asp Leu
 1365 1370 1375
 Ala Ile Lys Arg Arg Gln Ile Leu Ser Asn Lys Gly Leu Ser Glu Asp
 1380 1385 1390
 Glu Lys Asn Thr Lys Leu Gln Glu Leu Val Leu Lys Asn Pro Ser Gly
 1395 1400 1405
 Ser Gly Pro Pro Leu Ser Ile Glu Asp Leu Gly Leu Asn Phe Gln Phe
 1410 1415 1420
 Cys Pro Ser Ser Arg Ile Tyr Gly Phe Thr Ala Val Asp Leu Lys Pro
 1425 1430 1435 1440
 Ser Gly Glu Asp Glu Met Ile Thr Met Asp Asn Ala Glu Glu Tyr Val
 1445 1450 1455
 Asp Leu Met Phe Asp Phe Cys Met His Thr Gly Ile Gln Lys Gln Met
 1460 1465 1470

Glu Ala Phe Arg Asp Gly Phe Asn Lys Val Phe Pro Met Glu Lys Leu
 1475 1480 1485

Ser Ser Phe Ser His Glu Glu Val Gln Met Ile Leu Cys Gly Asn Gln
 1490 1495 1500

Ser Pro Ser Trp Ala Ala Glu Asp Ile Ile Asn Tyr Thr Glu Pro Lys
 1505 1510 1515 1520

Leu Gly Tyr Thr Arg Asp Ser Pro Gly Phe Leu Arg Phe Val Arg Val
 1525 1530 1535

Leu Cys Gly Met Ser Ser Asp Glu Arg Lys Ala Phe Leu Gln Phe Thr
 1540 1545 1550

Thr Gly Cys Ser Thr Leu Pro Pro Gly Gly Leu Ala Asn Leu His Pro
 1555 1560 1565

Arg Leu Thr Val Val Arg Lys Val Asp Ala Thr Asp Ala Ser Tyr Pro
 1570 1575 1580

Ser Val Asn Thr Cys Val His Tyr Leu Lys Leu Pro Glu Tyr Ser Ser
 1585 1590 1595 1600

Glu Glu Ile Met Arg Glu Arg Leu Leu Ala Ala Thr Met Glu Lys Gly
 1605 1610 1615

Phe His Leu Asn
 1620

<210> 5

<211> 1083

<212> PRT

<213> Homo sapiens

<400> 5

Met Phe Ser Phe Glu Gly Asp Phe Lys Thr Arg Pro Lys Val Ser Leu
 1 5 10 15

Gly Gly Ala Ser Arg Lys Glu Glu Lys Ala Ser Leu Leu His Arg Thr
 20 25 30

Gln Glu Glu Arg Arg Lys Arg Glu Glu Glu Arg Arg Arg Leu Lys Asn
 35 40 45

Ala Ile Ile Ile Gln Ser Phe Ile Arg Gly Tyr Arg Asp Arg Lys Gln
 50 55 60

Gln Tyr Ser Ile Gln Arg Ser Ala Phe Asp Arg Cys Ala Thr Leu Ser
 65 70 75 80

Gln Ser Gly Gly Ala Phe Pro Ile Ala Asn Gly Pro Asn Leu Thr Leu
 85 90 95

Leu Val Arg Gln Leu Leu Phe Phe Tyr Lys Gln Asn Glu Asp Ser Lys
 100 105 110

Arg Leu Ile Trp Leu Tyr Gln Asn Leu Ile Lys His Ser Ser Leu Phe
 115 120 125

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

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

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

<210> 6

<211> 823

<212> PRT

<213> Homo sapiens

<400> 6

```

Met Phe Tyr Val Ile Gly Gly Ile Thr Val Ser Val Val Ala Phe Phe
  1             5             10             15

Phe Thr Ile Lys Phe Leu Phe Glu Leu Ala Ala Arg Val Val Ser Phe
      20             25             30

Leu Gln Asn Glu Asp Arg Glu Arg Arg Gly Asp Arg Thr Ile Tyr Asp
      35             40             45

Tyr Val Arg Gly Asn Tyr Leu Asp Pro Arg Ser Cys Lys Val Ser Trp
      50             55             60

Asp Trp Lys Asp Pro Tyr Glu Val Gly His Ser Met Ala Phe Arg Val
      65             70             75             80

His Leu Phe Tyr Lys Asn Gly Gln Pro Phe Pro Ala His Arg Pro Val
      85             90             95

Gly Leu Arg Val His Ile Ser His Val Glu Leu Ala Val Glu Ile Pro
      100            105            110

Val Thr Gln Glu Val Leu Gln Glu Pro Asn Ser Asn Val Val Lys Val
      115            120            125

Ala Phe Thr Val Arg Lys Ala Gly Arg Tyr Glu Ile Thr Val Lys Leu
      130            135            140

Gly Gly Leu Asn Val Ala Tyr Ser Pro Tyr Tyr Lys Ile Phe Gln Pro
      145            150            155            160

Gly Met Val Val Pro Ser Lys Thr Lys Ile Val Cys His Phe Ser Thr
      165            170            175

Leu Val Leu Thr Cys Gly Gln Pro His Thr Leu Gln Ile Val Pro Arg
      180            185            190

Asp Glu Tyr Asp Asn Pro Thr Asn Asn Ser Met Ser Leu Arg Asp Glu
      195            200            205

His Asn Tyr Thr Leu Ser Ile His Glu Leu Gly Pro Gln Glu Glu Glu
      210            215            220

Ser Thr Gly Val Ser Phe Glu Lys Ser Val Thr Ser Asn Arg Gln Thr
      225            230            235            240

Phe Gln Val Phe Leu Arg Leu Thr Leu His Ser Arg Gly Cys Phe His
      245            250            255

Ala Cys Ile Ser Tyr Gln Asn Gln Pro Ile Asn Asn Gly Glu Phe Asp
      260            265            270

Ile Ile Val Leu Ser Glu Asp Glu Lys Asn Ile Val Glu Arg Asn Val
      275            280            285

Ser Thr Ser Gly Val Ser Ile Tyr Phe Glu Ala Tyr Leu Tyr Asn Ala

```

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

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

<210> 7

<211> 210

<212> PRT

<213> Homo sapiens

<400> 7

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

```
<210> 8
<211> 379
<212> PRT
<213> Homo sapiens
```

```

<400> 8
Met Gln Met Ser Glu Lys Lys Ala Tyr Met Leu Met His Glu Thr Ile
  1              5              10              15

Leu Gln Lys Lys Asp Glu Phe Pro Pro Ser Pro Arg Phe Ile Leu Arg
      20              25              30

Val Arg Arg Ser Arg Leu Val Lys Asp Ala Leu Arg Gln Leu Ser Gln
      35              40              45

Ala Glu Ala Thr Asp Phe Cys Lys Val Leu Val Val Glu Phe Ile Asn
  50              55              60

Glu Ile Cys Pro Glu Ser Gly Gly Val Ser Ser Glu Phe Phe His Cys
  65              70              75              80

Met Phe Glu Glu Met Thr Lys Pro Glu Tyr Gly Met Phe Met Tyr Pro
      85              90              95

Glu Met Cys Ser Cys Met Trp Phe Pro Ala Lys Pro Lys Pro Glu Lys
      100              105              110

Lys Arg Tyr Phe Leu Phe Gly Met Leu Cys Gly Leu Ser Leu Phe Asn
      115              120              125

```

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

<210> 9

<211> 947

<212> PRT

<213> Homo sapiens

<400> 9

Met Cys Val Asp Ser Leu Val Arg Ile Cys Ser Gly Leu Ser Tyr Gly
 1 5 10 15
 Arg Ile Arg Asn Ile Lys Ser Leu Ser Asp Ile Gln Ile Val Gln Val
 20 25 30

Ala Cys Gly Tyr Tyr His Ser Leu Ala Leu Ser Lys Ala Ser Glu Val
 35 40 45
 Phe Cys Trp Gly Gln Asn Lys Tyr Gly Gln Leu Gly Leu Gly Thr Asp
 50 55 60
 Cys Lys Lys Gln Thr Ser Pro Gln Leu Leu Lys Ser Leu Leu Gly Ile
 65 70 75 80
 Pro Phe Met Gln Val Ala Ala Gly Gly Ala His Ser Phe Val Leu Thr
 85 90 95
 Leu Ser Gly Ala Ile Phe Gly Trp Gly Arg Asn Lys Phe Gly Gln Leu
 100 105 110
 Gly Leu Asn Asp Glu Asn Asp Arg Tyr Val Pro Asn Leu Leu Lys Ser
 115 120 125
 Leu Arg Ser Gln Lys Ile Val Tyr Ile Cys Cys Gly Glu Asp His Thr
 130 135 140
 Ala Ala Leu Thr Lys Glu Gly Gly Val Phe Thr Phe Gly Ala Gly Gly
 145 150 155 160
 Tyr Gly Gln Leu Gly His Asn Ser Thr Ser His Glu Ile Asn Pro Arg
 165 170 175
 Lys Val Phe Glu Leu Met Gly Ser Ile Val Thr Glu Ile Ala Cys Gly
 180 185 190
 Arg Gln His Thr Ser Ala Phe Val Pro Ser Ser Gly Arg Ile Tyr Ser
 195 200 205
 Phe Gly Leu Gly Gly Asn Gly Gln Leu Gly Thr Gly Ser Thr Ser Asn
 210 215 220
 Arg Lys Ser Pro Phe Thr Val Lys Gly Asn Trp Tyr Pro Tyr Asn Gly
 225 230 235 240
 Gln Cys Leu Pro Asp Ile Asp Ser Glu Glu Tyr Phe Cys Val Lys Arg
 245 250 255
 Ile Phe Ser Gly Gly Asp Gln Ser Phe Ser His Tyr Ser Ser Pro Gln
 260 265 270
 Asn Cys Gly Pro Pro Asp Asp Phe Arg Cys Pro Asn Pro Thr Lys Gln
 275 280 285
 Ile Trp Thr Val Asn Glu Ala Leu Ile Gln Lys Trp Leu Ser Tyr Pro
 290 295 300
 Ser Gly Arg Phe Pro Val Glu Ile Ala Asn Glu Ile Asp Gly Thr Phe
 305 310 315 320
 Ser Ser Ser Gly Cys Leu Asn Gly Ser Phe Leu Ala Val Ser Asn Asp
 325 330 335
 Asp His Tyr Arg Thr Gly Thr Arg Phe Ser Gly Val Asp Met Asn Ala
 340 345 350

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

675					680					685					
Gly	Val	Ile	Cys	Gly	Leu	Ala	Ile	Tyr	Asn	Cys	Thr	Ile	Val	Asp	Leu
690					695					700					
His	Phe	Pro	Leu	Ala	Leu	Tyr	Lys	Lys	Leu	Leu	Lys	Lys	Lys	Pro	Ser
705					710					715					720
Leu	Asp	Asp	Leu	Lys	Glu	Leu	Met	Pro	Asp	Val	Gly	Arg	Ser	Met	Gln
				725					730					735	
Gln	Leu	Leu	Asp	Tyr	Pro	Glu	Asp	Asp	Ile	Glu	Glu	Thr	Phe	Cys	Leu
			740					745						750	
Asn	Phe	Thr	Ile	Thr	Val	Glu	Asn	Phe	Gly	Ala	Thr	Glu	Val	Lys	Glu
		755					760					765			
Leu	Val	Leu	Asn	Gly	Ala	Asp	Thr	Ala	Val	Asn	Lys	Gln	Asn	Arg	Gln
	770					775					780				
Glu	Phe	Val	Asp	Ala	Tyr	Val	Asp	Tyr	Ile	Phe	Asn	Lys	Ser	Val	Ala
785						790					795				800
Ser	Leu	Phe	Asp	Ala	Phe	His	Ala	Gly	Phe	His	Lys	Val	Cys	Gly	Gly
			805						810					815	
Lys	Val	Leu	Leu	Leu	Phe	Gln	Pro	Asn	Glu	Leu	Gln	Ala	Met	Val	Ile
		820						825					830		
Gly	Asn	Thr	Asn	Tyr	Asp	Trp	Lys	Glu	Leu	Glu	Lys	Asn	Thr	Glu	Tyr
		835					840					845			
Lys	Gly	Glu	Tyr	Trp	Ala	Glu	His	Pro	Thr	Ile	Lys	Ile	Phe	Trp	Glu
	850					855					860				
Val	Phe	His	Glu	Leu	Pro	Leu	Glu	Lys	Lys	Lys	Gln	Phe	Leu	Leu	Phe
865						870					875				880
Leu	Thr	Gly	Ser	Asp	Arg	Ile	Pro	Ile	Leu	Gly	Met	Lys	Ser	Leu	Lys
			885						890					895	
Leu	Val	Ile	Gln	Ser	Thr	Gly	Gly	Gly	Glu	Glu	Tyr	Leu	Pro	Val	Ser
		900						905					910		
His	Thr	Cys	Phe	Asn	Leu	Leu	Asp	Leu	Pro	Lys	Tyr	Thr	Glu	Lys	Glu
		915					920					925			
Thr	Leu	Arg	Ser	Lys	Leu	Ile	Gln	Ala	Ile	Asp	His	Asn	Glu	Gly	Phe
	930					935					940				
Ser	Leu	Ile													
945															

<210> 10
 <211> 1572
 <212> PRT
 <213> Homo sapiens

 <400> 10

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

325	330	335
Thr Ile Leu Gly Val Asn Ser Val Asn Gly Asp Leu Gly Ser Pro Ser		
340	345	350
Asp Asp Glu Asp Met Pro Gly Ser His His Asp Ser Gln Val Cys Ser		
355	360	365
Asn Gly Pro Val Ser Glu Asp Ser Ala Ala Asp Gly Thr Pro Lys His		
370	375	380
Ser Phe Arg Thr Ser Ser Thr Leu Glu Ile Asp Thr Glu Glu Leu Thr		
385	390	395
Ser Thr Ser Ser Arg Thr Ser Pro Pro Arg Gly Arg Gln Asp Ser Leu		
405	410	415
Asn Asp Tyr Leu Asp Ala Ile Glu His Asn Gly His Ser Arg Pro Gly		
420	425	430
Thr Ala Thr Cys Ser Glu Arg Ser Met Gly Ala Ser Pro Lys Leu Arg		
435	440	445
Ser Ser Phe Pro Thr Asp Thr Arg Leu Asn Ala Met Leu His Ile Asp		
450	455	460
Ser Asp Glu Glu Asp His Glu Phe Gln Gln Asp Leu Gly Tyr Pro Ser		
465	470	475
Ser Leu Glu Glu Glu Gly Gly Leu Ile Met Phe Ser Arg Ala Ser Arg		
485	490	495
Ala Asp Asp Gly Ser Leu Thr Ser Gln Thr Lys Leu Glu Asp Asn Pro		
500	505	510
Val Glu Asn Glu Glu Ala Ser Thr His Glu Ala Ala Ser Phe Glu Asp		
515	520	525
Lys Pro Glu Asn Leu Pro Glu Leu Ala Glu Ser Ser Leu Pro Ala Gly		
530	535	540
Pro Ala Pro Glu Glu Gly Glu Gly Gly Pro Glu Pro Gln Pro Ser Ala		
545	550	555
Asp Gln Gly Ser Ala Glu Leu Cys Gly Ser Gln Glu Val Asp Gln Pro		
565	570	575
Thr Ser Gly Ala Asp Thr Gly Thr Ser Asp Ala Ser Gly Gly Ser Arg		
580	585	590
Arg Ala Val Ser Glu Thr Glu Ser Leu Asp Gln Gly Ser Glu Pro Ser		
595	600	605
Gln Val Ser Ser Glu Thr Glu Pro Ser Asp Pro Ala Arg Thr Glu Ser		
610	615	620
Val Ser Glu Ala Ser Thr Arg Pro Glu Gly Glu Ser Asp Leu Glu Cys		
625	630	635
Ala Asp Ser Ser Cys Asn Glu Ser Val Thr Thr Gln Leu Ser Ser Val		
645	650	655

Asp Thr Arg Cys Ser Ser Leu Glu Ser Ala Arg Phe Pro Glu Thr Pro
 660 665 670
 Ala Phe Ser Ser Gln Glu Glu Glu Asp Gly Ala Cys Ala Ala Glu Pro
 675 680 685
 Thr Ser Ser Gly Pro Ala Glu Gly Ser Gln Glu Ser Val Cys Thr Ala
 690 695 700
 Gly Ser Leu Pro Val Val Gln Val Pro Ser Gly Glu Asp Glu Gly Pro
 705 710 715 720
 Gly Ala Glu Ser Ala Thr Val Pro Asp Gln Glu Glu Leu Gly Glu Val
 725 730 735
 Trp Gln Arg Arg Gly Ser Leu Glu Gly Ala Ala Ala Ala Ala Glu Ser
 740 745 750
 Pro Pro Gln Glu Glu Gly Ser Ala Gly Glu Ala Gln Gly Thr Cys Glu
 755 760 765
 Gly Ala Thr Ala Gln Glu Glu Gly Ala Thr Gly Gly Ser Gln Ala Asn
 770 775 780
 Gly His Gln Pro Leu Arg Ser Leu Pro Ser Val Arg Gln Asp Val Ser
 785 790 795 800
 Arg Tyr Gln Arg Val Asp Glu Ala Leu Pro Pro Asn Trp Glu Ala Arg
 805 810 815
 Ile Asp Ser His Gly Arg Ile Phe Tyr Val Asp His Val Asn Arg Thr
 820 825 830
 Thr Thr Trp Gln Arg Pro Thr Ala Pro Pro Ala Pro Gln Val Leu Gln
 835 840 845
 Arg Ser Asn Ser Ile Gln Gln Met Glu Gln Leu Asn Arg Arg Tyr Gln
 850 855 860
 Ser Ile Arg Arg Thr Met Thr Asn Glu Arg Pro Glu Glu Asn Thr Asn
 865 870 875 880
 Ala Ile Asp Gly Ala Gly Glu Glu Ala Asp Phe His Gln Ala Ser Ala
 885 890 895
 Asp Phe Arg Arg Glu Asn Ile Leu Pro His Ser Thr Ser Arg Ser Arg
 900 905 910
 Ile Thr Leu Leu Leu Gln Ser Pro Pro Val Lys Phe Leu Ile Ser Pro
 915 920 925
 Glu Phe Phe Thr Val Leu His Ser Asn Pro Ser Ala Tyr Arg Met Phe
 930 935 940
 Thr Asn Asn Thr Cys Leu Lys His Met Ile Thr Lys Val Arg Arg Asp
 945 950 955 960
 Thr His His Phe Glu Arg Tyr Gln His Asn Arg Asp Leu Val Gly Phe
 965 970 975

Leu Asn Met Phe Ala Asn Lys Gln Leu Glu Leu Pro Arg Gly Trp Glu
 980 985 990
 Met Lys His Asp His Gln Gly Lys Ala Phe Phe Val Asp His Asn Ser
 995 1000 1005
 Arg Thr Thr Thr Phe Ile Asp Pro Arg Leu Pro Leu Gln Ser Ser Arg
 1010 1015 1020
 Pro Thr Ser Ala Leu Val His Arg Gln His Leu Thr Arg Gln Arg Ser
 1025 1030 1035 1040
 His Ser Ala Gly Glu Val Gly Glu Asp Ser Arg His Ala Gly Pro Pro
 1045 1050 1055
 Val Leu Pro Arg Pro Ser Ser Thr Phe Asn Thr Val Ser Arg Pro Gln
 1060 1065 1070
 Tyr Gln Asp Met Val Pro Val Ala Tyr Asn Asp Lys Ile Val Ala Phe
 1075 1080 1085
 Leu Arg Gln Pro Asn Ile Phe Glu Ile Leu Gln Glu Arg Gln Pro Asp
 1090 1095 1100
 Leu Thr Arg Asn His Ser Leu Arg Glu Lys Ile Gln Phe Ile Arg Thr
 1105 1110 1115 1120
 Glu Gly Thr Pro Gly Leu Val Arg Leu Ser Ser Asp Ala Asp Leu Val
 1125 1130 1135
 Met Leu Leu Ser Leu Phe Glu Glu Glu Ile Met Ser Tyr Val Pro Pro
 1140 1145 1150
 His Ala Leu Leu His Pro Ser Tyr Cys Gln Ser Pro Arg Gly Ser Pro
 1155 1160 1165
 Val Ser Ser Pro Gln Asn Ser Pro Gly Thr Gln Arg Ala Asn Ala Arg
 1170 1175 1180
 Ala Pro Ala Pro Tyr Lys Arg Asp Phe Glu Ala Lys Leu Arg Asn Phe
 1185 1190 1195 1200
 Tyr Arg Lys Leu Glu Thr Lys Gly Tyr Gly Gln Gly Pro Gly Lys Leu
 1205 1210 1215
 Lys Leu Ile Ile Arg Arg Asp His Leu Leu Glu Asp Ala Phe Asn Gln
 1220 1225 1230
 Ile Met Gly Tyr Ser Arg Lys Asp Leu Gln Arg Asn Lys Leu Tyr Val
 1235 1240 1245
 Thr Phe Val Gly Glu Glu Gly Leu Asp Tyr Ser Gly Pro Ser Arg Glu
 1250 1255 1260
 Phe Phe Phe Leu Val Ser Arg Glu Leu Phe Asn Pro Tyr Tyr Gly Leu
 1265 1270 1275 1280
 Phe Glu Tyr Ser Ala Asn Asp Thr Tyr Thr Val Gln Ile Ser Pro Met
 1285 1290 1295
 Ser Ala Phe Val Asp Asn His His Glu Trp Phe Arg Phe Ser Gly Arg

1300	1305	1310
Ile Leu Gly Leu Ala Leu Ile His Gln Tyr Leu Leu Asp Ala Phe Phe 1315	1320	1325
Thr Arg Pro Phe Tyr Lys Ala Leu Leu Arg Ile Leu Cys Asp Leu Ser 1330	1335	1340
Asp Leu Glu Tyr Leu Asp Glu Glu Phe His Gln Ser Leu Gln Trp Met 1345	1350	1355 1360
Lys Asp Asn Asp Ile His Asp Ile Leu Asp Leu Thr Phe Thr Val Asn 1365	1370	1375
Glu Glu Val Phe Gly Gln Ile Thr Glu Arg Glu Leu Lys Pro Gly Gly 1380	1385	1390
Ala Asn Ile Pro Val Thr Glu Lys Asn Lys Lys Glu Tyr Ile Glu Arg 1395	1400	1405
Met Val Lys Trp Arg Ile Glu Arg Gly Val Val Gln Gln Thr Glu Ser 1410	1415	1420
Leu Val Arg Gly Phe Tyr Glu Val Val Asp Ala Arg Leu Val Ser Val 1425	1430	1435 1440
Phe Asp Ala Arg Glu Leu Glu Leu Val Ile Ala Gly Thr Ala Glu Ile 1445	1450	1455
Asp Leu Ser Asp Trp Arg Asn Asn Thr Glu Tyr Arg Gly Gly Tyr His 1460	1465	1470
Asp Asn His Ile Val Ile Arg Trp Phe Trp Ala Ala Val Glu Arg Phe 1475	1480	1485
Asn Asn Glu Gln Arg Leu Arg Leu Leu Gln Phe Val Thr Gly Thr Ser 1490	1495	1500
Ser Ile Pro Tyr Glu Gly Phe Ala Ser Leu Arg Gly Ser Asn Gly Pro 1505	1510	1515 1520
Arg Arg Phe Cys Val Glu Lys Trp Gly Lys Ile Thr Ala Leu Pro Arg 1525	1530	1535
Ala His Thr Cys Phe Asn Arg Leu Asp Leu Pro Pro Tyr Pro Ser Phe 1540	1545	1550
Ser Met Leu Tyr Glu Lys Leu Leu Thr Ala Val Glu Glu Thr Ser Thr 1555	1560	1565
Phe Gly Leu Glu 1570		

<210> 11
 <211> 706
 <212> PRT
 <213> Homo sapiens

 <400> 11

35/53

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

325					330					335						
Arg	Gln	Gly	Ser	Lys	Phe	Arg	Arg	Asn	Val	Ser	Thr	Leu	Leu	Ile	Glu	
340					345					350						
Leu	Gly	Phe	Gln	Ile	Lys	Lys	Lys	Thr	Lys	Arg	Leu	Tyr	Ile	Asn	Lys	
355					360					365						
Ala	Asn	Ile	Trp	Asn	Ser	Ala	Leu	Asp	Ala	Phe	Arg	Asn	Arg	Asn	Phe	
370					375					380						
Asn	Pro	Ser	Tyr	Ala	Ile	Glu	Val	Ala	Tyr	Val	Ile	Glu	Asn	Asp	Asn	
385					390					395					400	
Phe	Gly	Ser	Glu	His	Pro	Gly	Ser	Lys	Gln	Glu	Phe	Leu	Ser	Leu	Leu	
405					410					415						
Met	Gln	His	Leu	Glu	Asn	Ser	Ser	Leu	Phe	Glu	Gly	Ser	Leu	Ser	Lys	
420					425					430						
Asn	Leu	Ser	Leu	Asn	Ser	Gln	Ala	Leu	Lys	Glu	Asn	Leu	Tyr	Tyr	Glu	
435					440					445						
Ala	Gly	Lys	Met	Leu	Ala	Ile	Ser	Leu	Val	His	Gly	Gly	Pro	Ser	Pro	
450					455					460						
Gly	Phe	Phe	Ser	Lys	Thr	Leu	Phe	Asn	Cys	Leu	Val	Tyr	Gly	Pro	Glu	
465					470					475					480	
Asn	Thr	Gln	Pro	Ile	Leu	Asp	Asp	Val	Ser	Asp	Phe	Asp	Val	Ala	Gln	
485					490					495						
Ile	Ile	Ile	Arg	Ile	Asn	Thr	Ala	Thr	Thr	Val	Ala	Asp	Leu	Lys	Ser	
500					505					510						
Ile	Ile	Asn	Glu	Cys	Tyr	Asn	Tyr	Leu	Glu	Leu	Ile	Gly	Cys	Leu	Arg	
515					520					525						
Leu	Ile	Thr	Thr	Leu	Ser	Asp	Lys	Tyr	Met	Leu	Val	Lys	Asp	Ile	Leu	
530					535					540						
Gly	Tyr	His	Val	Ile	Gln	Arg	Val	His	Thr	Pro	Phe	Glu	Ser	Phe	Lys	
545					550					555					560	
Gln	Gly	Leu	Lys	Thr	Leu	Gly	Val	Leu	Glu	Lys	Ile	Gln	Ala	Tyr	Pro	
565					570					575						
Glu	Ala	Phe	Cys	Ser	Ile	Leu	Cys	His	Lys	Pro	Glu	Ser	Leu	Ser	Ala	
580					585					590						
Lys	Ile	Leu	Ser	Glu	Leu	Phe	Thr	Val	His	Thr	Leu	Pro	Asp	Val	Lys	
595					600					605						
Ala	Leu	Gly	Phe	Trp	Asn	Ser	Tyr	Leu	Gln	Ala	Val	Glu	Asp	Gly	Lys	
610					615					620						
Ser	Thr	Thr	Thr	Met	Glu	Asp	Ile	Leu	Ile	Phe	Ala	Thr	Gly	Cys	Ser	
625					630					635					640	
Ser	Ile	Pro	Pro	Ala	Gly	Phe	Lys	Pro	Thr	Pro	Ser	Ile	Glu	Cys	Leu	
645					650					655						

His Val Asp Phe Pro Val Gly Asn Glu Cys Asn Asn Cys Leu Ala Ile
 660 665 670

Pro Ile Thr Asn Thr Tyr Lys Glu Phe Gln Glu Asn Met Asp Phe Thr
 675 680 685

Ile Arg Asn Thr Leu Arg Leu Glu Lys Glu Glu Ser Ser His Tyr Ile
 690 695 700

Gly His
 705

<210> 12
 <211> 6200
 <212> ADN
 <213> Homo sapiens

<400> 12
 gggttttttagg cctggcgcgc atggcgctctc cttctagaaa ctcccagagc cgacgccgggt 60
 gcaaggagcc gctccgatac agctacaacc cggaccagtt ccacaacatg gacctcaggg 120
 gcggccccca cgatggcgctc accattcccc gctccaccag cgacactgac ctggtcacct 180
 cggacagccg ctccacgctc atggtcagca gctcctacta ttccatcggg cactctcagg 240
 acctggatcat ccaactgggac ataaaggagg aagtggagcg tggggactgg attggcatgt 300
 acctcattga tgaggtcttg tccgaaaact ttctggacta taaaaaccgt ggagtcaatg 360
 gttctcatcg gggccagatc atctggaaga tcgatgccag ctctactttt gtggaacctg 420
 aaactaagat ctgcttcaaa tactaccatg gagtgaagtgg ggccctgcga gcaaccaccc 480
 ccagtgtcac ggtcaaaaac tggcagctc ctatttttaa aagcattggg gctgatgaga 540
 ccgtccaagg acaaggaagt cggaggtga tcagcttctc tctctcagat ttccaagcca 600
 tgggggttgaa gaaagggatg tttttcaacc cagaccctta tctgaagatt tccattcagc 660
 ctgggaaaca cagcatcttc cccgccctcc ctccaccatg acaggagagg agatccaaga 720
 tcataggcaa caccgtgaac cccatctggc aggcagagca attcagtttt gtgtccttgc 780
 ccaactgacgt gctggaaatt gaggtgaagg acaagtttgc caagagccgc cccatcatca 840
 agcgcttctt gggaaagctg tcgatgcccg ttcaaagact cctggagaga cacgccatag 900
 gggatagagt ggtcagctac acacttgccc gcaggcttcc aacagatcat gtgagtggag 960
 agctgcaatt ccgatttgag atcacttctc ccatccaccc agatgatgag gagatttccc 1020
 tgagtaccga gcctgagtca gccc aaatc aggacagccc catgaacaac ctgatggaaa 1080
 gcggcagtg ggaacctcgg tctgaggcac cagagtcctc tgagagctgg aagccagagc 1140
 agctgggtga gggcagtgct cccgatcgtc cagggaacca aagcatagag ctttccagac 1200
 cagctgagga agcagcagtc atcacggagg caggagacca gggcatgggt tctgtgggac 1260
 ctgaaggggc tggggagctc ctggcccagg tgcaaaagga catccagcct gccccagtg 1320
 cagaagagct ggccgagcag ctggacctgg gtgaggaggc atcagcactg ctgctggaag 1380
 acggtgaagc cccagccagc accaaggagg agcccttgga ggaggaagca acgaccaga 1440
 gccgggctgg aagggaaaga gaggagaagg agcaggagga ggagggagat gtgtccaccc 1500
 tggagcaggg agagggcagg ctgcagctgc gggcctcggg gaagagaaaa agcaggccct 1560
 gctccttgcc tgtgtccgag ctggagacgg tgatcgctc agcctgcggg gaccccgaga 1620
 ccccgaggac acactacatc cgcattccaca ccctgctgca cagcatgccc tccgcccagg 1680
 gcggcagcgc ggcagaggag gaggacggcg cggaggagga gtccaccctc aaggactcct 1740
 cggagaagga tgggctcagc gaggtggaca cgggtggccg tgaccctgtc gccctggaag 1800
 aggacagaga agagcccag ggggctactc caggcacggc gcaccctggc cactccgggg 1860
 gccacttccc cagcctggcc aatggcgcg cccaggatgg cgacacgcac cccagcaccg 1920
 ggagcgagag cgaactccagc cccaggcaag gcggggacca cagttgagag ggctgtgacg 1980
 cgctctgctg cagccctcgt tgctacagct cctcgtgcta cagcacgtcc tgctacagca 2040
 gctcgtgcta cagcgctcgt tgctacagcc cctcctgcta caacggcaac aggttcgcca 2100
 gccacacgcg cttctcctcc gtggacagcg ccaagatctc cgagagcacg gtcttctcct 2160
 cgcaagacga cgaggaggag gagaacagcg cgttcagatc ggtacccgac tccatgcaga 2220
 gccctgagct ggaccggag tccacgaacg gcgctgggac gtggcaagac gagctggccg 2280
 cccctagcgg gcacgtggaa agaagcccgg aaggtctgga atccccctg gcagggtccaa 2340
 gcaatcggag agaaggtgaa tgcctatac tccataattc ccagccagta agccagcttc 2400

cttccttgag	gcctgaacat	catcactacc	caacaatcga	tgagcctctt	ccaccaaaact	2460
gggaagctcg	aattgacagc	cacgggcggg	tcttttatgt	ggaccacgtg	aaccgcacaa	2520
ccacctggca	gcgtccgacg	gcagcagcca	ccccggatgg	catgcggaga	tcgggggtcca	2580
tccagcagat	ggagcaactc	aacaggcggt	atcaaaaacat	tcagcgaaac	attgcaacag	2640
agaggtccga	agaagattct	ggcagccaaa	gctgcgagca	agccccagca	ggaggaggcg	2700
gaggtggagg	gagtgactca	gaagccgaat	cttcccagtc	cagcttagat	ctaaggagag	2760
aggggtcact	ttctccagtg	aactcacaaa	aaatcacctt	gctgctgcag	tccccagcgg	2820
tcaagttcat	caccaacccc	gagttcttca	ctgtgctaca	tgccaattat	agtgcctacc	2880
gagtccttac	cagtagcacc	tgcttaaagc	acatgattct	gaaagtccga	cgggatgctc	2940
gcaattttga	acgtaccag	cacaaccggg	acttggtgaa	tttcatcaac	atgttcgcag	3000
acactcggct	ggaactgccc	cggggctggg	agatcaaaac	ggaccagcag	ggaaagtctt	3060
ttttcgtgga	ccacaacagt	cgagctacca	ctttcattga	ccccgaatc	cctcttcaga	3120
acggtcgtct	tcccaatcat	ctaactcacc	gacagcacct	ccagaggctc	cgaagttaca	3180
gcgctggaga	ggcctcagaa	gtttctagaa	acagaggagc	ctctttactg	gccaggccag	3240
gacacagctt	agtagctgct	attcgaagcc	aacatcaaca	tgagtcattg	ccactggcat	3300
ataatgacaa	gattgtggca	tttcttcgcc	agccaaacat	ttttgaaatg	ctgcaagagc	3360
gtcagccaag	cttagcaaga	aaccacacac	tcaggagaaa	aatccattac	attcggactg	3420
agggtaatca	cgggcttgag	aagttgtcct	gtgatgcgga	tctgggtcatt	ttgctgagtc	3480
tctttgaaga	agagattatg	tcctacgtcc	ccctgcaggc	tgctttccac	cctgggtata	3540
gcttctctcc	ccgtgtttca	ccctgttctt	cacctcagaa	ctccccaggt	ttacagagag	3600
ccagtgaag	agccccctcc	ccctaccgaa	gagactttga	ggccaagctc	cgcaatttct	3660
acagaaaact	ggaagccaaa	ggatttggtc	agggctccgg	gaaaattaag	ctcattattc	3720
gccgggatca	tttggtggag	ggaaccttca	atcagggtgat	ggcctattcg	cggaaagagc	3780
tccagcgaaa	caagctctac	gtcacctttg	ttggagagga	gggcctggac	tacagtggcc	3840
cctcgcggga	gttcttcttc	cttctgtctc	aggagctctt	caacccttac	tatggactct	3900
ttgagtactc	ggcaaatgat	acttacacgg	tgcatatcag	ccccatgtcc	gcattttagt	3960
aaaaccatct	tgagtggttc	aggttttagcg	gtcgcacctt	gggtctggct	ctgatccatc	4020
agtaccttct	tgacgttttc	ttcacgaggc	ccttctacaa	ggcactcctg	agactgccct	4080
gtgattttgag	tgacctggaa	tatttgatg	aggaattcca	ccagagtttg	cagtggatga	4140
aggacaacaa	catcacagac	atcttagacc	tcactttcac	tgtaaatgaa	gaggtttttg	4200
gacaggtcac	ggaaggggag	ttgaagtctg	gaggagccaa	cacacagggtg	acggagaaaa	4260
acaagaagga	gtacatcgag	cgcatggtga	agtggcggtg	ggagcgcggc	gtggtacagc	4320
agaccgaggc	gctggtgcgc	ggcttctacg	aggttgtaga	ctcagaggctg	gtgtccgtgt	4380
ttgatgccag	ggagctggag	ctggtgatag	ctggcaccgc	ggaaatcgac	ctaaatgact	4440
ggcggaataa	cactgagtac	cggggagggt	accacgatgg	gcatcttgtg	atccgctggt	4500
tctgggctgc	ggtggagcgc	ttcaataatg	agcagaggct	gagattactg	cagtttgtca	4560
cgggaacatc	cagcgtgccc	tacgaaggct	tcgcagccct	ccgtgggagc	aatgggcttc	4620
ggcgcttctg	catagagaaa	tgggggaaaa	ttacttctct	ccccaggga	cacacatgct	4680
tcaacgcact	ggatcttoca	ccgtatccct	cgtactccat	gttgatgaa	aagctgttaa	4740
cagcagtaga	ggaaccagc	acctttggac	ttgagtgagg	acatggaacc	tcgcctgaca	4800
ttttcctggc	cagtgcacac	accttctctg	ggatgatccc	cttttccctt	tcccttaatc	4860
aactctctct	tgattttggt	attccatgat	ttttattttc	aaaccaaato	aggattgaca	4920
aaagctgtgc	atgaagaact	gccttctctt	aagatctaac	cttcaggctt	ctctcctctg	4980
ttttcaatga	actgctagcc	tgatgcaat	attaaaaaac	agctgtctca	aggtctgtgt	5040
atatctccac	atacctccat	tactaacaat	gaaatatgaa	tgcaagttaa	gctacacttg	5100
accaaattggt	aataaattgt	tacttccatt	tctatcattg	aagggaataat	gtgagcatta	5160
agcactccag	gctttcatat	gcccattgtc	tctgagcaga	gccaccattt	tttataattt	5220
ctaataacca	actccagAAC	taggagctga	tcaactcttt	gttttctctt	ccatctactt	5280
ttcctgtgct	ataatatcca	tccaaaggac	aacagtggca	aagctgaaat	ttttatacat	5340
tcaactcatg	attcacatgt	ggcatcagtc	ccatcagccg	gaactagcct	agacatacgg	5400
tgcaaatatg	acacttctaa	cgattaacaa	cagcaagaaa	acacctgctg	ctgatgcaat	5460
gcaatgcac	ccaatggttg	tggggattgt	gggtccaact	caagagaagt	ttaggagggg	5520
gagcatccct	agtgaatact	cacaccacaa	gaaggacaaa	cttgtgcaca	tgtccaagaa	5580
agaaagcttc	ttgattgagg	tagcatgaag	gatgaggctt	cagcccccat	tgtcttatgt	5640
agaatgtggc	aatgccaact	ggagaaaggg	aagaaggaca	tattaccttg	gtttgaatcc	5700
ctgagttctg	tactgttctg	ttttgtttag	tctagccaca	gttcttcaca	aaggaaaaaa	5760
aaatgtgtag	atgataccat	gacttttgtt	aaagccatga	cttttgtttg	cttggcagac	5820
aaaccctttt	tttaaaactt	tgatattttt	ttttcacatt	tttttccctt	ttcctttctt	5880
aatcatggag	ttcaagttcc	tttgcatctg	attgtccatc	gggaccacac	taggaagctg	5940
cagagagtga	tggtgcttgt	tagggatcaa	gggcaacata	gtacttctcc	ttcaccata	6000
gtaatcctcc	tggggcagaa	acataacacc	ccaaaggcac	gttgatttgt	atcaaaataa	6060

atatccagtt	tcttttagca	ttcagtgaaa	acatatctca	gaaaacttca	tgttgtcaga	6120
aaaacagctg	caggctccaa	agacagccta	acctctcaac	tacatttgaa	ataaaccxaa	6180
ccataatggt	aaaaaaaaaa					6200

<210> 13

<211> 7084

<212> ADN

<213> Homo sapiens

<400> 13

gaaaaccttt	gggcttcatg	gagtcgtctt	agatgttgat	tcagtgaatg	aactgggtgca	60
ggtagaaacg	tacctccgca	gtgaagggtg	gctgggtgca	tactgggtatc	ctattgacat	120
gttggaagg	ccccagcag	gctaccgaag	gactgccacc	aatgggctgg	tcacactgga	180
caataccaac	cttcaaattc	acagggagct	gctgcgctgc	gaggctgcgc	tggccaggct	240
gtactgccgc	atggccctgc	tcaatatctt	cgccccgaag	ttgcctcact	tgttcactcg	300
cctcttccac	atccctgcca	tccgggacat	taccctggag	cacctgcaac	tgctgtccaa	360
tcagctcctc	gcacctcctc	tcccagacgg	caccatcagc	tccagctcga	tccctctggc	420
gcagtcttta	cagcattgca	tccattccca	gaactgctcc	gccacggacc	tcttttacca	480
gggcaactcc	agacagtgga	gagagtggct	caacgtggcc	atcaccggga	ccctgcacca	540
gggcgaggag	agccttttag	agctgacgaa	acagatctgc	tctttcctgc	agacagcacc	600
agagcagttc	ccctccgaag	agttcccaat	ttccgaatcc	aaagtcaaca	tggacgtgaa	660
tttccccggg	gcagcttttg	ttgttggtgc	ttgtaaagaa	agtcaatctg	gattccgcaa	720
agactcctct	ctgtacaagg	caccgtgggc	acgggtgctc	gtatatggcc	tcgccacaa	780
agtgaagcga	aatggccagc	tgaacctcat	cgaggccgcc	tgttaccgcg	gggacgcgtc	840
cccagccaac	actgggcttg	cacctcccc	caccgctgac	cagtaccctc	ctgtggctct	900
ctccacagac	agggtccaca	tcaaactggg	ggtgtctcca	cctcctggag	cagtccctgt	960
gttacattcc	ctgcccctgg	agttcccact	ggctatggcc	ttcgcagagc	agctgctgtc	1020
ctggaaatca	gaggacagtg	aagggaagtc	cgaagatgag	cctgacacca	ttccgacatc	1080
cgctcctcct	caggtgggtg	agctgctagg	aaacttcttg	tggaccacgg	acatggcagc	1140
ctgcgtgaag	gagcttggtt	tccatctcct	ggcagagctc	ctgcgcacgg	tgcacacct	1200
ggagcagagg	cggcaccg	ctggcctgtc	ctcctcaatc	gccctccagc	tgaacccctg	1260
cctggccatg	ctgatggcct	tgcagtcgga	gctccacaag	ctgtacgacg	aggagacgca	1320
gaactgggtc	tcaggcggcg	cctgcggggg	ctccgggggg	gcggcggccg	gtgaccaggg	1380
caggttctct	acgtattttc	atgcactcat	ggaagggtgc	ctggctgtgg	ccgaagtgc	1440
cctgcctact	aacatgagtg	tcacagccag	tggggtgacc	tcagcgaccg	ccccaaatct	1500
cagtgaactg	tctcctcctc	cctcgtcctc	cccaggacag	acccacaga	gtcccagcct	1560
cctctccaag	aggaaaaaag	tcaagatgaa	gcgggaaaag	gcctcctcgt	cgggcaagcg	1620
ccagtcttcc	cgcaccgtgg	actcggaccc	caccgtgctc	agcatcggag	gcagcaagcc	1680
cgaggacatg	ctgtggttcc	accgcgcact	cacctgctc	atcatcctcc	gccacctcac	1740
caggaaggac	ccacaggggc	tgggcgtgac	gagtgacgcc	atcgccgatg	cctgccaggc	1800
cctggtgggc	cccaccgccc	acagccgctt	gctgggtgat	tccgggatcc	ccaccacct	1860
ggacgagggc	gtagtcaag	gcgccatccg	caaggcctgc	aacgcccacg	gcggggtctt	1920
caaagacgag	atctacatcc	cgctgcagga	agaagacacc	aagaagccaa	aagacaaggc	1980
cgagggcggg	gacgggaaag	ttgagccgca	gaagacactg	gccttccccg	gcacagacag	2040
catggaggtc	agcacgtcca	gcagcctgac	ccccgccatg	agcatcagcg	cctccgcctc	2100
caccagccag	gcctccatct	gcagctcgca	gggcatctcc	caaaccgtca	gcgacctctc	2160
tgtggatccg	ctgcctgccc	gcctcgagct	gcccacccct	ccgggctgt	tggagcccca	2220
cgcggtgtcc	agccaggaaa	gcctggacat	ttcctgtgct	agcaccggca	gcctgggcag	2280
cctgggcagc	ctgggggagc	ccctggacaa	tgcagagacg	gcctcggtgt	cggacatggg	2340
ctccatgtac	acagtcactt	ccctggacaa	ccagccctc	gccgcccggc	ccatcaaagg	2400
cttcgcagtc	gtggagataa	gatcccgcagc	caaaattgag	aaaattcgag	caagtttatt	2460
taacaataat	gacttgattg	gtttgtcaag	tttggtggt	gaagatgaat	tgatggaaat	2520
gtcaacagaa	gagattetaa	ccgtgtctgt	agtaaatacag	agtctctttg	atactcaagg	2580
gagcccagg	ttagaagatt	atttcaatga	taagtcaatt	aaaggggaga	agctgggtcc	2640
cggagccaga	gaggttctga	cggagatatt	taagagttgt	gcccattcag	agcagacgct	2700
gagcctgaca	ccagcgaagc	ccatcagagt	ctctgacatt	tatcttagca	aagagcagat	2760
caactcccag	accccaggca	acctcctcca	cctcttcttc	accaatgtcc	ggccgccaaa	2820
aaaggtgctg	gaggatcagc	tcacccagat	cctgaggaag	tatggcgtgc	caaagcccaa	2880
gtttgacaag	agcaagtaca	gcaaggccgg	gaaggagcag	caccccgtag	aggtggtgag	2940
caccaagcgg	cctatcacca	agccgcccgc	caaggacaag	gctgtgctca	acagcgtcag	3000

caggactgcc ttaagtgaga agaagccaac tgtgaagcca aaatcaccag aaaagagcaa 3060
accagatgaa aaggaccag aaaagtcacc taccaaaaaa caagaagtcc ctgaggaaaa 3120
atacctgacg ctggaaggat ttcacaaatt tgttattgac cgagccaggc aagatatccg 3180
tagcgtctgg agggcgatct tgtcctgtgg ttacgatctt cactttgaga ggtgcgcgatg 3240
catcgatgtc cgacatgcac agaaggcctc aagaaagtgg accctggaga tggacgtggc 3300
acttgtgcag tacatcaacc agctatgccg ccacctcgcc atcacaccg caccgctcca 3360
tccccatgag gtgtacctgg accccgcgga tgctgctgac ccagagtgg cctgtctcct 3420
gaacgtgccc atcgagagcc tgcgcctgcg cttcgccttg ctgcagtccc tcaacaccac 3480
actggagacc ttcttcctgc ccctgggtgga gctgcgccag acaccatgt ataccacag 3540
catcgccgcc ctgctgaagg aggccaaagg gctgatcttc tatgacacga aggtgaccgt 3600
gatgaatcga gtgctgaatg ccactgtgca gaggacagcg gaccacgcgg cccctgagat 3660
caccttgga cactggaaa ttgtgggagg ggaaatcaga gcttctgaaa actcctactt 3720
ctgtcaggct gccaggcagc tggcctcagt gccgtcgtct cagctctgcy tcaagctggc 3780
cagtggcgtg gacccacat atgccttcaa catccgcttc actggcgagg aggtccatgg 3840
caccagcggc tctttccgcc acttctctgtg gcaggtgtgt aaggagctgc agagttcctc 3900
gctgtcgtg ctgctgctgt gccccagctc agctgtcaat aagaacaagg gcaagtatat 3960
cctgaccccg agcccatca cctacgggga ggagcagctg ctgcacttcc tggggcagct 4020
gctggggatt gcaattcggg cagacgtccc gctgccctg gacctcctgc cctccttctg 4080
gaagacgctg gtgggcgagc ccttggaccc tgagcaagac ctgcaggaag cggatatcct 4140
cacctacaat tacgtcaaga agtttgagag catcaatgat gagaccgagc tggaggccct 4200
gtgcgctgag atgcctccc agcactggc caggagagc cctgacagcc ccaacaagcc 4260
ctgctgcagg ttccactacc tgaccatgac gggcgaggag gtggagctgt gcagccgggg 4320
ccgacatc tctgtggcgt gggagaacaa ggacatctac gggcagcca tccggagcct 4380
gcggtcgcgg gagctgcaga atgtggagtg cgtgacggcc gtgcggggcg gcctgggctc 4440
catcatcccc ctgcagctgc tgaccatgct cagccactg gagatggagc tgcgcacctg 4500
cggcctcccc tacatcaacc tcgagttcct caaggccac accatgtacc aagtggggct 4560
gatggagacg gaccagcaca tcgagttcct ctggggggcc ctggagatgt tcaaccagga 4620
ggagctgtgc aagttcatca agtttgctg caaccaggag cgcacccgt tcacctgccc 4680
ctgcaaagat gggggctccc aactgccc tgtgccccg taccatga agatcgcccc 4740
cccagatggc acagcaggtt cccagactc tcgtacatc cgcgtggaga cctgcatgtt 4800
catgatcaag ctccccagt actcctctct ggaatcatg ctggagaaac ttctgtgtgc 4860
cattcactac cgtgaagacc cctcagtggt ctgatgggag ggagccccag aattaggctg 4920
tcaactgagg accactctg ctggcttggg aagccaccac tgcggcccg cctccaggg 4980
ccctgcgtga ggagttggca acattttgct ttccaaact ttcgtccaca ttccagggcc 5040
tcctggaaga cttaaccttt tgtctttgta cgtttcgtga tgggttgggt cttttgctgc 5100
ctgtttgtgg tctatttgta ggatagtta gttcccagac agtttgtgtc taatttgatc 5160
ttcttgga ctgtccctca tggccaccag attctaagc cataaggccc cagctggttc 5220
cccatcacag aaccacccag caggaagccg ggggtacaat gacctcgcc acacagagcg 5280
ctcgtgcttt cagcagagcg cctccacaga caagcctgtg gccagccctg cctccagcc 5340
tcgagctgga gttgctccag aggcagcccc agcaggcctc cctccacata gaggtggctg 5400
ggccaattgc ccaaacacgt gagacacaca gggagcttta tgtgttcac ttctgtcttt 5460
agaacttgct cccggtcttg gttttctga ggaaaagccc acagtgggtc ccctgcccct 5520
cccatacctc cctatctgcc gtctgtctct ttgggcccgt gtctgacagt tttcatctga 5580
ggcctcctcg ggagcaggag acttaagggt tcatcccag cttagttcac agagcctagc 5640
tgggtgcac ccagccgcag ggggtgagaa gacaccaga gatgggggac aatgctcagg 5700
tgggggccc ctccgagact ctggttgaga ggaaagcaca tgattcaagc ctaccacaca 5760
aaaaaggagc aaaaaatgct cagaaatggc ccttcctccc tgcagtgcg taggacaagg 5820
gaaagggcca ccccatttt caaaaagagt gactgaggtg ctctgctcaa tttggggctc 5880
gctttagtga cccagacact caagtccaga tggagactcc aggggggtcc ttctgcccct 5940
ctggccagcc ccacccagca ctccccatca cccatgtaa gtgtgtatgc cagggtgag 6000
ctgcctccag gtcacccac tgcactcttg ccccaaatag gtctttgtcc agtgggtactt 6060
tgttcatgag cagcacaggc catcttgttg gacactgtca ctttgggccc tgccaggtct 6120
cacagaatca catcgtgtga tttattgtat aatattgtga atattggaag ttttgagtga 6180
gtgctgtcta gaaggttttg gagtgtggat gctggtccaa gaaagaagat gtgccctggc 6240
ctcttaggaa atgcatcaag gaccattata gaggcgtag catccatctc ctccctgtaa 6300
gacctagggt taatcatgac agctgcatca gttgtatttg atgcaagtc actctggagg 6420
tgagtggcgg caagcagttc cctgctttg aaagaagatg gtcagccgc tgccctccca 6480
gcgccacag ctcttttgag tgtggggggg cacagccagg atgtattcct gacaataagg 6540
atccttttat tttcatttct gcataatgaa agagcctttc gatatttag aagtggtttt 6600
tcctccctac atctttaaat tcttgggtct ctttagtct catccctga atggacttgt 6660

```

cagctacctc cccaggagga gttcagaagc ttctgttggc catcacttga gtagatgaca 6720
agctgtaagg tatctgctgc taacgtggct gcttctccga gagcacattg aggtccagac 6780
agatccaaac cttggcccta aaaaatgtgt gctgtaaagt tacttgatgt atatttatta 6840
taattattta tggttgcatt taacaaactt tcattcaat ctgatgctat ttacaccatg 6900
ctgtgtaaag gaagctcact aaaggaaaaa agtcacacca ataaataacc ttcatcctaat 6960
tttgatttac cttaagagtt gtgtgatcat ctactcttgc tgagcaagtt caaaggctgg 7020
catgcttgga tggctctgat atttaagtgc tgccaagtgg ctaggaatta aagtgtttct 7080
aatt 7084

```

<210> 14

<211> 5321

<212> ADN

<213> Homo sapiens

<400> 14

```

tagaattttc agtgaaggta tctgtcagta tttatggcag ctgggtgctt cttgcccttc 60
tttaagcagt tcaggggactc caaatctgtc aaccgttgtg agtatggaag agccctcatt 120
acgggtgaga gggatgctta ctttaacttc gaaggagggt ataattttctc ttaaaaaagg 180
aaacattgta ctttaaccga ctacctaagt catagtactg gttgagttag tgaactagggt 240
ttttgttgat ttagtagtg gtggttttgt tggttttgtt gtgtttttct gcttttttgt 300
tgaacgtgta aatgaaggcc tgaggtgatt taattatttg caaatcagc aagtaaaata 360
ttttcaaaaa gagaatatta cagaaatctg aagaaaggct gaaagtggca gtgaagggtat 420
ttgttagata ataattggga agaatttcct ggcagaatag ttcaaacatt ttctaagggt 480
ccaaatatgg ttgtataatc tccttccatg gcggtcttta aaagaaagat tatttctttc 540
cttgaaatat tttaaagggt agtctttctg tagcacatgg ttagtattgt agcattctgt 600
aatcatgtca gattcaaaaa gtgagataaa cttttatatt tatttgcaga aggatagcta 660
atcattctca ataagaatat ttaaaatagt gtaggatgaa gggattgttt aaccactgta 720
ccacttataa actatttaac tttatttggg attaaaaata tcctagcagg tttttttttt 780
tgcttgtttg tttgtttttt tttttttttt ggagacagta tctcactctt tcaccaggc 840
tggagtatag tgacatgatc acagctcact gtagccttga cttccgagac tcaagcaatc 900
ctcccccttc aacctctca gtaactggga ctacaagtgc atgctgtgt acccagctag 960
aataataata ggaaatagtt tattctctcc tcactttttt ttttttttta atagaaagggt 1020
gttaatcttt acttggaatt aagaatttgc tactcagtat tttttttttt ttttttgaga 1080
tggagtctcg cctgtgcacc caggctggag tgcagtgggt tgatcttagc tctactgcaac 1140
ctctgcctcc tgggttcaag tgatcgctct gcctcagcct cccgagtagc tgggaattata 1200
gggtgtgtgc accatgtctg gctaattttt tgtattttta gtagagatgg gatttcaatt 1260
caccatgttg ccaggctga tctcgaattc ctgacctcaa gtgatctacc cacctcggcc 1320
tcccaaagtg ctgggattac aggcgtgagc caccatctcc ggcctgctag tcagtatttt 1380
aatagaagtt tggtaattgt atctcttgtt ttaaaaaaac actaattatt agaaatggat 1440
tgtgtctgta agtatatgta tatacataaa tatgccacat tttgtaattt cttgtcaaaa 1500
attatccaaa ttctaggaat tatatgtaat gaaaggagca cagtaaaaaa gctgttcatt 1560
aggttcttta gggattctca ttttttaatt gaaaaaagac atttgtttaa ttgaaagttg 1620
ttgacattga tcatattaat acagaaaatg ttgagtgtat atacttaggg gttctcattt 1680
tttaactgaa aaaatcttat ttgtttaatt gaaagtgggt gacattgatc atattaatac 1740
agtaaatgtt aagggtgatt atatatagta gtaatgtgtg cttttaaaac taaacatttg 1800
aaaattaaaa gaagaaagat tgtgaatttt ttgttagatt tagtttttcc tttttattac 1860
aaggaataca tacatacaaa ttaaattatt gccataaatg aagaaattgt catttatgta 1920
gcatttgtat aattttatta cttaccttta ggcattacta tcacgttaac ttgggaagat 1980
agaaatgact agagtaaatt ttctcttcta taatatttgt ctagtgtgtt tttgtttttt 2040
ttttttttta tcctgtagtt taagaatgat gtttaaaaaa cacttttggt tctctcagcc 2100
tctggaattg ctttggcact cgtagatga atggctagtt ttaatagcca cagaattgat 2160
gaaaaacaaa agagactcaa cagagatcac ttctatttta ctgaaacaaa aaggccaaga 2220
tcaagatgct gcttccattc ctccatttga acctccagga cctgggagct atgaaaatct 2280
gtccactggc acaagggaat ctaaacgaga tgctcttgca gggagacagg aagccagttc 2340
agattgtcag gatgttattt ctatgacagc taaccggcta agtgctgtca ttcaagcttt 2400
ttacatgtgc tgttcttgtc agatgcctcc gggaaatgact tcacctcgtt tcattgaatt 2460
tgtctgcaaa catgatgaag ttttaaaatg ctttgttaat agaaatccca aaattatatt 2520
tgaccacttt cactttctcc ttgaatgtcc tgagttgatg tcaagattca tgcatatcat 2580
aaaagcacag ccttttaag atcgctgtga atggttctat gaacatttgc attcaggaca 2640
gccagattca gatatggtgc acaggccagt gaatgaaaat gatatcctgc tggttcacag 2700

```

```

agattctatt tttaggagta gctgtgaagt tgtgtcaaaa gcaaattgtg caaagctaaa 2760
gcaagggatt gctgtacggg tccatggaga agaaggcatg ggtcaagggt ttgtgcgtga 2820
gtgggtttgat attctgtcca atgagatagt caatcctgat tatgcattgt ttacccagtc 2880
agctgatgga acaacttttc agcctaatag caactcctat gtaaatcctg atcacttgaa 2940
ctattttcgg tttgctgggc agatcctggg attagcgttg aaccacaggc agctgggtcaa 3000
tatttacttc acacgatcct tctacaagca cattccttgg attcctgtaa attaccaaga 3060
tgtggcatcc attgatccag aatatgcgaa aaatttgcaa tggatttttag ataatagata 3120
aagtgatctg ggtctagaac taactttttc tgttgagact gatgtgtttg gagcaatgga 3180
agaggtgcct ttgaaacctg ggggtgggag tattccttgg acacaaaata ataaagcgga 3240
gtacgtccag cttgttactg aacttcgaat gacaagagcc attcagcctc agatcaatgc 3300
ttttttacag ggctttcata tgttcattcc accctccctc atacagcttt ttgatgaata 3360
tgaattggag ctactgcttt ctggcatgoc agaaattgat gtgagtgtt ggataaaaaa 3420
tacagaatac acaagtggct atgaaagaga agatccagtt attcagtggt tctgggaagt 3480
tgtagaagac attactcaag aggagagagt tcttctctta cagtttgtta cgggcagttc 3540
caggggtccc catgggtggg ttgctaatat catgggtgga agtggtattgc aaaactttac 3600
aatcgctgct gtgccatata ctccaaactc tttaccaact tcaagcacat gcatcaacat 3660
gctcaagtta cctgaatacc caagtaaaga aataactcaag gacagacttc ttgtggcact 3720
acattgtggc agctatgggt acacaatggc ataatagaat ctggaaaact cctctgacta 3780
ctgatgcaca attcagaatg gcagaagtaa tttgggaaaa tgtcaacaaa aaagcagcct 3840
aaatgcaacc cataggcagg gctgatgctt ccaatttata aaggatcatc aggttttctg 3900
ttctctctct ttccttttta tgttttctct gtttgtgata caattagaaa atataaaatc 3960
acagtagatt ttatttttta aaatgctaac tgaaagtaat agagactgtc ctttttcata 4020
attaatttta tccaagattg tattaaggca aaatctgatt ctacattcca cctctgctat 4080
gtaactgtct tgttaaaagg gtgttttctc ctaatttctg atataattata tgaggctatc 4140
cagctgggtg gttcttttgc atgtaaactg ccatttatat tttagaaaac tattgtatag 4200
aatggattta gattgtctat aaagccacaa atacgtattt tgccacagtg tattctatat 4260
tgcaatgatt ttttttagcat tttaatatatt taatatatat tgtaaaattt agactgatga 4320
tactaacagt tgatgaaatg acatataatt tatatatgaa agcttacgct atattgtatg 4380
aattatttgc atctttcagt ggccagtttt ccatatgtat atattatggt ctcaatgttt 4440
ttcttacgcc tcattttaat ttataatgaa ggtaaaatta aaatgtattt taccacgttt 4500
cttttcatta cttttatctg tgagctctga cacatctgaa aaagtaatct gatgtgcaaa 4560
ttataattta aatatgttaa ttttttggct tcttaaattt gcttttcac attaaaatgt 4620
caagttcaag tgatatgtgc ctaatatcac ttggatgttg gtgggttttt gaatttttgg 4680
gtgggttaatc agttttatatt tgaaaagacg tacttgaata gttacagcat atgtttgaac 4740
aggaagttag aacatgcata cacgaagaaa tgctaacgga aggatttgtt atgttttagga 4800
tcttccttg gaaactaaaa atagaatatt aatgacatta ctgttttag aatgacatat 4860
gcagattttc tcataagcag tcattgtgtt tgccagtaat gtttgagaga catgtaagt 4920
gaaagttttg ctaaattata aagctccttt aattcgttgg ttttgattct cttattctct 4980
tgtcttttct aaatgttaac aaaatatatc ttaacagatt acatgaaatt taggaattat 5040
ttaaaagtta ccattagctc taaaattaag attcggatgc tttatttata gtaactgaag 5100
ctaataatgt tttatgtttt gattttttga aatttaattg tagaagtcac tgcttcttga 5160
gttttcaaat agataaccac ctttaatat acactgctta taatacta atgtttacagat 5220
atgtttctgt ttataacat ataatacatt ggctttgtca tattagtttt ttttgcaagt 5280
agttatgtaa aagagataga taataaaata ttaataaact g 5321

```

<210> 15

<211> 5675

<212> ADN

<213> Homo sapiens

<400> 15

```

aacctccaga tacttacaag gagattacga tttcgggttg aacgtgcacc tgggtgaaact 60
gcattgattg acaggactgg cagaatgttg aagatggaac ctttggtctac agttgaatct 120
ctggaacagt accttttgaa aatggtagca aaacagtggt atgattttga ccgatcttca 180
tttgtttttg ttgaaaaatt aagagaagga caaaatttta tatttcggca ccagcatgat 240
tttgatgaaa atggaatcat ttactggatt ggaacaaatg caaaaactgc ttatgaatgg 300
gtaaatccag ctgcctatgg acttgtagta gtaacgtcat cagaagggaag aaatctacct 360
tatggccgct tagaagacat actaagtcgt gataattcag ctttaaatgt tcatagcaat 420
gatgataaga atgcctgggt tgccatagat ctgggtctct ggggtgatacc atcagcatat 480
acacttcgtc atgctcgtgg ttatggaagg tctgcactga gaaattgggt tttccaggta 540

```

tccaaagatg	gacagaactg	gacttctttg	tatacccatg	ttgatgactg	cagtctcaat	600
gaaccagggt	caactgcaac	ttggcctctt	gatccacca	aggatgagaa	acaagggtgg	660
agacatgtga	gaattaaaca	gatggggaaa	aatgccagt	gacaaacaca	ctacctctca	720
ttatctggat	tcgaacttta	tggcactgta	aatggagtat	gtgaagatca	gctagggaaa	780
gcagctaaag	aagcagaagc	taatcttaga	cggcagagac	gtctagtacg	ttcccagggt	840
ctgaaataca	tgggtccagg	agctcgtggt	atcagaggcc	tggattggaa	atggcgagat	900
caggatggca	gcccacagg	agaaggcact	gtcacaggag	aactacacaa	tggctggatt	960
gatgtcacct	gggatgctgg	tggctcaaac	tcttacgta	tgggcgcaga	aggaaaattt	1020
gacctcaagc	ttgcaccagg	gtacgacct	gatacagtgg	catcacccaa	acctgtttca	1080
tccactgttt	caggcacaac	gcaatcatgg	agcagcttgg	tgaaaaacaa	ctgtccagac	1140
aagacatctg	ctgctgcagg	ctcctcaagt	agaaaaggaa	gcagcagttc	tgtgtgtagc	1200
gtggccagta	gcagcgacat	cagcttgggt	tcgacaaaa	cggaacggag	atcagaaaat	1260
gtaattgga	acagtatagt	ttcaggagct	gatgtccatg	aaccaattgt	tgttctttca	1320
tctgctgaaa	acgtccctca	aacagaagta	gggtcatctt	ccagtgcagg	caccagcacc	1380
ttaacacggy	aaacgggaag	tgaaaatgct	gaaagggaag	tagggcctga	tagttctggt	1440
cgtactcctg	gggagtctag	tgcaatatcc	atgggaattg	tcagtgttag	ttctcctgat	1500
gttagttcag	tatctgaatt	aactaataaa	gaagcagctt	cacaacgacc	tcttagctct	1560
tcagcaagta	acagactgtc	agtgaagttc	ttgttggctg	ctggggcccc	tatgagctct	1620
agtgaagtg	tacctaacct	gtcctcaaga	gaaacatcta	gcttggagag	ttttgtaagg	1680
agagtggcaa	acatagcacg	gactaatgcc	acgaacaaca	tgaatctaag	ccgaagcagc	1740
agtgaataca	acactaatac	tttggggagg	aatgtgatga	gcacagcaac	ttctcctctt	1800
atgggtgctc	agagtttccc	taatttgacc	acacctggtg	ctacatcaac	agtgactatg	1860
tcaacatcca	gtgttactag	cagcagcaat	gtagctacag	caacaacagt	tttatcagtt	1920
gggtcaatctt	taagtaaac	tttaaccacc	agcctcacat	caacttccag	tgagagtgc	1980
acaggtcagg	aagcagaata	ttccttatat	gatttccctg	atagctgccg	tgccagtact	2040
ctattggctg	agctcgatga	tgatgaggac	ttacctgagc	cagatgaaga	agatgatgag	2100
aatgaagatg	acaatcagga	ggaccaagaa	tacgaggagg	ttatgattct	gagacgcccc	2160
tccctgcaac	gtcagagctg	ctcccgcctc	gatgtaacgc	atcatgctgt	tacctcgag	2220
ctaccacagg	tacctgctgg	agcaggaggc	cgacctattg	gggagcagga	agaagaagag	2280
tacgaaacta	aaggaggacg	ccggagaaca	tgggatgatg	attatgtgct	aaagagacag	2340
ttttctgcat	tggttcctgc	ttttgatcct	agacctggtc	gtactaatgt	ccagcagaca	2400
actgatctag	aaataccacc	cccaggggacc	cctcattcag	agctcttggg	agaagtcgaa	2460
tgtactccgt	cacctcgatt	agctctcact	ttgaaagtaa	caggctcttg	aacgactcgt	2520
gaagttgaat	taccactcac	caatttcaga	tcaaccatct	tttactatgt	acaaaaattg	2580
cttcaattgt	cctgtaattg	caatgtgaaa	tcagataaac	ttaggcgtat	ttgggagccc	2640
acatacacaa	tcatgtacag	agaaatgaag	gattctgata	aagaaaagga	aatggaaaa	2700
atgggttgct	ggtctataga	gcatgtggag	cagtaccttg	gcactgatga	attaccaaag	2760
aatgacttga	taacctacct	gcagaagaat	gcagacgctg	ctttcctgcg	ccactggaaa	2820
ttactggca	ctaataaaa	aacagaaatt	gttctcagct	accagggggc	catttctact	2880
tataaggatt	tttgtgagca	tggaacaaag	tctgggttaa	ctcaggccaa	agcaggcaat	2940
cttcaaagta	gtgatattct	taatttaaca	aaagaacaac	tcggtattct	gtatatagtt	3000
ggacagaact	cttgtggagt	agaagatgtc	cttcagcttc	tgcgatttct	tcagtttact	3060
gcaagtgacc	cttattcaag	aatatcccag	gaagatgggt	atgaacagcc	gcagattgag	3120
ttccaccag	atgaattcac	tagcaaaaa	attacaacaa	aaatattaca	gagatttgag	3180
gaaccattgg	cactggcaag	tggggctctg	ccagactggg	gtgaacaatt	aaccagcaaa	3240
tgtccttttc	taataccatt	tgaaactaga	cagctttatt	tcacatgtac	agcatttggc	3300
gcctcaagag	caatagtagc	gttacagaac	cgactggaag	ccactgtgga	gcgaacgaga	3360
accacaagca	gtgttagggc	agatgacct	ggagagtttc	gagttgggtc	tctcaagcat	3420
gaaagagtaa	aagtccacg	tggcgagtca	ctgatggaat	gggctgagaa	tgtcatgcaa	3480
atacatgcag	atcggaatc	agttcttgag	gttgaatttt	taggagaaga	aggaactggc	3540
ttgggaccca	cattagagtt	ttatgctctg	gtggcagcag	aattccagag	aactgacttg	3600
ggagcttggc	tttgtgatga	taattttcca	gatgatgaat	ctcgtcacgt	tgatcttggg	3660
ggtggattga	aacctcctgg	atattatgtg	cagaggctcat	gtggactggt	cacagcacca	3720
ttccacagg	atagtgatga	gcttgaaagg	atcacgaaac	tgtttcattt	ccttggaaat	3780
ttcttggcca	aatgcattca	agacaataga	cttgtggact	tacctatttc	taaacctttt	3840
tttaaactta	tgtgtatggg	tgacattaaa	agcaaatatga	gtaaactgat	ttatgagtca	3900
cgaggtgata	tggtattaca	ctgtactgaa	agtcagctctg	aagcttctac	agaagaaggt	3960
catgattcac	tctcggtagg	aagctttgaa	gaggattcaa	aatcagaatt	tattcttgat	4020
ccccctaaac	caaaaccccc	agcttgggtt	aatggaattt	tgacttggga	agactttgaa	4080
ttagtaaacc	cacacagagc	cagattttta	aaagaaatta	aagaccttgc	tatcaagagg	4140
cgccaaattt	taagcaacaa	aggtctttct	gaagatgaga	agaacacaaa	attacaggaa	4200

```

ctagtgtgta agaatccatc aggttctggg cctccactta gcatagagga tttagggtta 4260
aatttccagt tttgcccttc ctcaagaata tatggtttta cagctgtgga tctcaagcca 4320
agtgggtgaag atgagatgat aacaatggat aatgcagaag aatatgtgga tttgatgttt 4380
gactttttgta tgcatacggg tattcagaaa caaatggaag cctttagaga tgggtttaat 4440
aaagtttttc caatggagaa attaagttcc ttcagccatg aagaagtcca aatgattctt 4500
tgtggaaacc agtcaccatc ctgggcagca gaggatatta tcaattacac tgaacctaa 4560
ctgggttata cacgtgacag ccctgggttc ctgagggtttg tgagggtttt atgtggcatg 4620
tcttctgatg aaaggaaagc attcttgacg tttaccactg gttgttcaac tctaccccca 4680
ggtggactgg ctaacctgca tcccaggctc acggttgtac gcaagggtga tgctactgat 4740
gcaagctatc catcagtcac tacatgtgtg cattacctta agttgcctga atattcttcc 4800
gaggagatca tgagagagcg cctgctagct gctacaatgg agaaaggctt tcatctcaat 4860
tgagctttga agtgcaatgg gagacatcag agactttaaa aatactagtg aagcctcttg 4920
tgtttgtgtg cagagaagta tatgattcac catgctaatt acacttgcc ttttttccac 4980
cattaaggct ttaagaacat gtggaataag ttttttagct gctaatagaca aaacaaatcc 5040
tgtaactacc cagccagcaa gtatatagca cagaacactg tgttacttta caagggctta 5100
tgtgactgga ataagggtgg cccacttgac tgttccaaag agcagcttct cagatcttca 5160
gtgttccactg gtaaatctct aacagtgtat ttgtgtaaag tttgtcattt catactccat 5220
acactacagt tgctgtcact gatccctgtt ttgctggctt ttaagctact tgggtcaaaa 5280
tcctgcttcc ttaaaacata gagaattaat gagcatctca agctttttct tttccttttt 5340
aatgatgcct gcactatcaa gagtattcta gtgttctctc tttgtttggc atataatcat 5400
gcaccaaact ttttatttct ttaagggtgg agtatatttt tatttcttaa atgccatac 5460
atgaagatca aagtcttaag tgtgtttgca agtcaaaaat aaagatgtat taagggggga 5520
aaacctggtc taagtcaag gcacacttac agcaggtttt actttcggtt gtattttctt 5580
tgtatattat aaacatttat ttaacttggt gccgtttgaa gtaaaaaatt tccaaaatgt 5640
atgctcaaca ataattatta aatgtttgac agcgt 5675

```

<210> 16

<211> 5160

<212> ADN

<213> Homo sapiens

<400> 16

```

gttccagggtg caagcgccgg gtttgctgcc cgctggggcg cctgcagcg gcccgagctg 60
tgccggcggt ggatgagggg caggcgaggc agggcgcccc ctccagtatt gccgccccctc 120
ccgccccagg gcagggtgg gagggtacag cccggggggc ggctcggttc gcctccgggc 180
cgcccgcttc tcgctgcccc gggccggggc ggcggggcgc gagagcctcc cagcccgccc 240
cgtgccccgc ccgcccggct gcttcgcgg cgccagatgg ctaggtgccc 300
aggatgttca gcttcgaagg cgacttcaag acgcgcccca aggtgtccct tggcgggcgc 360
agcaggaagg aggaaaaggc ttctctttta catcgctactc aggaagaaag aagaaagaga 420
gaggaagaaa ggcgaagggt gaaaaatgca ataattatcc agtcatttat tcgaggctat 480
agagacagaa aacagcaata ttccatccaa agaagtgcac ttgatcgctg tgctaccttg 540
tcacagtcgg gggcgcttt tcccattgct aatggcccca accttacctt tttggttaagg 600
cagcttctgt ttttttaca acaaaatgaa gactcaaaac gtttgatatg gctgtatcag 660
aacttaatta aacacagctc tctgtttgtc aagcagttgg atggatctga gagacttaca 720
tgcttatttc agataaaaag attgatgagc ctctgttgca ggttgctgca aaactgtaat 780
gatgacagtt tgaatgttgc acttccaatg agaatgcttg aagtattttc gtctgagaat 840
acttacttgc ctgttttaca agatgctagc tatgtggtgt cagtgtattga acaaattttg 900
cactacatga ttcaaatgg gtattatagg tctctatatt tggtgattaa cagcaagctt 960
ccatcaagta ttgaatattc tgatttatct cgagttccta tagcaaaaat tttgctagag 1020
aatgttctaa aaccattgca ctttacttac aactcctgtc cggaagggtg gaggaacaa 1080
gtttttacag ccttcacaga ggagtttctg gcagcacctt ttacagatca gatttttcat 1140
ttcatcattc cggcgcttgc agatgcgcag accgttttcc cttaagagcc ctttctgaat 1200
gcactgttgt taatagagag tagatgttca agaaagagt gtggagcacc ctggcttttc 1260
tatttctgtt taactgttgg cgaaaattat ttggggggcc tctctgagga agggctgctg 1320
gtgtatttgc ggtgctgca gaccttctc tctcagttac cagtctctcc tgccagcgcg 1380
agctgtcacg actcagccag tgactctgag gaggagagt aagaagccga caagccctca 1440
agcccgagg atggcagact gtcagtatca tacataacag aggaatgcct gaagaagctg 1500
gacacaaagc agcagaccaa caccctgtc aacctggtgt ggagggactc tgcagcgag 1560
gaggctctca ccaccatggc ctccgtctgc cacacgctga tgggtgcagca ccgcatgatg 1620
gtacccaaag tcaggcttct ctacagttta gcctttaatg ccagggttct gagacatctt 1680

```

tggtttctaa	tatcttccat	gtcaacacgg	atgatcacag	ggtctatggt	accgttgctt	1740
caggtgatat	ccaggggttc	tcttatgtct	tttgaagatt	ctagtcgaat	catcccactc	1800
ttttaccttt	ttagctcctt	gtttagtcac	tcactaattt	ccatacatga	taacgaattc	1860
ttcgggtgatc	ccatagaagt	tgtagggtcaa	agacaatcat	caatgatgcc	ttttacttta	1920
gaagagctga	taatgttgct	tcgatgcctt	cgagatgcat	gcctggggat	catcaagtgt	1980
gcttatccag	aaaccaagcc	agaagttcga	gaagaatata	ttacagcatt	tcagagtatt	2040
ggagttacta	ctagctctga	aatgcaacaa	tgcatacaga	tggaaacagaa	aagatggatt	2100
cagttattta	aggttatcac	caatctagt	aaaatgttga	agtccagaga	cacgaggaga	2160
aatttttgct	ctccaaacca	ctggctgtca	gaacaagaag	atattaaagc	agataagggtc	2220
actcagctct	atgtgccagc	atccagacat	gtgtggagggt	tccggcggat	ggggaggata	2280
ggcccgcctgc	agtccaccct	ggacgtgggt	ttggagtccc	cgcgcctgtc	tgtgtctgag	2340
gaaagacagc	ttgctgtcct	gacagagttg	ccttttgtgg	ttccatttga	ggaacgagta	2400
aagatctttc	agaggttgat	ttatgcagat	aagcaagaag	ttcaaggaga	tgggtccattt	2460
ctggatggaa	ttaatgtcac	aataagaaga	aattacattt	atgaagatgc	ttatgacaaa	2520
ctttctccag	aaaatgagcc	tgatttgaaa	aagcggatcc	gtgtgcactt	gctcaatgcc	2580
catggcctgg	atgaagctgg	cattgatgggt	gggtggtattt	tcagagagtt	tttaaataaa	2640
ctactgaagt	caggattttaa	ccccaacccag	gggttcttta	agactactaa	tgaagggtct	2700
ctgtaccccca	accgcgctgc	tcagatgctt	gtgggagatt	cttttgccag	acattactac	2760
ttcctagcca	gaatgcttgg	aaaggctctc	tatgagaaca	tgctgggtgga	gctgcccttt	2820
gcaggcttct	ttctttccaa	gttgccttga	accagtgcgc	acgtggacat	tcaccacctc	2880
gcctccctag	accctgaggt	gtataagaat	ttgctctttc	tgaagagcta	cgaagacgat	2940
gtggaggagc	ttgggctgaa	cttcaactgtg	gtgaacaatg	acctgggaga	ggcgaggtga	3000
gttgaactaa	aattcgggtg	gaaagacatc	cctgtcacca	gcgccaaccg	gattgcgtac	3060
atccacttgg	tggcagacta	caggctgaac	aggcagatcc	gccagcactg	cctggctttc	3120
cgccaggggc	ttgccaatgt	cgtcagcctc	gagtggctcc	gaatgtttga	tcagcaagaa	3180
attcagggtat	taattttctgg	tgcacaagtt	cccataagcc	tagaggacct	aaaatccttt	3240
acaaactatt	caggaggcta	ttctgcagac	catcctgtta	ttaagggtctt	ctggagagtt	3300
gtggaagggt	tcaactgatga	agaaaagcgc	aaactgctga	agtttgtaac	aagctgctct	3360
cgaccccttc	tcttgggggtt	taaggagttg	tatcccgcat	tttgtattca	caacggaggc	3420
tcgagccttg	agcggctccc	cacagccagc	acctgctgaa	gctccccgag	gctccccgag	3480
ttctatgacg	agacactttt	gcgaagttaa	cttctctatg	cgattgaatg	tgcgctggc	3540
tttgagctga	gctgaagctg	atgctgggggt	cagaccctta	cagagaacca	gtgcttccct	3600
cgtcagcagc	gcctccccag	acccacgagg	atactcacac	tgcacgcctg	aggctctcct	3660
aagctccttc	tttcattctg	ccattcctcc	ctcccttcc	tttttaaatg	atttttatta	3720
cgggtgtggtc	acttatattag	atggacattg	cttttcaaat	aacttaaaat	aacacgttat	3780
gtgccatgtg	gctacttttag	taatatgtcc	aagaagagca	cagttttttac	actagtggca	3840
tctcagtga	attaacccaaa	gatgaagctt	tggctttgct	ggtgagatca	gagccctcct	3900
gagcaggcag	gcgccactcca	gggttcagac	aggcgtgcac	aggcggcaga	gatacagggg	3960
ctgagggctg	agacgccatg	gggcccgtgc	tgccttatgtg	gttggattgt	ttacaagcct	4020
cattattaaa	actgaaggca	tttttttttt	ctgctgcctt	tcccaaagtg	gttaggtttg	4080
gaaaagagat	gatgatggta	atattttatt	tgtgcttttt	aagccatttc	cccaaattgg	4140
actagcatgc	ttgtttttcag	tataccgtgg	cctgcctcat	gatggtttgg	agatactgtc	4200
tgtggatgtg	agggtggggac	ttcattoatt	gtcctatttc	tatctccact	ttgtgcctgg	4260
agagctttca	ggggagggtg	aggaggagg	tctgccaaag	tactgcaaca	tctgtcacc	4320
actataccca	gttacttggg	ggaggacaga	cactgtgggt	tcattaaagt	tgtttgaacc	4380
aaagtggcgg	ctgcactctt	gtcccgtatg	tagccgtgcc	ggtctcccat	catccgctcg	4440
ccctcccttc	ccctgggctg	cgccacttg	tcttccgtga	tatttggggg	tgactcgcca	4500
tgcttggcac	cctctgcttc	ctgggtgctg	tctgactoga	agacgggaca	gtccctgggt	4560
cacatccagg	gaagaggagt	gtcggtagtt	cttgacgtag	gcactttatc	aggacctgac	4620
ctgttgctgg	gtgatttttag	tctctacaaa	cagaaagcgt	ttcaaagcgt	cagctgtggg	4680
agcagagtga	ccctttgctg	atgctggggg	gaggggatct	aaatcctcat	ttatctcttc	4740
tatgtctagt	attttactgt	cactggaggc	tctgtgggct	gtcatagtta	attgaccata	4800
attagcaata	tactttttaa	gtgggaaagc	tgaatgacac	ttttaagaca	atgaacatta	4860
tcaaaacaaa	atgtataatt	tcttaatttg	aataataagc	gtttaaatgc	tattttagtg	4920
cttgatatac	agaaataaaa	taattagggg	tggctctttt	tatttttaggt	tttatgttga	4980
atgtctata	tcttattagt	taatttgat	tatttttag	tattttggaa	atagcatata	5040
tgagactgag	gagaaattga	caattcactt	atgtgtgggt	tttttctcag	ctattctgag	5100
cttattttatt	tatttgtatg	ttctaattgg	taaacattta	cattaaatat	tttttttccc	5160

<211> 5402
 <212> ADN
 <213> Homo sapiens

<400> 17

```

gccggttccc aggcgggggtg gagaccaact ctgggggtctc ttgggtgggag agtgggggagc 60
ccgtcttctg ctgagtgtctg ccgccccttc ccaggacagc ggaggtggaa cttcgtcgcg 120
ctgcaacccc cggctcggga tcctggggcg tcctttggca gttgattgca ccctgcacta 180
gagagaagtc caggaatgat atattcttgg gctgtatttc totaccctgg agttcatcca 240
gttcatccca gaaacactga cctgattact gtgacaaagt ctgatggata gaccctttct 300
gttgaccttg tggcgttttt ctttcatgtg gaagtggaa gacaagggtga aagggggccaa 360
aagttacctg ctttgggtgaa taggagtgtc ctacttgggt ctcagaacat gggccctcgt 420
ttaaagctgc agctgtgac ctcggctgtc tgttggcatt gacgggacct gatgttttac 480
gttattggtg gaatcacagt gtctgtgggt gcattcttct tcacaattaa gtccctcttt 540
gagcttgccg cagctgtagt cagcttcctc cagaatgagg accgcgagcg ccgaggggac 600
cggactatct atgactacgt gcggggaaat tacctggatc cccggctctg caaagtctcc 660
tggtattgga aggaccctta tgaggtgggc cacagcatgg ccttccgagt gcattttatt 720
tataagaacg ggcagccttt ccctgcacat cggcctgtgg gactaagagt tcacatctct 780
catgtcgagc tagcagtgga aattccagtg acccaggaag tccttcagga gcccaattcc 840
aacgtagtaa aagtggcctt cactgtgcgc aaggctgggc gttatgaaat cacagtgaag 900
cttgggtggat taaatgtggc atatagtcct tactacaaaa tttttcaacc tggaatgggtg 960
gttccttcta agaccaaaat tgtgtgccac ttttctactc ttgtattgac ctgtgggcag 1020
ccgcacaccc ttcaaatagt accccgagat gagtatgata atcccaccaa caattccatg 1080
tccttgagag atgagcacia ttacaccttg tccattcatg agctcggccc tcaagaagaa 1140
gagagtactg gtgtctcatt tgagaaatca gtaacatcca acaggcagac tttccagggtg 1200
ttcttgcgac tcaccctgca ttctcgaggg tgcttccatg cttgcatttc ataccaaaat 1260
cagccaatca ataatggtga atttgacatt attgtcctaa gtgaggatga gaagaatatc 1320
gtcgaacgca atgtgtccac ttcaggcgtg agcatttact ttgaggctta tctttataat 1380
gctaccaact gtatgcagc tccatggcac ctgccacca tgcacatgac ctcttcccag 1440
cgcgggccat ccactgctgt tgacgaggaa gatgaagact cgccctctga gtgccacacc 1500
cctgagaagg tgaagaaacc gaagaagggt tactgtctatg tgtcaccaaa gcaattctca 1560
gtgaaggagt totacctgaa gatcatcccc tggcgccctt acaccttcog agtgtgtcca 1620
ggaacaaaat tttcatacct tggtcctgac cctgtccata agctgtctac actggtgggtg 1680
gatgatggca ttcaacctcc tgtggagctc agctgtaagg agaggaacat tctagcagcc 1740
acttttatcc gctccctgca taagaacata ggaggctctg agacctttca ggacaagggtg 1800
aactttttcc agcgagagct tcggcaggta catatgaaaa gaccacattc caaagtcacc 1860
ctgaaggcca gcagacatgc cttgttgtaa tcgtctctga aagccactcg gaatttctcc 1920
atctcagatt ggagcaagaa ctttgagggt gttttccagg atgaagaagc tctggactgg 1980
ggagggcctc gccgggaatg gtttgagcta atctgcaaag cactatttga taccaccaat 2040
cagctcttca cccggttcag tgacaacaac caagcattag tgcacccaa ccctaatacg 2100
ccgctcatc tgccctgaa aatgtatgag tttgcgggac ggctcgtggg caagtgtctc 2160
tatgagtcct ctctaggagg agcctacaag cagttgggtc gagctcgctt caccgcctct 2220
ttcctggccc aaatcatagg actgcgtatg cattacaagt actttgaaac agatgacca 2280
gaattctaca aatctaaagt ttgttttatc ctcaacaatg acatgagtga gatggagctg 2340
gtctttgcag aagagaaata taataaatca ggtcaattgg ataaggttgt agaactcatg 2400
acagggtggag ctcaaactcc agtcaccaat gcgaataaaa tcttctattt aaatttgctg 2460
gccaatatc ggctggccag tcaagtgaag gaggaggtgg aacatttcct aaaaggcctg 2520
aatgaattgg tccctgagaa ccttttggct atttttgatg agaatgagct tgagctgctg 2580
atgtgtggga ctggagacat cagtgtgtct gacttcaaag cccatgcagt agttgttgg 2640
ggctcatggc atttcagaga aaaggctcat aggtgggttt ggactgtgg 2700
accaggagg agttggctcg gctacttcag ttcacaacag gctcctctca gctaccacct 2760
ggaggctttg ccgcccctcg tccctcattt cagattattg ccgctccgac ccatagcacg 2820
ctgcctactg cacacacatg ttttaaccag ctgtgcctcc ctacatatga ctccatgaa 2880
gaggtgcaca ggatgctgca gctggccatc agcgagggtt gcgagggtt tggcatgctc 2940
tgaccactct cctgtcatcc agttggctcc ctactctct ggagcttctg ggcgcaagtt 3000
acagacatca taacctga ttctaacaca cataaccatc agccagaaga tgccgcatgc 3060
tcccctgtgt ctggaggatt ttgtcaccta caagccttgt ctttacctca cctgctccct 3120
gcccatactc accacaggcc actttggcat ggtatgtaag ctgagctctt cattctgtca 3180
tgagaagagg accatgctgc tatcatttat ttggtccttt gaagatctca gtagctgagg 3240
gagatggcac accggggtca gccctgttgg gaaactggtg tggagaccct taaatccaca 3300
ctgtgtccca aaactccctc tgcgtgatact ccttgagagc accctctttg gccctcacta 3360

```

```

cttgaccaga ctggtacttg agtccttctc atgggtgggg tgattgcctc ttctcatcag 3420
gagccaggag agagggggac agataggagg tggcccatag gagcagtcct gctgcacaat 3480
ggtaggcata ggccatggca ctggactgoc tctaaggact gctaaaaaga atattttttt 3540
gtggtgtcag aactggaaaa agcactttcc cttcgggcat ttctggaaat gattattaat 3600
cacaaagaag aactctgtaa gctttttctt gaattgtagc cagtggagaa agcagataga 3660
ctgaagaata tgaaggatag ctgagctgta gcctccagag tggggcatgc ctaggcatat 3720
ggctggcttg gagactactg atgcttttcc ctgagtttgt tattggcact gaagtatggc 3780
cggcttgggc cactgacttc cccattatgt agtctgctaa aagctgggga tccttttagca 3840
ttctactgaa gaaaattttg tagcaaaaga ttagaacagt aagaatagta tgccagcaat 3900
ccctgattct ctctctgctt ggtgttctct ggcagcacta agacaaagga acagggacta 3960
ggagtttacg tgcttatcac aggtccttgg cgtcaggaca ctagggatga acatggaggt 4020
gatattgtac acagtaaaaca cctgccaacc cctgtactcc cctttgctcc atcagttatc 4080
aggaaggaga actaaggagg gaggaagctt attaggttca ctgttgaaat atgattgaag 4140
agtacctgct gctgtatctt ggaagatgac aacatccttt ttctattactg ttgattgaa 4200
aataatttca agattcaaaa tctcattgac ttcccaaatt tgtgttttta agaaggcttt 4260
gctgcaagct accattccca atggggcagt caactctgag aatattacaa gccatcttta 4320
ttgccaaaac cagcaagtac acagagtctc gaaacaaggc aggactttgg caatgccttg 4380
gcctcctgga agcatgtctg agccctcctg ttccgaggga tcatgaggaa caagccctcc 4440
tgagctgcta cagttcctgc ctgacctcag tctttcgagc tcatcatttg tcctttcttg 4500
gcacatgggt gctgacctca ctcatccttg gggtcgggga tccagcatca gggctctctat 4560
acccaggat cctccatgcc acatggtcac tgctctctc agggaccagg acaagggtgct 4620
gctgcttgcc caaatgtttt cccatgggat atgtaccgga aggtttatca ccaccctggg 4680
aaacataatg ctgccccttg ggcccagaaa ggggtttcca gctgggggca cgtggactgg 4740
ttctgctgtt tttggcagtc gcttttctga attctccct ccggcagcct tccagagact 4800
gatcctggag attgagttga ttgtcttgcc tagacctgtc attttaagct ctagtcatag 4860
cactttttca gagcatctta ggcaccattg caaccacaac agggaagctg catccctgtg 4920
gtggtcctta ggcaccagtc tttgttaaac aaaacccttt ggcactattg tggttttcta 4980
ttctctgtct gaactctatt caaaagtatc tttgctctct tgggcctttt cttttactgt 5040
tttgttttt ttttctaate ctgctttcat actagccagt gtggggaaaa ggtacaatat 5100
gtcaagaga tgagagatg ttatttcttg ggcaatttct tattagtgtt tcttattttg 5160
gccagttctt ttatttatgt ccttgtgacc caggtaactg gggggccagc taccctctctg 5220
gccttttagc gtctttgaag gagaccagac atgagtgaat acctaggaga gtgtcagcat 5280
gtttctggaa aattggcaga gaccaagccc tgctgcagat tcgtcaggcc aggtgaaagg 5340
gccaggcagt tgcagctgat gatgtaaata ttttgtacag tagataaata aatgtttaaa 5400
ag 5402

```

<210> 18

<211> 1764

<212> ADN

<213> Homo sapiens

<400> 18

```

agctttgcgc agtggcagta tcgtagccaa tgaggtttat ccgaggcgcg attattgcta 60
attgaaaact tttcccaata ccccgccatg acgacttgaa atatagtcgg cattggcaat 120
ttttgacagt ctctacggag actgcatgtg ctggtgaagc tcctggaagt gatggaagga 180
atggacaagg agacgttttg gttcaagttt gggaaggaa taacattcac cactgtactg 240
agtgaccaac aggtgggtgga gctgatccct ggggggtgcag gcatcgctgt gggatatggg 300
gaccgttctc gtttcatcca actggtccag aaggcacggc tagaggagag caaggagcag 360
gtggcagcta tgcaggcagg tctgctgaag gtggtaccac aggtctgtgt ggacttgctg 420
acctggcaag agttggagaa gaaagtgtgt ggggatccag aggtcactgt ggatgctctg 480
cgcaagctca cccggttttg ggaacttcgag ccatctgact cgcgggtgca gtatttcttg 540
gaggcactga acaacttcac caacgaggac cggagccgct tcctgcgctt tgtcacgggc 600
cgcagtcgcc tgcagcacg gatctacatc taccagaca agctgggcta cgagaccaca 660
gacgcgtcgc ccgagtcttc cacttgctcc agcacctctc tcctgccaca ctatgccagt 720
gccaaagtat gcgaggagaa gctccgctat cgggcctaca actgcgtggc catcgacact 780
gacatgagcc cttgggagga gtgaggcgtg ccgcccgtg tgggaccagc aagactgcac 840
gtgtccctct tggccttgcc cagggcgaag acaccttccc tgccttggtt tggctgacgt 900
gctcagcaaa accccatgtg cctgctcct gtgtgcagtt ggggtagggg cagctggcat 960
ggtcaggtaa cactagtggc ccagccccgc agaccacaa gccctaccg tgctggggct 1020
tgcttcccga ggtatttcac ctcttaagag ggaatcttcc acaagcccag cacaagctgc 1080

```

```

caggcctgag ctacttgaag gggggccatct aggtccccc aa cccatggact ttgectccat 1140
tttcagctcc gccttttttc tcctattttc tctctggctt tcttcagcca tgactcacia 1200
ctaaaaacat aaaacactgg aggttagtggt agggccctcc ccaagcaggg agcctgggat 1260
gggcaggagg tgatagccaa actccttggt cacctgctcc aagaaggag cagtagctga 1320
gcacctgccc tcacatactg ctcttttccc ctctccctcc acaccagaga tgtggtgagc 1380
tctgttcttc taccaaccca gtctcaacac acaaagtgc accaccttcc ctgactcaga 1440
acccacatcc actcaatgtg aactctacta ccacgacctc cccatattcc tcacttctcc 1500
atcacctcca gcctgactcc ctgtctgccc tttcaccccc aagattttgc acagggttaag 1560
gccagttatg gcctttttga aatctgtaat agctccctt tccccaactc taaagcctag 1620
accttaaacc tgttcctaga gctatgcaca ccctgcccc agtttaccgt tcctccctca 1680
gggcctccgt gacactccat gaaaagaagt tcttgcatat cggaaagtgt aataaatgga 1740
tgaattcaaa aaaaaaaaaa aaaa

```

```

<210> 19
<211> 2010
<212> ADN
<213> Homo sapiens

```

```

<400> 19
gtcgactacc agaaaatact ttcaacataa atgaactctc caacttatta aactttttata 60
tagatagagg aagacagctc tttcgggata accacctgat acctgcagaa acccccagtc 120
ctgttatttt cagtgtattt ccattttatct ttaattcgct atccaaaatt aaattattgc 180
aagctgattc acatataaag atgcagatgt cagaaaagaa agcatacatg cttatgcatg 240
aaacaattct gcaaaaaaag gatgaatttc ctccatcacc cagattttata cttagagtca 300
gacgaagtgc cctgggttaa gatgctctgc gtcaattaa tcaagctgaa gctactgact 360
tctgcaaagt attagtgggt gaattttatta atgaaatttg tcctgagtct ggagggggtta 420
gttcagagtt cttccactgt atgtttgaag agatgacca gccagaatat ggaatgttca 480
tgtatcctga aatgtgttcc tgcagtgtgt tctctgcca gcctaaacct gagaagaaaa 540
gatatttctt ctttggaatg ctgtgtggac tctccttatt caattttaa gttgctaacc 600
ttcctttccc actggctctg tataaaaaaac ttctggacca aaagccatca ttggaagatt 660
taaaagaact cagtcctcgg ttggggaaga gtttgcaaga agttctagat gatgctgctg 720
atgacattgg agatgcgctc tgcatacgt tttctataca ctgggaccaa aatgatgttg 780
acttaattcc aaatgggatc tccatacctg tggaccaaac caacaagaga gactatgttt 840
ctaagtatat tgattacatt ttcaacgtct ctgtaaaagc agtttatgag gaatttcaga 900
gaggatttta tagagtctgt gagaaggaga tacttagaca tttctaccct gaagaactaa 960
tgacagcaat cattggaaat actgattatg actggaaaca gtttgaacag aattcaaagt 1020
atgagcaagg atacaaaaaa tcacatccta ctatacagtt gttttggaag gctttccaca 1080
agctaaccct ggatgaaaaa aaaaaattcc tctttttcct tacaggacgt gataggctgc 1140
atgcaagagg catacagaaa atggaaatag tatttcgctg tcctgaaact ttcagtgaag 1200
gagatcaccc aacatcaata acttgtcata atattctctc cctccctaag tattctacaa 1260
tggaagaagt ggaggaagca ctccaagtag ccatcaacaa caacagagga tttgtctcac 1320
ccatgctcac acagtcataa tcacctctga gagactcagg gtgggctttc tcacacttgg 1380
atccttctgt tcttctttac acctaaataa tacaagagat taatgaatag tggttagaag 1440
tagttgaggg agagattggg ggaatgggga gatgatgatg atgggtcaa ggtgcaaaat 1500
ctcacacaag actgaggcag gagaataggg tacagagata gggatctaag gatgacttgg 1560
acacactccc tggcactgaa gactctgaac actggcctgt gattgggtcca ttccaggacc 1620
ttcatttgca taaggatat aaccacatca gcctctgatt ggccatgggc cagacctgca 1680
ctctggccaa tgattgggtt attccaggac attcatttgc ataaggagtc aaaccacacc 1740
agtcttggat tggctgtgag ccaattcacc tcagtctcta attggctgtg agtcagtctt 1800
tcattttacat aggtgtgaac catcaagaaa cctctacagg gtacttaagc cccagaagat 1860
tttgetacca gggctcttga gccacttgct ctagcccact cccaccctgt ggaatgtact 1920
ttcacttttg ctgcttcaat gccttgtgct ccaataaatc cactccttca ccacccaaaa 1980
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

```

```

<210> 20
<211> 4362
<212> ADN
<213> Homo sapiens

```

<400> 20

```

gtgggttagta gcgataaatg ctctgaagga aaaaacatac agtactgtga gaatgtagta 60
ttttgaaatg ttttcatttgc aaaattgtac ttgatactca gaatagtgtt tttttttcct 120
aaaatactta atatgctgaa caggtcaatg tatcatctgc ctagtatgca taaacaataa 180
cagtaaatgt ttagtaattt gagcacagt gggatgttc atgtattttt tattccttac 240
tctgttttga ttgccagtg ttgattttgc tattttcaaag aaatggcatt ttgttgagtt 300
tgaaaatttg aagggtgcatt ggatcctgct aatttgttgg attctgctaa attgtagcag 360
tacagctctc aggtttttgta gttaagagca taacttctact aaataaaaaat atttttaata 420
caataaatatg gaatcagagt aactctcctt tttcctgctc aaactttgat ttgctgaagc 480
tgcttcccga aaatcatgtg tctcgactct ctggtcagaa tttgctccgg actctcttat 540
ggcagaatca gaaatattaa aagtttgtca gatatccaga ttgtacaggt tgcttgtggg 600
tactatcatt cacttgcact ttctaaagca agtgaagtct tctgttgggg acagaataaa 660
tatggccaat tgggttttag tactgactgt aaaaagcaaa cttcaccgca gctgcttaag 720
tctttgcttg gaatcccttt catgcaagtt gcagcaggag gagcccatag ttttgtactc 780
accctttctg gagctatctt tggatgggga cgcaacaagt ttggtcagct aggtcttaat 840
gatgaaaatg ataggtatgt tcctaattta ctaaagtcac taagatctca gaaaatagtt 900
tatatttgtt gtggagaaga tcatactgct gctctaacca aggaaggtgg agtgtttact 960
tttgagagctg gaggggtatgg tcagtggggc cataattcta ccagtcatga aataaaccca 1020
aggaaagttt ttgaacttat ggggaagcatt gtcactgaga ttgcttgtgg acggcagcac 1080
acttctgctt ttgttccttc atcaggacga atttactctt ttgggcttgg tggtaatggg 1140
cagctgggaa ccggttcaac aagcaacagg aaaagtcctt ttactgtaaa aggaaattgg 1200
taccctata atgggcagtg tctaccagat attgattctg aagaatattt ctgtgtaaaa 1260
agaattttct cagggggaga tcaaagcttt tcacattact ctagtcccca gaactgtggg 1320
ccaccagatg acttcagatg tcccaatccg acaaagcaga tctggacagt gaatgaagct 1380
ctaattcaga aatggctgag ctatccttct ggaaggtttc ctgtggagat agccaatgag 1440
atagatggaa cgttttcttc ctctggttgc ctaaatggaa gtttttttagc tgtagcaat 1500
gatgatcact atagaacagg taccagattt tcaggggttg atatgaatgc tgctaggctt 1560
ttattccaca aacttatata acctgatcat ccgcagatat ctccagcagg ggcagctagt 1620
ttggaaaaga atcttatttc taaactgact agtcccttac ctgatgttga agcattgagg 1680
ttttatctta ctctaccaga atgtcccttg atgagtatt ccaacaattt cacaacaata 1740
gcaattccct tctgtgaac ctgaaaaagg caccactgaa agtacttgaa 1800
aactgggtgt cagtacttga acctccaacta ttctcaaga tagtagaact ttttaaggaa 1860
gttgtgttac atcttttgaa actctacaag atcgggtatt ccccttctga aagaagaatt 1920
ttcaacagtt ttcttcatac tgcattaaag gtttttagaaa tactacatag ggtaaatgag 1980
aaaatgggac agattatata gtatgataaa ttttatatac atgaagtaca agaattgata 2040
gacataagaa atgattatat caactgggtc caacagcagg cctatggaat ggatgtcaac 2100
catggattaa ctgagttggc agatatccct gttacaatct gtacatatcc atttgtattt 2160
gatgccaaag caaaaactac tctgttacag accgatgcag tcttacagat gcagatggct 2220
attgaatcag cccacaggca gaatgtctcc tctcttttcc tccagtgat tgaatctgtg 2280
aatccctgct taattctagt ggtgcgtaga gaaaatattg taggagatgc aatgggaagtc 2340
cttaggaaaa caaagaacat agattacaag aagccactca aggttatatt tgttgagaa 2400
gatgctgtgg atgcaggagg ggtgcgcaaa gaatttttct tgctcatcat gagggaatta 2460
ttggatccta aatacggcat gtttaggtat tatgaagatt ccaggctcat ttggttttct 2520
gataagacat ttgaagacag tgatttgttc catttgattg gtgttatctg tggcttagca 2580
atataaatt gtaccattgt ggacctccat tttccttttg ctttatataa gaaactactg 2640
aaaaagaagc catccttgga tgatttgaaa gaactaatgc ctgatgttgg gagaagcatg 2700
caacagttac tggattatcc agaagatgac atagaggaaa cattttgtct taattttactg 2760
atcacagttg aaaacttttg tgcaacagaa gtgaaaagagc tggttctaaa tgggtgcagac 2820
acagctgtta acaaacaaaa tcggcaagag tttgtcgatg cttatgtgga ttacatattc 2880
aataaatcag tggcttctct atttgatgct tttcatgcgg gctttcataa ggtctgtgga 2940
ggaaaagtcc ttctgctctt tcagcctaag gaactacaag caatggctcat tggaaataca 3000
aattatgatt ggaaggaact ggaaaagaat acagaataca aaggggaata ttgggcagaa 3060
catcctacga taaaaatttt ttgggaagta tttcacgaat taccattgga aaagaagaaa 3120
cagtttctgt tatttttgac aggtagtgtat cgcattccta ttcttgggtat gaagagtctg 3180
aaactagtca tccagtcac aggaggtggg gaggagtatc tccagtttc ccatacttgt 3240
tttaactctt tggatcttcc aaaatatata gaaaaagaaa ctctacgctc taaactgatc 3300
caagctattg atcacataga aggttccagt ttaataatac tttggagtta taactattca 3360
gtttagtga aaagcattaa actatttgtg ttttcttctg ggtgatgaat tcagcaaggt 3420
gacagaggta ctattataat tcttacttgc agaatgttca atctacgagt gttcatggaa 3480
gccaaaaaat attaaaggaa aatgaacaaa ctgttaatat tattgtacag aaccatggat 3540
tttttttgac catcttctaa taaacatagc aagtattatg aatacattaa agttttacta 3600

```

```

acatgaattt taagagtttg catatttcag aaatgatctg gtgtgagtgc atggaaatat 3660
tgcttaattt ttcttcaatc attgagttaa aaacctttta ctttggcctg caatagtcac 3720
ttgattattt ttctattttg taaataatgt taagttttgt aataaaatag ttatgttctg 3780
ataccagtac agtttctatg gttgtaattg aacttgaaca gactttttaa ggttaaaaat 3840
tatgatttaa atcttactct gagactctat aaaaagaaaa aaaaggtagc atgggtgaaa 3900
cactatcttt tcttttttgc tagaataagt gtctttgtgc aaacctaaat cacagatagg 3960
gaatatacta cataacctgc agttcttttc tttgtgaatc ctgacacact gatagatggg 4020
ggattgtcga tcagagaact tattaatatt tagtactgga agaactctgt ctccacagtt 4080
gccagttaata aaaagaaaca ttggctacta tgagcaccaa tcaactgggtt atagctttca 4140
aaattattga tgctgcagtg ctttagagct atttccttga atttaagaaa caaatcttaa 4200
cagttttatg gtgctgatgc ttagttgtct catgccatta aattgtaaaa gtgagttgat 4260
gcaatacatg tgactttctg ctataatgca aatattttatt ttttaaattt attttaaaat 4320
gccgtgaaaa ctgtttaata aagattttatt gttttaatat tt 4362

```

<210> 21

<211> 6926

<212> ADN

<213> Homo sapiens

<400> 21

```

atttgagcat cagttgaggg agaggacgct tcatttcctt tttggttggg aatgtgccgt 60
gaacatcctc tgctggagct tggggcctgg agcagactgg ggctgcagat ttagacgtag 120
acgtttccga ggacgcgaga ctgcagagct ttgtaggcag ccaatcccag agaagcagac 180
gggatggcta gttcagcccg ggagcacctg ctttttgtga ggctgcgaaa tccccagatg 240
cgttacacat tgagcccaga gaacctccag agccttgccg cccagagctc catgccagag 300
aacatgaccc tgacgcgggc caacagcgac accgacctgg tgacttctga gagccgctcc 360
agcttaactg ccagcatgta cgagtacacg ctggggcaag ccagaaacct cattatcttc 420
tgggacatta aagaggaggt ggaccccagt gattggattg gactttatca tatagatgag 480
aattctccag ccaacttctg ggattctaag aacaggggtg tgactggaac acaaaaaggg 540
caaattgtat ggagaattga gcctgggccc tatttcatgg aaccggagat aaaaatctgt 600
tttaaatatt accacggcat tagtggagcc ctgcgagcca cgacccctg catcacctg 660
aagaacccag ctgtgatgat gggggcagaa ggcatggagg gaggtgcttc aggaaacctg 720
cattctcgaa aacttgtag ctttacattg tcagatctta gggcagttgg gctaaagaaa 780
gggatgttct tcaatcctga cccttatctt aagatgtcaa ttcagccagg aaagaagagc 840
agtttcccca cctgtgcccc ccacgggcag gagagacggg ctactatcat cagtaacacc 900
accaatccaa tttggcaccg agagaaatat tccttttttg cacttcttac tgatgtctta 960
gaaattgaaa ttaaagacaa atttgccaa agcgtgcccc tcatcaagcg ttttctgggg 1020
aaactaacca ttccagtcca gaggtgctg gagcgacaag ccctcgggtg tcaaattgctc 1080
agctacaacc ttggcagaag gctcccagct gaccacgtga gtgggtacct ccagttttaa 1140
gtggaggtta cgtcttctgt tcatgaagat gcctctccag aagctgttgg cacaatactt 1200
ggagtcaatt ctgtgaatgg agacttaggt agccttccg atgacgagga catgccaggg 1260
agccatcacg acagccaggt gtgctctaatt gggccagttt ctgaggacag tgctgccgat 1320
ggaaccccca agcattcatt caggactagc tccactctgg aaatagacac agaggaatta 1380
acctctacgt cttcaaggac ctacacctcc agaggacgtc aggattcaat caatgattac 1440
ttagatgcta tcgaacacaa tggccactcc agggcaggga cagcgacctg ctctgagagg 1500
tccatgggtg cctctccgaa actgaggagt agttttccca ctgacacaag actcaatgcc 1560
atgcttcata ttgattcaga tgaagaagat cacgagttcc agcaagacct gggttacca 1620
tcttccttgg aggaggaggg aggcctgac atgttttagc gggcatccag agctgatgat 1680
ggaagcctga catctcagac aaagctggag gacaacctg ttgagaatga ggaagcctcc 1740
acacacgaag ctgcttctt tgaggataag ccagaaaatc ttccagagct tgcagagagc 1800
tccttacctg caggcccagc cccagaggaa ggtgaaggcg gccagagacc tcaacctcag 1860
gctgaccagg gcagtgtgta actgtgtggc tctcaagagg tagatcagcc cacaagtggc 1920
gcagacacag ggacttccga tgcctcagga ggaagccgaa gagccgtcag tgagaccgag 1980
tctctcgatc agggctctga gccttcccag gtgtcctctg aaacagaacc cagtgtacct 2040
gccaggacag agagtgtgag cgaagccag accaggcctg agggagagag tgacctgga 2100
tgctgtgaca gctctgcaa tgagagtgtg accacgcagc tgtcctctgt ggacacacgg 2160
tgctcctctc ttgagagcgc acggtttctc gaaaccccag ctttttcttc tcaggaggag 2220
gaagacggag cctgtgcagc agagcccacc agcagtggcc ctgccgaagg gtcgcaggaa 2280
tcctgtgtga ctgctggttc ttacctgtg gtacaggtgc ccagtgggga ggatgaaggg 2340
ccagggggcag aatcgccac tgtacctgac caggaggagc tgggggagggt ctggcagcgg 2400

```

```

agggggagcc tggagggagc tgcagctgct gccgagagcc caccgcaaga agaaggcagt 2460
gctggggagg cccaaggcac ctgtgaaggg gcaactgccc agggaggagg cgctactgga 2520
ggttctcaag ccaacggcca ccagccactg cgatcactac cttcagtgcg ccaggatgtt 2580
agccgggtacc agagggtgga cgaggtcttc ccaccaaact gggaggcacg cattgacagc 2640
cacggcagga tcttctacgt ggatcacgta aacagaacca cgacgtggca gcgaccgaca 2700
gtccccagc ccccgcaggt gctgcagaga tctaactcca tacagcaaat ggagcagctg 2760
aaccggcgat atcagagtat ccgcagaacc atgaccaatg agaggcctga ggaaaatacc 2820
aatgctattg atggggcagg ggaggaagct gactttcacc aagccagtgc agatttcoga 2880
cgggagaaca tcttgccctca ctctacctcc agatccagga tcacgttgct gttacagtct 2940
cccccgtag agttcctcat cagcccagag ttcttcaccg tgcctgcatc taaccctagt 3000
gcctaccgca tgtttacaaa caacacgtgt ttgaagcaca tgatcaccaa agtccggagg 3060
gacaccacc actttgaacg ctaccagcat aaccgcgacc ttgtgggatt cctcaacatg 3120
ttcggaaca aacagctaga gctgccgcgg ggatgggaaa tgaaacatga tcaccagggc 3180
aaggcattct ttgtggacca caactccgcg accaccactt tcattgatcc ccggctccca 3240
cttcagagca gtagaccac aagtgcgctg gttcatcggc aacacctgac aaggcaacgc 3300
agccacagtg cgggtgaggt aggagaagat tctcgacatg caggaccacc agttcttccc 3360
aggccatcca gtacattcaa tacagtcagt aggccacagt accaggacat ggttccagtg 3420
gcttacaatg acaagattgt tgcatttcta cgtcaaccca atatcttga aattctacag 3480
gagcgtcaac ctgatcttac caggaatcac tcaactcagg agaagatcca atttatccga 3540
actgaaggga ctccaggatt ggtgcgcctt tcaagcgatg cagacctgtg tatgttgctg 3600
agcttatttg aagaagagat aatgtcatat gtgcctccac atgccttact ccacccagc 3660
tactgtcagt cccacgtgg ctctcccggt tcatctcctc agaactcgcc aggtaccag 3720
cgtgccaatg ccgggcttcc agccccttac aagcgggatt tcgaagccaa actgaggaac 3780
ttttacagga agttagagac taaaggatat ggacaaggcc cagggaagtt aaagttaatt 3840
atccgaagag atcacttact agaagatgct tttaatcaga ttatgggcta ctccagaaaa 3900
gacctgcaga gaaataagct atatgtcacc ttcggtgggg aggaagggtt ggattacagt 3960
gggccttcta gagagttttt ctctctggta tccagagaac tctttaaccc atattatggc 4020
ttatttgaat attcagccaa tgacacatac acagtacaaa taagtcctat gtctgctttt 4080
gtagacaatc accatgaatg gttccgattc agtggttaga tcttgggtct tgcactaata 4140
caccagtatt tgttgatgc cttcttcaca cggccctttt ataaggctct tctcagaatt 4200
ctatgtgacc tgagtgcact agaatacctt gatgaagatg tccatcagag cctgcagtgg 4260
atgaagaca atgatatcca tgacatccta gacctcacgt tcaactgtga cgaagaagta 4320
tttgggcaga taactgaacg agaattaaag ccaggggggtg ccaatatccc agttacagag 4380
aagaacaaga aggagtacat cgagaggatg gtgaagtgga ggattgagag ggggtgtgta 4440
cagcaaacag agagcttagt gcgtggcttc tatgaggtgg tggatgccag gctggtatct 4500
gtttttgatg caagagaact ggaattggtc atcgaggcca cagctgaaat agacctaaat 4560
gattggagaa acaacacaga atatagagga ggataccatg acaatcatat tgtaattcgg 4620
tggttctggg ctgcagtgga aagattcaac aatgaacaac gactaagggt gttacagttt 4680
gttacaggca catccagcat tccctatgaa ggatttgctt cactccgagg gagtaacggc 4740
ccaagaagat tctgtgtgga gaaatgggg gcctccctac ccctctccag agcgcataca 4800
tgttttaacc gtctggatct gcctccctac ccctctcttt ccatgcttta tgaaaaactg 4860
ttgacagcag ttgaagaaac cagtactttt ggacttgagt gacctggaag ctgaatgccc 4920
atctctgtgg acaggcagtt tcagaagctg cttcttagaa gaatgattga acattggaag 4980
tttcaagagg atgcttctct taggataaag ctacgtgctg ttgttttcca ggaacaagtg 5040
ctctgtcaca tttggggact ggagatgagt cctcttgga ggatttgggt gagcttgatg 5100
cccagggaac aacccaaccg tctttcaatc aacagttctt gactgccaaa ctttttccat 5160
ttgttatgtt ccaagacaaa gatgaacca tacatgatca gctccacggg aatttttagg 5220
gactcaggag aatcttgaaa cttacccttg aacgtgggtc aagccaaaact ggcagcattt 5280
ggcccaatct ccaaattaga gcaagttaaa taatataata aaagtaataa tatttcctga 5340
aagtacattc atttaagccc taagttataa cagaatatcc atttcttgct tatgagtgcc 5400
tgcatgggtg gcaccatagg tttccgcttt catgggacat gagtgaataa gaaaccaagt 5460
caatatgagg tacctttaca gatttgcaat aagatgggtc gtgacaatgt atatgcaagt 5520
ggtagtgtgt taattatggc taaagacaaa ccattattca gtgaattact aatgacagat 5580
tttatgcttt ataattgcatg aaaacaattt taaaataact agcaattaat cacagcatat 5640
caggaaaaag tacacagtga gttctgttta tttttttag gctcattatg tttatgttct 5700
ttaagatgta tataagaacc tacctatcat gctgtatgta tcaactatc cattttcatg 5760
ttccatgcat actcgggcat catgctaata tgtatccttt taagcactct caaggaaaca 5820
aaagggcctt ttatttttat aaaggtaaaa aaaattcccc aaatattttg cactgaatgt 5880
accaaaagggt aaggacattt acaatatgac taacagcaac tccatcactt gagaagtata 5940
atagaaaata gcttctaaat caaacttctt tcacagtgcc gtgtctacca ctacaaggac 6000
tgtgcactca agtaataatt ttttaagatt cactatatgt gatagtatga tatgcattta 6060

```

```

tttaaaatgc attagactct cttccatcca tcaaatactt tacaggatgg catttaatac 6120
agatatttcg tatttccccc actgcttttt atttgtacag catcattaaa cactaagctc 6180
agttaaggag ccatcagcaa cactgaagag atcagtagta agaattccat tttccctcat 6240
cagtgaagac accacaaatt gaaactcaga actatatttc taagcctgca ttttactga 6300
tgcataaattt tcttattaat attaagagac agtttttcta tggcatctcc aaaactgcat 6360
gacatcacta gtcttacttc tgcttaattt tatgagaagg tattcttcat ttttaattgct 6420
tttgggatta ctccacatct ttgtttattt cttgactaat cagattttca atagagtga 6480
gttaaatttg gggtcataaa agcattggat tgacatatgg tttgccagcc tatgggttta 6540
caggcattgc ccaaacattt ctttgagatc tatatttata agcagccatg gaattcctat 6600
tatgggatgt tggcaatctt acattttata gaggtcatat gcatagtttt catagggtgt 6660
ttgtaagaac tgattgctct cctgtgagtt aagctatgtt tactactggg accctcaaga 6720
ggaataccac ttatgttaca ctctgcact aaaggcatgt actgcagtgt gaagaaatgt 6780
tctgaaaaag ggttatagaa atctggaat aagaaaggaa gagctctctg tattctataa 6840
ttggaagaga aaaaaagaaa aacttttaac tggaaatgtt agtttgact tattgatcat 6900
gaatacaagt atatatttaa ttttgc 6926

```

<210> 22

<211> 2601

<212> ADN

<213> Homo sapiens

<400> 22

```

gaatttgaat cctgtgggccc gttgaatgtg gctgctcgcg gtcggcggtgc cccgacgtac 60
agcggggccg gaaaagtggc actgaggctc tggaaacttct gccagctct ccttgtaaaa 120
tgaatgaaag taaacctggt gactcacaga accttgcttg tgttttctgt cgaaaacatg 180
atgactgtcc taataatac ggagaaaaga aaactaagga gaaatggaat ctactgtac 240
attactactg tttgttgatg tcaagtggaa tttggcagag aggcaaagaa gaagaaggag 300
tttatggttt tctaatagaa gatatcagga aggaagtga tagagcttct aaactgaaat 360
gctgtgtttt caagaaaaat ggtgcttcaa ttggatgtgt tgcaccccg tgtaaacgaa 420
gttatcattt cccatgtgga cttcagagag aatgtatttt ccagtttact ggcaattttg 480
cgctcatttt ttgggacct cgacctgttc aaataattac atctaataat tatagagagt 540
ccttaccatg caccatttgc ttggaattta ttgagcctat tccaagttat aacatattac 600
gaagtccctg ttgtaagaac gcttggtttc atagagactg tttacagggt caagcaataa 660
atgcgggagc gtttttcttt aggtgtacaa tatgcaataa tagtgacatc tttcagaaag 720
agatgttgag aatgggaatt catattcctg aaaaagatgc ttctgggaa ttagagggaa 780
acgcttatca agagcttctg cagcactatg agcgttgtga tgttcgaaga tgcgttgca 840
aagaagggcg agactataat gcacctgata gcaaatggga aataaagcgc tgcagtgtt 900
tggtgtccag tggcacacat ttagcctgct cctcattacg gtcacgggag caaatggg 960
agtgttttga atgtagggtt attatctaca attcaggaga gttccaaaaa gccaaaaaac 1020
atgtattacc caattctaat aatgtgggga ttacagattg tttgttgga gagtcatcac 1080
ctaaattacc cagacagtca cctggatccc agagtaaaga tctactgagg caaggcagca 1140
aatttagaag aaatgtatca acactattaa tagagttagg attccaaatt aaaaaaaaaa 1200
ctaaaagatt gtatatcaac aaagccaata tctggaatag tgcccttagat gcattcagaa 1260
atcgaaactt taatccttca tatgcaattg aagtagcata tgttattgaa aatgataatt 1320
ttggaagtga gcatcctgga tcaaagcaag aatttctgag tctcttaatg caacatcttg 1380
agaactcatc attgtttgaa gggtccttgt caaagaactt gtctctaaat tctcaagctc 1440
tgaaagagaa tctttactat gaagctggca aaatgcttgc catttcttta gttcacgggtg 1500
gtccttcacc tggtttcttt tctaaaacct tgtttaactg ccttgtttat ggaccagaaa 1560
ataccagacc aatttttagat gatgtttcag actttgatgt ggcacagatt ataactcagga 1620
taaatactgc aacaactgta gctgacttaa agtcaataat aaatgaatgc tataactacc 1680
ttgagttaat tggatgtctc agacttataa cgacattaag tgataaatat atgttagtaa 1740
aagacatact tggctaccat gtaattcaga gagtcacac accctttgaa agttttaagc 1800
agggctctgaa aacccttggt gttttggaga aaattcaggc ttatccagaa gcattttgta 1860
gcatcctgtg tcataaacct gagagcttt ctgcaaaaat ccttagtgag ctttttacag 1920
tacacacatt acctgatgtg aaagctttgg ggttttgaa cagttactta caggctgttg 1980
aagatggtaa atctacaaca acaatggaa acattcttat ttttgcaact ggttgcaatt 2040
ccattcctcc agctggattt aaaccactc cttcaattga gtgtctgcat gtggattttc 2100
ctgttggaac cgagtgtaat aactgtttag caattcccat caccaataca tataaagagt 2160
ttcaagaaaa tatggacttc accataagaa acactctaag actagaaaag gaagaaagt 2220
ctcattacat tggacattaa aatgtttcct tgaacaaaga gaagcttctt taaaagctgc 2280

```

```
tattgataac tctcttttat ttcactaacc ttttcatcat cactttgaac aaactagtta 2340
gcttcttgac ctaataaaaat ttatgatatg aatataaagg aatattatag ccacttaggc 2400
taaaaaaagg ttttttttgt taaagcatac tagtcaaaat tggtgataat ttctcatttt 2460
cttgatatag atacttcata aacctcaata taaatactct caaaatgggt gtatggattt 2520
tatttaggta tggtagcatg gttaataaaa gtaaaaactg ttttactcaa aaaaaaaaaa 2580
aaaaaaaaa aaaaaaaaaa a                                     2601
```