

(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  
12 octobre 2006 (12.10.2006)

PCT

(10) Numéro de publication internationale  
**WO 2006/106193 A1**

(51) Classification internationale des brevets :  
*C12N 15/62* (2006.01)    *C12N 15/10* (2006.01)  
*C07K 14/00* (2006.01)    *G01N 33/50* (2006.01)

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(21) Numéro de la demande internationale :  
PCT/FR2006/000257

(22) Date de dépôt international : 3 février 2006 (03.02.2006)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :  
0503366                      5 avril 2005 (05.04.2005)    FR

(84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Déposant (pour tous les États désignés sauf US) :  
**BIOMETHODES** [FR/FR]; Genavenir 8, 5 Rue Henri Desbruères, F-91030 Evry (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : **DEL-COURT, Marc** [FR/FR]; 43 Rue De Tolbiac, F-75013 Paris (FR). **DUCHER, Laurent** [FR/FR]; 6 Rue Grognard, F-94120 Fontenay Sous Bois (FR). **BLESA, Stéphane** [FR/FR]; 6 Rue Du Général Leclerc, F-77166 Grisy-sus-isnes (FR).

**Déclaration en vertu de la règle 4.17 :**

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

**Publiée :**

— avec rapport de recherche internationale

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

(74) Mandataires : **GALLOIS, Valérie** etc.; Cabinet BECKER & ASSOCIES, 25 Rue Louis Le Grand, F-75002 Paris (FR).

(54) Title: METHOD FOR OBTAINING SOLUBLE OR BETTER EXPRESSED VARIANTS OF A PROTEIN OF INTEREST

(54) Titre : METHODE D'OBTENTION DE VARIANTS SOLUBLES OU MIEUX EXPRIMES D'UNE PROTEINE D'INTERET

(57) Abstract: The invention concerns a method for selecting soluble or better expressed variants of a protein of interest of a peptide of interest, from a bank of mutants of the protein or of the peptide of interest fused to the gene of resistance to kanamycin.

(57) Abrégé : La présente invention concerne une méthode permettant la sélection de variants solubles ou mieux exprimés d'une protéine ou d'un peptide d'intérêt, à partir d'une banque de mutants de la protéine ou du peptide d'intérêt fusionnés au gène de la résistance à la kanamycine.

WO 2006/106193 A1

## METHODE D'OBTENTION DE VARIANTS SOLUBLES OU MIEUX EXPRIMES D'UNE PROTEINE D'INTERET

### Introduction

5 La présente invention concerne la biologie moléculaire et plus particulièrement l'évolution des protéines *in vitro*. La technologie a pour objet une méthode permettant la sélection, à partir d'une banque de mutants d'une protéine générée dans une première étape, des variants présentant une solubilité améliorée ou un niveau d'expression supérieur.

10

L'insolubilité des protéines recombinantes constitue un problème majeur dans le domaine des biosciences : lorsque exprimées en bactéries, une grande parties des protéines ne peuvent être exprimées à un niveau important sans s'agréger sous forme de corps d'inclusion. Elles ne pourront alors pas être purifiées, et seront dans la plupart des cas inactives.

15

Lorsque l'on exprime une protéine eucaryote (notamment une protéine humaine) dans un hôte procaryote, les différentes étapes de maturation (glycosylation, clivage de peptides, etc...) n'ont pas lieu. La protéine a donc une structure générale assez distante de sa structure naturelle, et présente le plus souvent une solubilité réduite.

20

Dans le cas de protéines recombinantes bactériennes produites en bactéries, ce phénomène d'insolubilité n'est observé que dans une moindre mesure. Cependant, si l'on souhaite exprimer une grande quantité de la protéine, une fraction des protéines s'agrège : il existe une concentration limite, fonction de la solubilité de la protéine, que l'on ne peut dépasser sans former de tels corps d'inclusion.

25

La nature des protéines agrégées est parfois réversible : si l'on traite les corps d'inclusion de façon à les dissoudre, soit en utilisant des dénaturants, soit en modifiant le pH ou la température, puis qu'on replace la solution dans des conditions plus physiologiques, on peut retrouver des protéines dans leur forme native et soluble, c'est-à-dire ayant récupéré leurs fonctionnalités. Néanmoins, le plus souvent, une fraction très

30 minoritaire seulement des protéines peut ainsi être récupérée.

Cette limite à l'expression des protéines représente un obstacle lorsque l'on souhaite exprimer une protéine en grande quantité : c'est le cas notamment des enzymes

industrielles, - utilisées dans les industries textiles, agroalimentaires et chimiques notamment -, ou celui des fragments d'anticorps recombinants à une seule chaîne, les ScFv. Cette dernière classe de molécules, qui s'expriment très mal en bactéries pour ces raisons de limite de solubilité, sont utilisables en thérapeutique humaine et doivent alors  
5 être produits dans des quantités non négligeables.

Cette limite à l'expression que représente l'insolubilité des protéines recombinantes est également un frein à l'expression des protéines utilisées en petite quantité à des fins de recherche. Par exemples, les cibles thérapeutiques doivent, pour être caractérisées, être produites en quantité suffisante pour permettre la réalisation  
10 d'études structurales. Ces études ne peuvent être réalisées sur des protéines insolubles car la formation des cristaux destinés à ces analyses se fait en milieu liquide. L'insolubilité des protéines est donc un phénomène qui limite les progrès dans la caractérisation de nombreuses protéines.

15 Des efforts importants pour améliorer la production des protéines recombinantes ont été menées ces dernières années.

L'approche générale consiste à rechercher des variants de la protéine ayant acquis une meilleure solubilité, tout en ayant conservé ses autres fonctionnalités (son activité pour une enzyme, sa capacité à lier un antigène pour un anticorps recombinant).  
20 Dans une première étape, des mutants sont générés à partir du gène codant pour la protéine de départ. Dans un deuxième temps, ces mutants sont criblés pour identifier ceux qui présentent le meilleur niveau de solubilité ou d'expression. Dans un troisième temps, les mutants présentant un niveau de solubilité ou d'expression amélioré sont testés pour leurs fonctionnalités, afin de vérifier que la ou les mutations introduites ne  
25 les ont pas affectées.

Différentes techniques de mutagenèse existent ; elles permettent de modifier artificiellement la séquence nucléotidique d'un fragment d'ADN. Elle peuvent être séparées en quatre grands groupes : la mutagenèse aléatoire, la recombinaison de gènes,  
30 la mutagenèse dirigée et la mutagenèse massive<sup>TM</sup> (Massive Mutagenesis®).

Dans la technique de mutagenèse aléatoire, l'ADN ciblé (contenant le gène à muter) est amplifié dans des conditions particulières altérant les capacités de la polymérase à répliquer fidèlement l'ADN. Celle-ci introduit des mutations au fil des cycles, c'est-à-dire des différences par rapport à la séquence initiale. A la fin de la

réaction, un grand nombre de copies de la molécule initiale sont obtenues, chacune de ces molécules comportant des mutations différentes. Ces molécules sont présentes sous forme de banque, c'est-à-dire d'un mélange de molécules de nature différente (différant par la nature et la position de leurs mutations). L'ADN est ensuite inséré dans un vecteur d'expression.

La mutagenèse par recombinaison d'ADN s'inspire du processus de recombinaison à l'œuvre dans l'évolution Darwinienne. La mutagenèse par mélange d'ADN consiste à recombiner des séquences partiellement homologues, isolées à partir de différents organismes. Par exemple, si l'on travaille sur une enzyme, la première étape d'une approche par mélange d'ADN reviendra à isoler un grand nombre de gènes homologues à cette enzyme, soit à partir de collections de souches, soit à partir de gènes directement isolés à partir d'échantillons naturels (par une approche maintenant décrite sous le terme de « metagénome »). Différentes approches existent alors pour mélanger les domaines de ces gènes homologues et générer une banque de molécules d'ADN « chimériques », c'est-à-dire constituées de plusieurs domaines ayant des provenances différentes (Brevet US 6 132 970 ; Stemmer WP. *et al*, Nature, vol.370, p.389-391, 1994 ; Aguinaldo AM. *et al*, Methods Mol Biol., vol.192, p.235-239, 2002 ; Zhao H. *et al*, Nat Biotechnol., vol.16, p.258-261, 1998 ; Shao Z. *et al*, Nucleic Acids Res., vol.26, p.681-683, 1998 ; Kawarasaki Y. *et al*, Nucleic Acids Res., vol.31, 2003 ; Brevet US 5 965 408 ; Brevet WO 00 09 679 ; Brevet WO 02 48 351). Il est attendu que ces molécules contiennent ainsi des caractéristiques nouvelles, et notamment des propriétés cumulées de deux ou plusieurs gènes parentaux. Cette approche par recombinaison d'ADN est basée sur l'idée générale que la combinaison nouvelle de mutations naturelles, ayant donc été pré-criblées par la nature pour le maintien de l'activité de l'enzyme -, a plus de chance d'être porteuse d'amélioration que l'introduction de mutations générées au hasard. En même temps que l'on restreint la génération de diversité à un espace de séquence 'raisonnable' car 'présélectionné', on est cependant limité par les séquences d'origine, qu'il faut posséder physiquement, par la nécessité d'utiliser des gènes ayant un niveau d'homologie suffisant, et par l'impossibilité de générer des séquences autres que des combinaisons des séquences initiales. Ces approches de mélange d'ADN se sont révélées être particulièrement efficaces dans le domaine de l'amélioration de l'activité ou de la stabilité des enzymes.

La mutagenèse dirigée vise à introduire une ou quelques mutations (substitutions, mais aussi délétions ou insertions) de nature et de position connue dans

un fragment d'ADN. Un oligonucléotide est utilisé pour introduire cette mutation. Cet oligonucléotide est classiquement constitué d'une trentaine de bases, et est homologue en tout point à la séquence ciblée sur le fragment d'ADN, à l'exception d'une ou quelques positions localisées dans sa partie médiane. Cet oligonucléotide est ensuite  
5 utilisé pour amorcer une réaction de réplication en utilisant le fragment d'ADN comme matrice. Dans de nombreux protocoles, un second oligonucléotide, de sens opposé, est utilisé de façon à améliorer le rendement de la réplication. L'ADN synthétisé lors de cette réaction est ensuite sélectionné, notamment selon le critère de son contenu en bases méthylées, de façon à éliminer l'ADN parental non mutant. Différentes  
10 techniques de mutagenèse dirigée permettent d'obtenir assez rapidement des mutants, à un rythme qui reste cependant limité : seuls quelques mutants peuvent être générés par jour et par personne. Ce faible débit n'est que très partiellement adapté aux problématiques d'évolution moléculaire : il est en effet rare de pouvoir prédire précisément l'effet des mutations introduites et, lorsqu'on recherche un mutant amélioré  
15 sur tel ou tel critère, il est généralement préférable d'en générer un grand nombre dont la plupart n'auront pas l'effet escompté.

La Massive Mutagenesis® permet de réaliser en un temps très court un grand nombre de mutations dirigées sur un même fragment d'ADN cible. La matrice contenant l'ADN cible est préparée, et mélangée avec les oligonucléotides mutants  
20 initialement fabriqués. La matrice est dénaturée par la chaleur de façon à disposer temporairement d'ADN simple brin. Lors du retour à une température plus basse, certains ou tous les oligonucléotides présents dans le mélange se fixent sur la matrice au niveau de leur site d'homologie. Tous les éléments nécessaires à réaliser une réplication de la matrice à partir des oligonucléotides mutants sont ajoutés, et notamment une  
25 polymérase, le tampon permettant son activité, des nucléotides triphosphates en quantité suffisante, les éventuels cofacteurs nécessaires. La réaction de réplication a ensuite lieu dans les conditions de température correspondant à l'activité maximale de la polymérase. L'objectif est de disposer de mutants contenant chacun une mutation ou une combinaison de plusieurs mutations différentes. Le nombre moyen de mutations par  
30 molécule est modulable à loisir. La diversité possible s'accroît très rapidement avec le nombre de mutations moyen par molécules : l'utilisation de Massive Mutagenesis® permet d'obtenir des diversités très larges.

Ces diverses technologies permettent de générer de la diversité. Selon le contexte précis, l'une ou l'autre peut être préférée.

Si l'on recherche des mutants présentant par rapport à la molécule d'origine une amélioration de solubilité ou d'expression, il est nécessaire de disposer d'une technologie permettant de mesurer rapidement cette caractéristique chez chacun des mutants générés. Bien entendu, on pourrait effectuer les mesures une par une en observant chez chacun des mutants le taux de protéines solubles. Cette approche serait cependant très laborieuse, et c'est pourquoi des techniques ont été développées pour permettre de tester ce paramètre de façon parallèle, directement à partir des banques de mutants.

Ces techniques sont pour la plupart basées sur la fusion, en phase, d'une banque de mutants de la protéine d'intérêt d'une part, et d'autre part d'un gène rapporteur. Ces protéines de fusion sont ensuite criblées sur le critère de l'activité du gène rapporteur.

Les protéines de fusion sont utilisées depuis de nombreuses années dans des contextes autres que celui de la recherche de variants solubles : principalement, ces protéines de fusion sont utilisées pour localiser les protéines au niveau cellulaire ou intracellulaire.

Une des premières protéines de fusion à avoir été utilisées est la  $\beta$ -galactosidase, dont l'activité convertit le X-gal, un sucre non naturel, en un composé bleu. Cette activité peut donc facilement être tracée. (Casadaban MJ. *et al*, *Methods Enzymol.*, Vol.100, p.293-308, 1983)

La CAT (chloramphénicol-acétyl-transférase) a également été utilisée dans ce contexte des protéines de fusion depuis de longues années. Le chloramphénicol est un antibiotique qui se lie aux ribosomes de la bactérie, et bloque ainsi la biosynthèse des protéines, ce qui entraîne rapidement la mort de la cellule. La CAT est une enzyme qui transfère un groupement acétyle d'un co-enzyme A sur le chloramphénicol (qui est alors inactivé). A partir de cette différence d'un groupement acétyle, des tests colorimétriques ont été mis au point : l'activité CAT est alors tracée par un signal jaune (Robben J. *et al*, *Gene*, vol.126, p.109-113, 1993).

Ces approches basées sur l'utilisation de protéines rapporteur enzymatiques présentent un obstacle non négligeable : le substrat doit pouvoir accéder au compartiment dans lequel se situe l'enzyme rapporteur. Dans de nombreux cas, le

marquage en utilisant ces enzymes n'est donc pas possible. De plus, le produit coloré marqué par l'enzyme est une petite molécule chimique, qui a une tendance naturelle à la diffusion : la localisation intracellulaire des protéines de fusion est donc imprécise.

C'est pour ces raisons qu'ont été développées des protéines de fusion naturellement fluorescentes. Le rapporteur le plus couramment utilisé est aujourd'hui la Green Fluorescent Protein (GFP) (Garamszegi N. *et al*, Biotechniques, vol.23, 1997). La GFP, isolée en 1962 et clonée en 1992, se compose de 238 acides aminés et est issue de la méduse *Aequorea victoria*. Elle a pour caractéristique de posséder un chromophore lui permettant d'absorber les rayonnements bleus (400nm) et de ré-  
10 émettre dans le vert (509nm). Il existe maintenant de nombreux variants de la GFP, qui fluorescent à d'autres longueurs d'ondes et émettent ainsi différentes couleurs (jaune, bleu, violet). Plusieurs protéines de fusion peuvent donc être observées simultanément. Une seconde protéine naturellement fluorescente a été isolée et clonée à partir de l'anémone de mer : le DsRed. Cette protéine émet naturellement un signal rouge.  
15 L'utilisation combinée de cette protéine et de la GFP permet là encore de tracer plusieurs protéines en même temps.

Des GFP modifiées ont été réalisées de façon à permettre une expression optimale chez de nombreux organismes tels que les bactéries, les levures, ou les cellules mammifères, afin de permettre les études dans chacun des système couramment  
20 employés en biologie.

Lorsqu'on réalise de telles protéines de fusion, la liaison entre le gène d'intérêt et celui codant pour la protéine rapporteur doit se faire sans interrompre la phase de lecture. Il est également nécessaire de prendre soin aux contraintes stériques de l'une et  
25 l'autre protéines. En effet, il est possible que la fusion de la protéine d'intérêt et de la protéine rapporteur ne permette pas à l'une et à l'autre d'adopter leur conformation naturelle. Leur structure en serait modifiée, et leur activité annulée. Afin d'éviter ce possible écueil, il peut être nécessaire d'introduire un espaceur entre le gène codant pour la protéine d'intérêt et celui codant pour la protéine rapporteur. Cet espaceur est  
30 une séquence d'ADN codant pour un peptide de structure la plus neutre possible, et ne présentant pas en lui-même d'activité particulière. Il est généralement composé de 8 ou 9 acides aminés neutres qui ne doivent pas influencer sur la structure des protéines et ne doivent pas être la cible des protéases. Il est également préférable qu'il soit flexible et hydrophile (Sieber V., Methods Mol Biol., vol.230, p.45-55, 2003 ; Waldo G.S. *et al*,

Brevet US 6,448,087, 1999). Ce fragment d'ADN doit bien entendu être introduit de telle sorte que la phase soit conservée sur l'ensemble constitué par la protéine d'intérêt, l'espaceur, et la protéine rapporteur.

5 La position relative de la protéine d'intérêt et de la protéine rapporteur n'est pas neutre. Dans certains cas, il est préférable que la protéine de fusion soit positionnée en C-terminal. Dans certains autres, le positionnement en N-terminal est préférable. Fréquemment, l'homme du métier effectue les deux constructions, et retient celle qui présente le meilleur signal à l'issue d'une étape expérimentale préliminaire.

10 L'utilisation principale des protéines de fusion est donc associée à des travaux de localisation des protéines. Cependant, ces protéines de fusion ont également été utilisées comme système de sélection pour identifier des variants plus solubles d'une protéine : il a en effet été observé que lorsqu'on fusionne une protéine d'intérêt à une protéine rapporteur, la conformation de l'une et de l'autre sont liées. Une conformation  
15 incorrecte de la protéine située du côté NH<sub>2</sub> est ainsi corrélée à une conformation également incorrecte de la protéine localisée à l'extrémité COOH. L'inverse n'est généralement pas vrai : c'est la conformation de l'extrémité NH<sub>2</sub> de la protéine qui influe sur l'ensemble. La conformation de la protéine étant mise en place de façon concomitante à sa synthèse, et celle-ci ayant lieu de l'extrémité NH<sub>2</sub> vers l'extrémité  
20 COOH, on peut comprendre que la conformation globale de la protéine dépend de la conformation acquise par le premier segment synthétisé. Ces effets de conformation lors de la synthèse sont parfois décrits sous le terme de « nucléation ».

Cette corrélation entre la conformation de la protéine d'intérêt et celle de la protéine lui étant fusionnée a donc pu être utilisée pour mettre en place des systèmes de  
25 sélection de variants plus solubles. C'est naturellement que les protéines de fusion les plus classiques ont été utilisées : la CAT, la  $\beta$ -galactosidase, et la GFP.

L'équipe de Maxwell *et al.* (Protein science, vol.8, p.1908-1911, 1999), a pu exprimer une protéine de fusion réunissant une protéine insoluble avec la CAT : le domaine catalytique d'une intégrase HIV insoluble a été fusionné au domaine N-  
30 terminal de l'enzyme. D'un autre côté, la même construction contenant cette fois-ci un mutant de la protéase plus soluble a été réalisée. Une augmentation significative de la résistance à l'antibiotique chloramphénicol est observée chez la forme mutée. Cette technique permettrait donc de sélectionner une protéine soluble au milieu d'un ensemble de protéines insolubles.

Peu de publications décrivant des travaux basés sur cette approche ont été faites à ce jour.

5 Dans le brevet EP1479694, il est décrit une méthode pour isoler un fragment scFv soluble. La première étape consiste en la création d'une bibliothèque de mutants du fragment. Puis, il y a construction de deux types de protéines de fusion. Les gènes de ces variants sont, dans la première construction, fusionnés avec une partie du gène nécessaire à l'expression de la  $\beta$ -Galactosidase. La deuxième construction consiste à  
10 fusionner le gène codant pour l'antigène correspondant avec la dernière partie du gène de la  $\beta$ -Galactosidase. Ainsi, dans les conditions de sélection permettant l'expression de l'antigène et du ScFv mutant, seules les formes les plus solubles du ScFv présenteront une interaction et permettront le fonctionnement de la  $\beta$ -Galactosidase. La différenciation entre deux clones de solubilité différente repose ici sur le fait que la  
15 spécificité du fragment scFv est proportionnelle à sa solubilité. Plus le fragment sera soluble, plus la reconnaissance de l'antigène sera élevée et donc, plus la  $\beta$ -Galactosidase fonctionnera. Cette technique, basée sur la complémentation de deux domaines de la  $\beta$ -Galactosidase, est limitée à l'utilisation dans le cadre d'une liaison anticorps-antigène. La nécessité de co-transformer deux plasmides présente une complication additionnelle  
20 par rapport aux systèmes basés sur un seul plasmide. Par ailleurs, ce système sélectionne à la fois les ScFv plus solubles et ceux présentant une meilleure affinité pour l'antigène, ce qui peut introduire également une complexité non souhaitable.

Le système le plus utilisé aujourd'hui pour sélectionner les protéines plus  
25 solubles est basé sur l'utilisation de la GFP.

Dans le brevet US6448087 (Waldo G.S. *et al*), les auteurs décrivent une technique dans laquelle la GFP est fusionnée avec une banque de mutants d'une protéine d'intérêt. Les bactéries exprimant les gènes correspondant à des protéines plus solubles émettent la plus forte fluorescence, et peuvent aisément être sélectionnés en  
30 utilisant un trieur de cellules.

Cette approche a permis à plusieurs équipes d'obtenir des résultats positifs, c'est-à-dire d'identifier des mutants présentant une solubilité améliorée. Cependant, cette technique est limitée par le fait qu'il ne s'agit pas ici d'un système basé sur la

survie des clones positifs, mais sur une variation d'un signal qui devra être mesurée en utilisant un appareillage complexe.

5 Dans la demande de brevet 20040170976 (Scott L.A. *et al*), il est décrit un système dans lequel l'expression de la protéine d'intérêt sous forme insoluble entraîne un changement d'expression du gène rapporteur, en l'absence de protéine de fusion. Pour ce faire, on utilise un promoteur décrit comme spécifiquement activable si des protéines solubles sont présentes dans la bactérie. Le gène rapporteur placé sous le contrôle de ce promoteur particulier peut être par exemple choisi pour permettre  
10 d'observer un signal de fluorescence. Si le gène rapporteur code pour une enzyme, il faudra quantifier le produit des modifications enzymatiques d'un composé substrat. Le taux de protéines rapporteur produites sera directement proportionnel au taux de protéines d'intérêt solubles dans le cytoplasme.

15 Le brevet US 6,727,070 (Thomas P.J. *et al*) décrit un système de complémentation entre deux fragments d'une même protéine rapporteur : une protéine de fusion réunissant la protéine d'intérêt avec le premier fragment d'une protéine rapporteur qui, par lui-même, n'est pas actif, est utilisée. La deuxième partie du marqueur est exprimée au sein de la cellule hôte à partir du chromosome. Il y a  
20 manifestation de l'expression de la protéine rapporteur dans le cas où les deux parties peuvent se compléter. Cette protéine rapporteur peut être une enzyme, un inhibiteur de protéine, un fluorophore ou un chromophore. Une meilleure solubilité ou un niveau d'expression augmenté peuvent être détectés en utilisant les appareillages nécessaires.

25 Ainsi, donc, l'observation de la fluorescence, de l'expression d'une enzyme catalysant une réaction, ou la simple survie cellulaire, peut permettre de réaliser une ségrégation entre les types de variants solubles et insolubles d'une bibliothèque de mutants.

Les systèmes basés sur la survie ont sur ceux basés sur la fluorescence ou la  
30 colorimétrie la supériorité suivante : elles ne nécessitent pas la mesure d'un signal. Elles permettent donc la sélection directe des mutants souhaités.

La seule technique simple de sélection de mutants de solubilité ou d'expression à être basée sur la survie repose, dans l'art antérieur, sur l'utilisation du gène de

résistance au chloramphénicol. Cette approche n'est apparue que dans très peu de documents, ce qui suggère une efficacité faible.

Il est donc utile de développer une technique simple permettant de sélectionner de façon sensible et fiable les mutants de solubilité ou d'expression présents dans une  
5 banque.

## Description de l'invention

La présente invention est une technique permettant d'obtenir des mutants d'une protéine ou d'un peptide d'intérêt présentant une solubilité améliorée ou associés à un  
10 niveau d'expression supérieur. Le procédé selon l'invention est caractérisé par la génération d'une banque de mutants du gène encodant la protéine ou le peptide d'intérêt fusionné en phase avec le gène de résistance à la kanamycine, et par l'utilisation comme crible sélectif d'un milieu solide ou liquide contenant une concentration de kanamycine stringente. La kanamycine est un antibiotique peu coûteux et couramment utilisé dans  
15 les laboratoires de recherche. Il s'agit d'un bactéricide puissant, agissant indépendamment de la densité bactérienne, et ayant un effet à la fois très intense et très rapide *in vitro*.

De nombreuses familles de protéines peuvent conférer une résistance à  
20 l'antibiotique kanamycine par différents mécanismes. Parmi elles, trois familles de protéines agissent en modifiant directement l'antibiotique, détruisant ainsi son activité anti-bactérienne :

- Les kanamycine nucléotidyltransférases par O-nucléotidylation,
- Les kanamycine phosphotransférases par O-phosphorylation, et
- 25 - Les kanamycine acétyltransférases par N-acétylation.

La famille la plus couramment rencontrée et utilisée est la première famille, celle des kanamycine nucléotidyltransférases (Kat). Ces enzymes ont des structures très différentes et bien que conférant toutes une résistance à la kanamycine. Les inventeurs ont établi qu'elles ne sont pas toutes utilisables dans la méthode de la présente  
30 invention. En effet, elles ne fonctionnent pas toutes lorsqu'elles sont fusionnées avec une protéine en 5', même soluble. Notamment, la kanamycine nucléotidyltransférase (Kat) fonctionne très mal sous forme de fusion alors que la kanamycine phosphotransférase fonctionne très bien.

C'est pourquoi le gène de résistance à la kanamycine utilisé dans la présente invention code pour la kanamycine phosphotransférase. De préférence, la kanamycine phosphotransférase est la kanamycine kinase II, un variant ou un fragment fonctionnel de celle-ci. De préférence, le variant est une kanamycine kinase améliorée, présentant  
5 par exemple une plus grande résistance à l'antibiotique.

Par « kanamycine kinase II » est entendue la protéine de *Klebsiella pneumoniae* dont le code enzyme est EC 2.7.1.95, le numéro d'accès Genbank est U32991, et le numéro d'accès SwissProt est P00552. La kanamycine kinase II est également appelé aminoglycoside-3'-phospho-transférase-IIa, néomycine-kanamycine phosphotransférase  
10 II, APH(3')II, nptII or APH(3')-II. De préférence, la protéine présente la séquence décrite dans la SEQ ID No 14 et un exemple de séquence nucléotidique codant pour celle-ci est décrit dans la SEQ ID No 13.

Par « fragment fonctionnel » est entendu un fragment capable d'assurer une résistance à la kanamycine. Notamment, comme la résistance à la kanamycine est due à  
15 la capacité de l'enzyme à transférer sur la kanamycine un groupement phosphate, un fragment fonctionnel est capable de phosphoryler la kanamycine. De préférence, le fragment comprend le site catalytique de l'enzyme en position 190. De préférence, le fragment comprend au moins 100 acides aminés consécutifs. En particulier, un fragment préféré peut comprendre les résidus 23-257 (domaine APH-phosphotransférase). Dans  
20 un mode de réalisation préféré, la kanamycine kinase II peut comprendre une délétion de la Méthionine en position 1. Cette kanamycine kinase II est décrite dans les SEQ ID Nos 1 et 2.

Par « variant » de la kanamycine kinase II est entendu un polypeptide capable d'assurer une résistance à la kanamycine, notamment capable de transférer sur la  
25 kanamycine un groupement phosphate, et qui présente au moins 90 % d'identité avec la kanamycine kinase II de *Klebsiella pneumoniae* ou une séquence de 100, 150 ou 200 acides aminés consécutifs de celle-ci. De préférence, le variant présente au moins 95 %, 96 %, 97 %, 98 % ou 99% avec celle-ci ou une séquence de 100, 150 ou 200 acides aminés consécutifs de celle-ci. Les pourcentages d'identité auxquels il est fait référence  
30 ici sont établis, comme indiqué ci-dessus, à l'aide du logiciel BlastX, en utilisant les paramètres par défaut.

Dans un mode de réalisation particulier de la présente invention, la kanamycine kinase II utilisée est un variant améliorée de celle-ci permettant d'obtenir une résistance à des concentrations de kanamycine supérieures à la protéine sauvage. Plus

particulièrement, la kanamycine kinase II utilisée est une kanamycine kinase II améliorée, un variant ou un fragment fonctionnel de celle-ci comprenant au moins une substitution sélectionnée parmi le groupe consistant en Q155W et S114H.

5 Par « Q155W » est entendu que la glutamine en position 155 dans la séquence décrite SEQ ID No 14, ou en position correspondante dans une autre séquence, est substituée par un tryptophane. Par « S114H » est entendu que la sérine en position 114 dans la séquence décrite SEQ ID No 14, ou en position correspondante dans une autre séquence, est substituée par une histidine.

10 Dans un premier mode de réalisation de la présente invention, la kanamycine kinase II améliorée, un variant ou un fragment fonctionnel de celle-ci comprend une substitution Q155W. Un exemple de kanamycine kinase II comprenant une substitution Q155W est décrite en SEQ ID No 16. La séquence peut comprendre en outre une autre substitution, de préférence S114H.

15 Dans un second mode de réalisation de la présente invention, la kanamycine kinase II améliorée, un variant ou un fragment fonctionnel de celle-ci comprend une substitution S114H. Un exemple de kanamycine kinase II comprenant une substitution S114H est décrite en SEQ ID No 18. La séquence peut comprendre en outre une autre substitution, de préférence Q155W.

20 Les variants exprimés sous une forme soluble permettent à la bactérie de survivre en présence de kanamycine. De manière surprenante, contrairement au système de fusion avec la CAT, cette sélection directe par survie est remarquable en ce qu'on obtient une discrimination très nette entre les différents variants de la protéine, ce qui permet la sélection directe des clones exprimant les variants solubles ou mieux exprimés de la protéine ou du peptide d'intérêt.

25 Ainsi, la présente invention concerne un procédé d'obtention de variants solubles ou mieux exprimés d'une protéine ou d'un peptide d'intérêt, caractérisé par la génération d'une banque de mutants du gène codant pour ladite protéine ou ledit peptide d'intérêt fusionnés au gène de résistance à la kanamycine dans un vecteur d'expression, et par la sélection des gènes mutants correspondant à des protéines ou des peptides plus  
30 solubles ou mieux exprimées par culture des bactéries transformées par ladite banque de vecteurs d'expression dans un milieu de culture contenant de la kanamycine. De préférence, le gène de résistance à la kanamycine est un polynucléotide codant une kanamycine phosphotransférase, de préférence la kanamycine kinase II, un variant ou

un fragment fonctionnel de celle-ci. Dans un mode de réalisation préféré, le gène de résistance à la kanamycine est un polynucléotide codant une kanamycine kinase II améliorée comprenant au moins une substitution sélectionnée parmi le groupe consistant en Q155W et S114H. De préférence, le gène code pour une protéine d'intérêt. Le polypeptide d'intérêt peut être par exemple une cytokine, une enzyme, ou tout autre polypeptide/protéine présentant un intérêt industriel ou thérapeutique. En particulier, le vecteur d'expression est un vecteur d'expression procaryote. Facultativement, celui-ci peut comprendre les éléments permettant également une expression dans un hôte eucaryote.

10 Dans un mode de réalisation particulier, le procédé comprend :

i) Génération de mutants d'un gène codant pour la protéine ou le peptide d'intérêt ;

ii) Préparation d'une banque de mutants par clonage des mutants obtenus en i) en phase avec le gène de résistance à la kanamycine dans un vecteur d'expression ;

15 iii) Transformation de bactéries compétentes par la banque de mutants obtenue en ii) ; et

iv) Sélection, en utilisant un milieu de culture contenant de la kanamycine, des bactéries ayant été transformées par un vecteur comprenant un mutant correspondant à une protéine ou un peptide d'intérêt mutant(e) plus soluble ou mieux exprimé(e) que la protéine d'intérêt de départ.

De préférence, les mutants sont générés à l'étape i) par mutagenèse aléatoire. Dans une alternative également préférée, les mutants sont générés à l'étape i) par recombinaison de gènes.

Dans un autre mode de réalisation, le procédé comprend :

25 i) Fabrication d'un vecteur d'expression contenant le gène codant pour la protéine ou le peptide d'intérêt fusionné en phase avec le gène de résistance à la kanamycine ;

ii) Génération d'une banque de mutants du gène codant pour la protéine ou le peptide d'intérêt sur le vecteur d'expression obtenu en i) ;

30 iii) Transformation de bactéries compétentes par la banque de mutants obtenue en ii) ;

iv) Sélection, en utilisant un milieu de culture contenant de la kanamycine, des bactéries ayant été transformées par un vecteur comprenant un mutant correspondant à

une protéine ou un peptide d'intérêt mutant(e) plus soluble ou mieux exprimé(e) que la protéine d'intérêt de départ.

De préférence, les mutants sont générés à l'étape ii) par mutagenèse massive.

Facultativement, le gène codant pour la protéine ou le peptide d'intérêt et le gène de résistance à la kanamycine sont séparés par un espaceur ne changeant pas la phase de lecture.

De préférence, les bactéries sont transformées par électroporation ou par choc thermique. Le milieu de culture contenant de la kanamycine peut être solide ou liquide. De préférence, le milieu de culture contient une quantité de kanamycine stringente. De préférence, le vecteur d'expression est un plasmide.

La présente invention concerne en outre l'utilisation d'un vecteur d'expression codant une protéine de fusion entre un mutant d'une protéine ou d'un peptide d'intérêt et le gène de la résistance à la kanamycine pour sélectionner un variant de cette protéine ou de ce peptide présentant une solubilité améliorée ou une expression augmentée. De préférence, le gène de résistance à la kanamycine est un polynucléotide codant une kanamycine phosphotransférase, de préférence la kanamycine kinase II, un variant ou un fragment fonctionnel de celle-ci. Dans un mode de réalisation préféré, le gène de résistance à la kanamycine est un polynucléotide codant une kanamycine kinase II améliorée comprenant au moins une substitution sélectionnée parmi le groupe consistant en Q155W et S114H.

Le procédé de l'invention est un procédé d'obtention ou de sélection de mutants d'une protéine ou d'un peptide d'intérêt plus solubles ou mieux exprimés chez les bactéries, caractérisé par le fait qu'il contient les étapes suivantes dans cet ordre ou dans un ordre différent :

- i) On construit un plasmide contenant le gène d'intérêt, fusionné au gène de résistance à la kanamycine. L'ensemble est placé sous la dépendance d'un promoteur bactérien.
- ii) On introduit des mutations au niveau du gène d'intérêt.
- iii) On transforme des bactéries compétentes avec ces constructions.
- iv) On sélectionne, en utilisant une concentration de kanamycine stringente, les mutants ayant acquis une solubilité ou un niveau d'expression augmentés.

En particulier, les étapes i) et ii) peuvent être inversées.

Dans un mode de réalisation particulier, on peut éventuellement réaliser une étape supplémentaire iii)bis avant l'étape iv) afin de présélectionner les bactéries ayant  
5 intégré un plasmide. On utilise alors la résistance à un antibiotique autre que la kanamycine, tel que l'ampicilline. Dans ce mode de réalisation, le vecteur utilisé devra donc contenir, en plus du gène de résistance à la kanamycine fusionné aux gènes d'intérêt mutants, une cassette d'expression contenant le gène de résistance à l'antibiotique retenu (l'ampicilline par exemple), et le promoteur en permettant  
10 l'expression.

Lors de l'étape i), le gène de résistance à la kanamycine peut être fusionné à la protéine en N-terminal ou en C-terminal (figure 1) de la protéine d'intérêt. Dans un mode de réalisation préféré, ce gène de résistance est fusionné en C-terminal de la  
15 protéine d'intérêt, la conformation générale de la protéine étant principalement gouvernée par son extrémité NH<sub>2</sub>. La séquence codante pour le gène de résistance à la kanamycine est bien connu de l'homme du métier. Des exemples de séquences utiles dans la présente invention sont donnés en SEQ ID No 1, 13, 15 et 17. Il est entendu qu'un variant ou un fragment conservant l'activité biologique du gène de résistance à la  
20 kanamycine peut également être utilisé dans le procédé selon la présente invention.

Dans un mode de réalisation particulier, on insère un espaceur peptidique entre la protéine d'intérêt et le produit d'expression du gène de résistance à la kanamycine afin que les deux protéines soient distantes l'une de l'autre et puissent se replier sans  
25 interférer.

L'introduction des mutations à l'étape ii) peut être réalisée en utilisant l'une des méthodes de mutagenèse existantes.

Dans un premier mode de réalisation, on utilise la mutagenèse aléatoire. Dans le  
30 cas où aucun élément ne permet de définir a priori la nature et la position des changements susceptibles d'apporter une amélioration de la propriété physique étudiée, il est nécessaire de produire un grand nombre de molécules possédant des mutations de position et de nature différentes, afin de maximiser les chances que se trouve parmi elles une molécule correspondant à une solubilité améliorée. Les mutants sont générés sous

forme d'un amplifiat PCR, destiné ensuite à être cloné en phase avec le gène de résistance à la kanamycine (figure 1). Comme dans toute étape de ligation double-brin, on ne peut éviter que certains des clones obtenus soient constitués d'un vecteur reliqué sur lui-même. Ces clones sont ici très gênants, parce qu'ils seront associés à l'expression du gène de résistance à la kanamycine, et apparaîtront lors de l'étape ultérieure de sélection comme des faux positifs. Il est néanmoins possible de limiter ce taux de faux positifs en effectuant une stratégie de clonage telle que les clones reliqués sur eux-mêmes ne pourront pas s'exprimer, par manque de codon initiateur ou de promoteur, par exemple, ceux-ci étant alors présent dans le fragment à cloner. Ces méthodes sont bien connues de l'homme du métier. Cependant, ces manœuvres compliquent la stratégie de clonage.

Dans un deuxième mode de réalisation, on utilise la recombinaison d'ADN pour générer la diversité. Cette approche s'apparente à la mutagenèse aléatoire parce qu'on ne maîtrise pas précisément la diversité que l'on introduit, mais en diffère cependant parce que ladite diversité est sélectionnée pour contenir un taux important de mutations améliorantes. En effet, seules les mutations présélectionnées par la nature pour conserver au moins une partie de l'activité sont retenues. Comme dans le cas de la mutagenèse aléatoire, une étape de ligation double-brin est généralement nécessaire, avec les conséquences en terme de taux de faux positifs décrits précédemment.

Dans un troisième mode de réalisation, on peut utiliser la mutagenèse dirigée pour générer la diversité. La mutagenèse dirigée permet d'introduire des mutations d'une façon parfaitement précise, mais à un faible débit. Un avantage supplémentaire de la mutagenèse dirigée dans le contexte est l'absence d'étape de ligation double brin, et le taux de faux positifs est donc limité. Cependant, la génération des mutants un par un n'est pas vraiment adaptée à un système de criblage ultérieur par sélection directe sur le critère de la survie. De plus, la connaissance des sites de mutations pour l'augmentation de la solubilité d'une protéine est une situation peu courante. La mutagenèse dirigée n'est donc pas la technique idéale pour l'amélioration du caractère soluble d'un polypeptide.

Dans un quatrième mode de réalisation, on peut utiliser la Massive Mutagenesis® (FR2813314). Ce procédé combine les avantages de la mutagenèse dirigée (contrôle de la nature des modifications obtenues en une position donnée), et de la mutagenèse aléatoire (obtention d'un grand nombre de mutations différentes réparties sur de nombreuses positions du fragment d'ADN à muter). Comme la mutagenèse

dirigée, cette technique n'intègre pas d'étape de ligation double-brin, ce qui est positif dans le contexte de l'invention pour éviter de produire des clones faux-positifs.

Lors de l'étape iii), la transformation peut être réalisée selon différentes techniques et principalement en utilisant une approche par choc thermique ou par électroporation. C'est notamment la taille de la diversité à obtenir qui orientera sur l'une ou l'autre de ces deux approches : la technique par électroporation permet en effet d'obtenir un nombre de transformants supérieur.

De même, différents types de bactéries peuvent être utilisés tels que *E. coli*, *Streptomyces* et *Bacillus*. Le choix se fera selon les conditions de culture, la protéine à exprimer et la quantité à produire. Dans le mode de réalisation le plus classique, c'est *E. Coli* qui est le plus souvent utilisé (figure 3).

A l'étape iv), on aura préalablement déterminé la concentration stringente de kanamycine à ajouter. Cette concentration sera déterminée par des tests utilisant la construction plasmidique contenant le gène d'intérêt non mutant fusionné au gène de résistance à la kanamycine. La concentration stringente sera suffisamment élevée pour ne plus permettre aux bactéries transformées par cette construction de pousser, ou pour ne leur permettre qu'une croissance ralentie. Dans ces conditions, les bactéries exprimant un variant plus soluble auront un taux de croissance accéléré. Dans un premier mode de réalisation, les cultures de la banque de mutants transformés seront faites en milieu liquide : la sélection des mutants les plus solubles se fera alors selon la vitesse de pousse (figure 4). Dans un deuxième mode de réalisation, elles peuvent aussi se faire en milieu solide : la sélection se fera en fonction du nombre et de la taille des colonies (figure 3).

Dans un premier mode de réalisation, la mutagenèse est réalisée par mutation aléatoire, et le fragment contenant les gènes d'intérêt mutants est introduit en 5' du gène de résistance à la kanamycine.

Une molécule d'ADN contenant le gène à muter est amplifiée dans des conditions particulières altérant les capacités de l'enzyme à répliquer fidèlement l'ADN. Celle-ci introduit au fil des cycles des mutations, c'est-à-dire des différences par rapport à la séquence initiale. A la fin de la réaction, un grand nombre de copies de la molécule initiale est obtenu, chacune de ces molécules comportant des mutations différentes. Ces

molécules sont présentes sous forme de banque, c'est-à-dire d'un mélange de molécules de nature différente (différant par la nature et la position de leurs mutations). Cette banque de molécules linéaires est introduite en 5' du gène de résistance à la kanamycine, en prenant garde à ce que le taux de molécules de vecteur relié sur lui même, susceptible d'exprimer le gène de résistance à la kanamycine, soit minimal.

Après transformation, les bactéries sont mises en présence d'une concentration sélective de kanamycine permettant de différencier les variants solubles des variants insolubles, préalablement déterminée. Seules les cellules exprimant un variant soluble peuvent survivre.

Dans un second mode de réalisation, la mutagenèse est réalisée en utilisant la Massive Mutagenesis®. Un plasmide est d'abord construit (figure 2), contenant une origine de répllication, un gène de résistance à l'ampicilline, et le gène d'intérêt fusionné en 5' et en phase avec le gène de résistance à la kanamycine.

- La matrice est préparée en utilisant un kit de préparation d'ADN plasmidique adapté.

- Les oligonucléotides mutants sont synthétisés.

- La matrice est mélangée avec l'ensemble des oligonucléotides mutants.

- La matrice est dénaturée par la chaleur de façon à disposer temporairement d'ADN simple-brin. Lors du retour à une température plus basse, certains ou tous les oligonucléotides présents dans le mélange se fixent sur la matrice au niveau de leur site d'homologie.

- Tous les éléments nécessaires pour réaliser une répllication de la matrice à partir des oligonucléotides mutants sont ajoutés, et notamment une polymérase, le tampon permettant son activité, des nucléotides triphosphates en quantité suffisante, les cofacteurs nécessaires. La réaction de répllication a ensuite lieu dans les conditions de température correspondant à l'activité maximale de la polymérase soit 68°C. L'objectif est de disposer de mutants contenant chacun des combinaisons de plusieurs mutations dirigées différentes.

Des bactéries *E. coli* utilisées comme cellules hôtes sont transformées. Les clones obtenus sont ensuite mis en culture dans un milieu contenant de l'ampicilline afin de ne sélectionner que les bactéries ayant intégré un plasmide.

Les bactéries ainsi présélectionnées sont ensuite incubées dans un milieu contenant une concentration stringente de kanamycine, de façon à ce que seules les

cellules exprimant un variant présentant une solubilité améliorée ou un niveau d'expression supérieur survivent.

Les protéines ou peptides d'intérêt peuvent être par exemple des enzymes  
5 industrielles, - utilisées dans les industries textiles, agroalimentaires et chimiques  
notamment -, ou des anticorps recombinants, en particulier des anticorps à chaîne  
unique.

#### **Description des Figures :**

10 Figure 1 : Schéma représentant la préparation d'une banque de mutants de la  
protéine ou du peptide d'intérêt. Dans un premier temps, une banque de mutants est  
préparée (mutagenèse) puis cette banque est clonée en phase avec le gène de résistance  
à la kanamycine dans un vecteur (insertion dans un plasmide).

15 Figure 2 : Schéma représentant la préparation d'une banque de mutants de la  
protéine ou du peptide d'intérêt. Dans un premier temps, le gène codant la protéine ou le  
peptide d'intérêt est cloné en phase avec le gène de résistance à la kanamycine dans un  
vecteur (insertion dans un plasmide), puis une banque de mutants est préparée  
(mutagenèse).

20 Figure 3 : Schéma représentant la transformation des bactéries compétentes par  
la banque de mutants de la protéine ou du peptide d'intérêt suivie par la culture en  
milieu solide sélectif contenant la kanamycine.

Figure 4 : Schéma représentant la transformation des bactéries compétentes par  
la banque de mutants de la protéine ou du peptide d'intérêt suivie par la culture en  
milieu liquide sélectif contenant la kanamycine.

25 Figure 5 : Schéma représentant le contrôle négatif.

Figure 6 : Graphique représentant la résistance à la kanamycine des bactéries  
transformées avec ContK, Pet23, TA2K ou TB2K6 (résultats de l'exemple 1).

30 Figure 7 : Graphique représentant la résistance au chloramphénicol des bactéries  
transformées avec ContCam, Pet+Cam, TA2Cam ou TB2Cam (résultats de l'exemple  
comparatif de l'exemple 1).

Figure 8 : Tableau représentant les résultats de l'exemple 2.

Figure 9 : Schéma de la construction pET11-KanK.

Figure 10 : Croissance des cellules transformées par un plasmide codant une kanamycine kinase II sauvage ou mutante (Q155W) en fonction de la concentration en kanamycine.

Figure 11 : Croissance des cellules transformées par un plasmide codant une kanamycine kinase II sauvage ou mutante (Q155W, S114H et Q155W+S114H) en fonction de la concentration en kanamycine.

Figure 12 : Efficacité de résistance à la kanamycine de Kat et KK sous forme de protéine de fusion.

Figure 13 : Croissance des cellules transformées par un plasmide codant une kanamycine kinase II mutante Q155W et deux protéines de fusion TA ou TB avec la kanamycine kinase II mutante Q155W en fonction de la concentration en kanamycine. TB et TA signifient respectivement un variant soluble et insoluble de la protéine FE65.

Les exemples qui suivent illustrent la présente demande.

15

## Exemples

### Exemple 1 : la protéine Fe65-PTB2

Les gènes codant pour un variant soluble et un variant insoluble de la protéine FE65-PTB2 ont été introduits dans le plasmide pet11. La construction a été faite de façon à fusionner ladite protéine à un gène de résistance à la kanamycine, en l'absence d'espaceur peptidique. Le vecteur possède par ailleurs un gène de résistance à l'ampicilline.

On a obtenu différentes constructions :

- Pet : plasmide pet11 qui est à la base des différentes constructions ; il ne contient pas de gène de résistance à la kanamycine et sert donc de contrôle négatif (figure 5).

- ContK : plasmide pet11 sur lequel a été cloné le gène de résistance à la kanamycine (SEQ ID Nos 1 et 2); cette construction sert de contrôle positif.

- TA2K : plasmide pet11 sur lequel a été cloné une protéine de fusion TA2-gène de résistance à la kanamycine (SEQ ID Nos 3 et 4); la protéine TA2 est un variant insoluble de la protéine prise comme modèle.

30

- TB2K : plasmide pet11 sur lequel a été cloné une protéine fusion TB2-gène de résistance à la kanamycine (SEQ ID Nos 5 et 6); la protéine TB2 est un variant soluble de la protéine prise comme modèle.

Des bactéries *E. Coli* BL21 DE3 ont ensuite été transformées par choc thermique, puis incubées en milieu liquide contenant de l'ampicilline à 100µl/ml à 37°C pendant 12h.

Après 12h de culture,  $10^{-9}$  UOD ont été utilisées pour ensemercer des cultures contenant une concentration variable de kanamycine.

On a testé une gamme d'antibiotique kanamycine allant de 100µg/ml à 0,75µg/ml afin de trouver la concentration stringente pour cet ensemble de mutants, c'est-à-dire la concentration permettant de différencier les mutants présentant une plus grande solubilité.

On a incubé les cultures et mesuré la DO au spectrophotomètre à 600nm, de façon à mesurer la croissance bactérienne.

15

On a alors constaté une différence de résistance d'un facteur 2,7 entre le variant soluble (demi-résistance = 90 µg / ml) et le variant insoluble (demi-résistance = 33 µg / ml) (figure 6).

20

Afin de pouvoir réaliser une comparaison, on a effectué la même expérience avec le gène de résistance au chloramphénicol.

- Contcam : plasmide pet11 sur lequel a été cloné le gène de résistance au chloramphénicol (SEQ ID Nos 7 et 8); cette construction sert de contrôle positif.

25

- TA2cam : plasmide pet11 sur lequel a été cloné une protéine de fusion TA2-gène de résistance au chloramphénicol (SEQ ID Nos 9 et 10); la protéine TA2 est un variant insoluble de la protéine prise comme modèle.

- TB2cam : plasmide pet11 sur lequel a été cloné une protéine de fusion TB2-gène de résistance au chloramphénicol (SEQ ID Nos 11 et 12); la protéine TB2 est un variant soluble de la protéine prise comme modèle.

30

L'expérience a été réalisée suivant les mêmes étapes que pour l'antibiotique kanamycine.

Les différentes cultures bactériennes ont été ensemercées à  $10^{-4}$  UOD / ml (préalablement testé comme représentant l'ensemencement minimum).

Une gamme de chloramphénicol allant de 100µg/ml à 12,3µg/ml a été testée afin de trouver la concentration stringente pour ce mélange de mutants, c'est-à-dire la concentration permettant de différencier les mutants présentant une plus grande solubilité.

5 Les cultures ont été incubées et la DO a été mesurée au spectrophotomètre à 600nm, de façon à mesurer la croissance bactérienne.

Dans cette expérience, on a constaté une résistance au chloramphénicol accrue d'un facteur 1,4 du variant soluble de la protéine (demi-résistance = 35µg/ml) par rapport au variant insoluble (demi-résistance = 25 µg/ml) (figure 7).

Ces résultats ont démontré que le gène de résistance à la kanamycine, en particulier la kanamycine kinase II, est utilisable comme rapporteur de la solubilité d'une protéine lui étant fusionnée. Le gène de résistance à la kanamycine est bien meilleur que le gène de résistance au chloramphénicol, pour lequel on constate que la fenêtre de variation de résistance des deux variants n'est pas très étendue et ne permet donc pas la sélection de mutants au sein d'une banque.

### Exemple 2 :

20 Afin d'appliquer la technique au criblage d'une banque de mutants, on a réalisé la même expérience que précédemment mais avec un mélange réunissant les deux variants : soluble et non soluble.

La concentration stringente de kanamycine déterminée dans l'exemple 1 est utilisée afin de ne sélectionner que les mutants exprimant la protéine soluble.

25 Les cultures ont été réalisées en milieu solide afin de pouvoir comptabiliser les colonies.

Plusieurs types de cultures ont été réalisées : une culture avec uniquement le variant exprimant la protéine soluble, une culture avec uniquement le variant exprimant la protéine insoluble et plusieurs cultures avec des mélanges des deux variants dans différentes proportions (figure 8).

30 On a obtenu un nombre de colonies proportionnel au mélange de protéines : pour un ensemencement ne contenant que des clones exprimant la protéine soluble, on a obtenu 17000 colonies ; pour un ensemencement ne contenant que des clones exprimant la protéine non soluble, on a obtenu 4 colonies ; pour un ensemencement réunissant

10% de clones exprimant la protéine soluble et 90% de clones exprimant la protéine insoluble, on a obtenu 4000 colonies.

Cette technique a permis de mettre en évidence la présence de mutants exprimant une protéine soluble dans un mélange de variants à une proportion pouvant  
5 descendre jusqu'à 0,1%.

### **Exemple 3:**

Dans le but de créer des mutants solubles d'une protéine d'intérêt, on réalise une banque de mutants d'un plasmide construit de telle sorte que le gène de résistance à la  
10 kanamycine soit fusionné en phase et en C-terminal de celui permettant d'exprimer la protéine d'intérêt.

La mutagenèse est réalisée en utilisant la technologie Massive Mutagenesis® :

- Le plasmide est mélangé avec un ensemble d'oligonucléotides mutants.
- Le mélange est dénaturé par la chaleur.
- 15 - Tous les éléments nécessaires à réaliser une réplication de la matrice à partir des oligonucléotides mutants sont ajoutés.

Les bactéries *E. coli* utilisées comme cellules hôtes sont transformées par électroporation. Les clones obtenus sont ensuite mis en culture en milieu liquide en présence d'ampicilline, de façon à présélectionner les bactéries ayant effectivement  
20 intégré un plasmide. La culture est ensuite utilisée pour ensemencher un milieu liquide contenant une concentration stringente de kanamycine, préalablement déterminée, de façon à sélectionner les cellules exprimant un variant soluble.

### **Exemple 4:**

25 Dans le but d'identifier des mutants solubles d'une protéine d'intérêt, on génère dans un premier temps de la diversité par mutagenèse aléatoire : une molécule d'ADN contenant le gène à muter est amplifiée par "error-prone PCR". Les molécules obtenues sont clonées dans un plasmide construit de façon à ce que les gènes mutants soient fusionnés en phase et en 5' du gène de résistance à la kanamycine.

30 La banque de mutants est ensuite criblée comme exposé dans l'exemple précédent, ou en utilisant un milieu solide pour différencier sur le critère du nombre ou de la taille les colonies correspondant à un variant plus soluble de la protéine d'intérêt.

**Exemple 5 :**

Dans le but de créer des mutants solubles d'une protéine d'intérêt, on réalise une banque de mutants d'un plasmide construit de telle sorte que le gène de résistance à la kanamycine soit fusionné en C-terminal de celui permettant d'exprimer la protéine d'intérêt.

On réalise des mutations par recombinaison d'ADN sur le gène d'intérêt :

- On isole plusieurs gènes homologues entre eux, correspondant à l'activité recherchée.
- On amplifie ces gènes au moyen d'oligonucléotides flanquants, en utilisant des cycles d'amplification très courts (STEP).
- On clone le fragment d'ADN obtenu de telle sorte que les gènes mutants soient fusionnés en phase et en 5' du gène de résistance à la kanamycine présent sur le vecteur.

La banque de mutants est ensuite criblée comme exposé dans l'exemple précédent.

**Exemple 6 :**

Le gène codant pour la kanamycine kinase II (APH(3')-IIa) a été cloné dans les sites NcoI et NotI du vecteur pET11 (Novagen) derrière le promoteur T7 pour donner la construction appelée pET11-KanK (Figure 9).

Une banque de mutants de la kanamycine kinase a ensuite été construite grâce à la technologie Massive Mutagenesis® (FR2813314): une diversité totale a été introduite sur chaque position 3 tours de mutagenèse ont été réalisés pour obtenir une banque avec une diversité finale de  $5.10^6$  clones. La souche utilisée était *E. coli* DH10B. Le séquençage de 24 clones pris au hasard a mis en évidence un nombre moyen de mutations par molécule de 2,5. De plus, il n'y avait pas de biais majeur dans les séquences.

La banque a ensuite été transformée dans la souche *E. coli* BL21(DE3) pour permettre l'expression de la kanamycine kinase et étalée sur différents milieux :

- milieux non sélectifs : ampicilline 100 µg/ml et kanamycine 25 µg/ml pour estimer l'efficacité de transformation de la banque
- milieu de sélection : kanamycine 300 µg/ml pour permettre la sélection de mutants résistants à des concentrations élevées en kanamycine

Le nombre de clones après transformation a été estimé à  $1.10^6$  par dénombrement sur milieu ampicilline. Le milieu contenant 300  $\mu\text{g/ml}$  de kanamycine a permis de sélectionner 170 clones. Après minipréparation des plasmides isolés de 12 d'entre eux, une étape de séquençage a été réalisée grâce aux oligos T7pro et T7ter.

5 La séquence des 12 clones a permis de mettre en évidence une seule et unique mutation sur la position 155 où un résidu glutamine est remplacé par un résidu tryptophane.

CAG  $\rightarrow$  TGG

Gln (Q)  $\rightarrow$  Trp (W)

10 Ce mutant Q155W a ensuite été reconstruit sur la matrice pET11-KanK par mutagenèse dirigée. La croissance en milieu liquide de ce mutant en présence de différentes concentrations en kanamycine a alors été comparée à celle du WT. Brièvement, des concentrations croissantes de kanamycine sont ajoutées à des cultures d'*E.coli* BL21(DE3) transformées par pET11, pET11-KanK et pET11-KanK\_Q155W  
15 diluées jusqu'à une  $\text{DO}_{600\text{nm}}$  de  $10^{-4}$ . Après 15h de croissance à  $37^\circ\text{C}$ , les DO des différentes cultures ont été mesurées.

Le résultat est présenté dans la Figure 10.

La structure de la kanamycine kinase II étant connue (Nurizzo *et al.*, 2003), on peut placer la mutation dans la structure de l'enzyme. Le résidu Q155 est placé sur une  
20 boucle flexible (résidus 151-166) qui borde la cavité de fixation du substrat kanamycine. Sur cette boucle, sont également présents des résidus cruciaux pour l'activité de l'enzyme comme D159 et E160. La proximité de ce résidu par rapport au site actif peut donc expliquer une augmentation de l'activité de l'enzyme.

Une nouvelle banque a ensuite été construite avec le mutant pET11-  
25 KanK\_Q155W comme matrice de départ. Le protocole suivi est le même que lors de la première étape de mutagenèse (3 tours successifs, séquençage 24 clones pour le contrôle qualité). Après transformation en BL21(DE3), la banque a été sélectionnée sur 1 mg/ml de kanamycine pour isoler des variants encore résistants à la kanamycine que le Q155W.

30 Après séquençage de quelques clones repiqués sur milieu sélectif, une nouvelle mutation apparaît : la sérine 114 est mutée en Histidine.

TCC  $\rightarrow$  CAC

Ser (S)  $\rightarrow$  His (H)

Ce nouveau mutant a été construit par mutagenèse dirigée sur une matrice pET11-KanK (simple mutant S114H) et sur une matrice pET11-KanK\_Q155W (double mutant Q155W, S114H). Le résultat du test de croissance en milieu liquide des différents mutants et du WT est présenté sur la Figure 11.

5 L'EC50 pour chaque clone a été définie comme la concentration de kanamycine pour laquelle la DO mesurée correspond à 50% de la DO maximale.

Dans ces conditions (dilution  $10^{-6}$  pour l'inoculum de départ), le calcul des EC50 est le suivant :

	EC50 (µg/ml)	Facteur d'amélioration / WT
pET11-KanK (WT)	47	
pET11-KanK::Q155W	170	3,6
pET11-KanK::S114H	320	6,4
pET11-KanK::Q155W,114H	720	14,4

10 Le facteur d'amélioration est important. En effet, il est compris entre 3 et 7 pour les mutants simples et est supérieur à 14 pour le mutant double. Il est à noter que l'effet des deux mutations est plus qu'additif, il constitue une synergie entre les deux mutations.

#### 15 **Exemple 7 :**

L'efficacité de résistance à la kanamycine de deux types de gènes de résistance à la kanamycine, à savoir la Kat (kanamycine nucléotidyltransférase) et la KK (kanamycine kinase II), sous forme de protéines de fusion avec des gènes rapporteurs (GFP ou l'interféron gamma) a été testée. Le résultat est présenté dans la figure 12. On  
 20 peut ainsi constater que, de manière surprenante, l'efficacité de résistance de Kat sous forme de protéine de fusion est sérieusement réduite, alors que l'efficacité de la KK est globalement préservée.

#### **Exemple 8 :**

25 Les résultats présentés ici reposent sur utilisation d'une protéine codée par le gène de résistance à l'antibiotique kanamycine (gène de la kanamycine kinase II) comme rapporteur de solubilité d'une protéine-cible (Fe65) fusionnée à ce dernier. Ce système est fondé sur l'hypothèse de transmission du caractère soluble ou insoluble de la protéine d'intérêt à la protéine conférant la résistance à la kanamycine.

Le but des expériences présentées ici était de générer une banque de mutants de la protéine d'intérêt Fe65 grâce à Massive Mutagenesis® et de vérifier que le système de sélection grâce à la résistance à la kanamycine permet de sélectionner directement des variants qui apportent une amélioration de solubilité.

5

#### Constructions

Plusieurs constructions sont utilisées pour réaliser les tests :

**pET11** : plasmide qui est à la base des différentes constructions ; c'est le contrôle négatif.

10 **pET11K** : plasmide pET11 dans lequel a été cloné le gène de résistance à la kanamycine. Cette construction sert de contrôle positif.

**TA2K** : plasmide pET11 dans lequel a été cloné le gène hybride Fe65TA2-gène de résistance à la kanamycine. La protéine TA2 constitue le variant insoluble de la protéine Fe65 modèle.

15 **TB2K** : plasmide pET11 dans lequel a été cloné le gène hybride Fe65TB2-gène de résistance à la kanamycine. La protéine TB2 est le variant soluble de la protéine Fe65.

#### Conditions déterminées précédemment

20 Concentration de kanamycine sélective pour discriminer les protéines solubles et insolubles (en milieux liquide et solide): 50 µg/ml

#### Expériences

##### *Criblage d'une mini-banque*

25 Le but des expériences présentées ici était de réaliser une mini-banque de variants sur le gène codant pour la protéine modèle Fe65 insoluble (TA2) grâce à Massive Mutagenesis® et de vérifier que le système kanamycine permet de sélectionner directement le mutant soluble. Ce dernier était contenu dans les oligonucléotides choisis pour la mutagenèse.

30

-13 oligonucléotides introduisant des mutations ont été dessinés, un d'entre eux correspondant à la mutation TB2 (SolPhe).

Oligonucléotides utilisés pour la construction de la banque :

	SoIN01D	GGCAGCGCCTAAGGATGAGTTGGTC	SEQ ID No 19
	SolQ02E	GAATGAGTTGGTCGAGAAGTTCCAAGTC	SEQ ID No 20
	SoIN03V	TATTACCTGGGGGTAGTACCTGTTGCTAA	SEQ ID No 21
5	SoIN04D	TAGATGTGATTGATGGGGCCCTCGAGT	SEQ ID No 22
	SolQ05E	AGCAGCCGTGAAGAATGGACCCCAAG	SEQ ID No 23
	SolQ06E	ACCATCTTGCACGAGCAGACAGAGGCAGT	SEQ ID No 24
	SolQ07E	ACCATCTTGCACCAGGAGACAGAGGCAGT	SEQ ID No 25
	SoIN08D	TTCTGGTGCAGCCCGATGCTGCCAGC	SEQ ID No 26
10	SolQ09E	TCAGAGGCTGTGGAGGCTGCGTGTCAT	SEQ ID No 27
	SolQ10E	ATGCTTCGCTACGAGAAGTGTGCGG	SEQ ID No 28
	SolQ11E	TCAGAGGCTGTGGAGGCTGCGTGTCAT	SEQ ID No 29
	SolPhe	AGTGTCGGGTGCGTTNCTCCTTCCTGG	SEQ ID No 30
	SolPro	ATGTCAGTGTGGCCNNGCTACCCTCACCA	SEQ ID No 31

15

- Un tour de Massive Mutagenesis® (PLCR) a été réalisé sur la construction pET11TA2 avec les oligonucléotides décrits précédemment; le résultat de la réaction a été transformé dans la souche d'*Escherichia coli* DH10B puis étalé sur boîte contenant de l'ampicilline. Le comptage du nombre de colonies obtenues a permis de vérifier que toute la diversité de la banque était représentée.

20

- Les colonies obtenues ont été raclées et une préparation d'ADN a été réalisée à partir de ces colonies.

25

- La préparation d'ADN correspondant à la banque a ensuite été transformée dans la souche d'*E. coli* BL21 (DE3) pour permettre l'expression des protéines de fusion.

- Le résultat de la transformation a été étalé sur milieu contenant de l'ampicilline 100 µg/ml (comme témoin de l'efficacité de transformation) et sur milieu sélectif contenant de la kanamycine à 50 µg/ml (qui doit permettre de sélectionner uniquement les mutants solubles).

30

Résultats: Sur milieu ampicilline, 1500 clones ont été comptés. La même quantité d'ADN transformée sur kanamycine a donné seulement 12 clones.

Des minipréparations d'ADN ont été réalisées sur chacun des 12 clones et les gènes hybrides ont été séquencés dans leur totalité pour déterminer quelles mutations ont été retenues.

5        Résultat : Sur les 12 clones séquencés, 11 possédaient la mutation correspondant au variant TB2.

- 7 n'ont que la mutation TB2 (TTC→TCC: oligo SolPhe)
- 3 ont TB2 + autre mutation (+ SolN08D, SolQ09E, SolQ10E)
- 1 a une seule mutation autre que TB2 (SolQ09E)

10

**Exemple 9 :**

Les gènes codant pour un variant soluble (TB) et un variant insoluble (TA) de la protéine FE65 ont été introduits dans le plasmide pET11. La construction a été faite de façon à fusionner ladite protéine à la kanamycine kinase possédant la mutation Q155W.

15        On a obtenu les constructions suivantes:

- pET11\_KanK:Q155W : plasmide pET11 dans lequel est clonée la séquence de la kanamycine kinase mutée sur la position 155 (Q155W)

20        - pET11\_TA-KanK:Q155W : plasmide pET11 dans lequel est clonée le variant insoluble de FE65 fusionné à la séquence de la kanamycine kinase mutée sur la position 155 (Q155W)

- pET11\_TB-KanK:Q155W : plasmide pET11 dans lequel est clonée le variant soluble de FE65 fusionné à la séquence de la kanamycine kinase mutée sur la position 155 (Q155W)

25        Des bactéries *E. Coli* BL21 DE3 ont ensuite été transformées par choc thermique, puis incubées en milieu liquide contenant de l'ampicilline à 100µl/ml à 37°C pendant 12h.

30        Après 12h de culture, 10<sup>-6</sup> UOD ont été utilisées pour ensemercer des cultures contenant une concentration variable de kanamycine (gamme d'antibiotique kanamycine allant de 400µg/ml à 12,5µg/ml). On a incubé les cultures et mesuré la DO au spectrophotomètre à 600nm, de façon à mesurer la croissance bactérienne.

On a alors constaté une différence de résistance d'un facteur 4 entre le variant soluble (demi-résistance = 156 µg / ml) et le variant insoluble (demi-résistance = 39 µg / ml) (Figure 13).

Ces résultats ont démontré que le gène de résistance à la kanamycine muté sur la position 155 est utilisable comme rapporteur de la solubilité d'une protéine lui étant fusionnée. La différence d'un facteur 4 entre le variant soluble et insoluble montre que le système est d'une grande sensibilité.

## REVENDICATIONS

1) Procédé d'obtention de variants solubles ou mieux exprimés d'une protéine ou d'un peptide d'intérêt, caractérisé par la génération d'une banque de mutants du gène codant pour ladite protéine ou ledit peptide d'intérêt fusionnés au polynucléotide codant une kanamycine phosphotransférase dans un vecteur d'expression, et par la sélection des gènes mutants correspondant à des protéines ou des peptides plus solubles ou mieux exprimées par culture des bactéries transformées par ladite banque de vecteurs d'expression dans un milieu de culture contenant de la kanamycine.

10

2) Procédé selon la revendication 1, caractérisé en ce que la kanamycine phosphotransférase est la kanamycine kinase II, un fragment ou un variant fonctionnel de celle-ci.

15

3) Procédé selon la revendication 2, caractérisé en ce que la kanamycine phosphotransférase est une kanamycine kinase II améliorée, un fragment ou un variant fonctionnel de celle-ci comprenant au moins une substitution sélectionnée parmi Q155W et S114H.

20

4) Procédé selon l'une des revendications 1-3, caractérisé en ce que le procédé comprend :

i) Génération de mutants d'un gène codant pour la protéine ou le peptide d'intérêt ;

25

ii) Préparation d'une banque de mutants par clonage des mutants obtenus en i) en phase avec le polynucléotide codant une kanamycine phosphotransférase dans un vecteur d'expression ;

iii) Transformation de bactéries compétentes par la banque de mutants obtenue en ii) ; et

30

iv) Sélection, en utilisant un milieu de culture contenant de la kanamycine, des bactéries ayant été transformées par un vecteur comprenant un mutant correspondant à une protéine ou un peptide d'intérêt mutant(e) plus soluble ou mieux exprimé(e) que la protéine d'intérêt de départ.

5) Procédé selon la revendication 4, caractérisé en ce que les mutants sont générés à l'étape i) par mutagenèse aléatoire.

6) Procédé selon la revendication 4, caractérisé en ce que les mutants sont  
5 générés à l'étape i) par recombinaison de gènes.

7) Procédé selon l'une des revendications 1-3, caractérisé en ce que le procédé comprend :

10 i) Fabrication d'un vecteur d'expression contenant le gène codant pour la protéine ou le peptide d'intérêt fusionné en phase avec le polynucléotide codant une kanamycine phosphotransférase ;

ii) Génération d'une banque de mutants du gène codant pour la protéine ou le peptide d'intérêt sur le vecteur d'expression obtenu en i) ;

15 iii) Transformation de bactéries compétentes par la banque de mutants obtenue en ii).

iv) Sélection, en utilisant un milieu de culture contenant de la kanamycine, des bactéries ayant été transformées par un vecteur comprenant un mutant correspondant à une protéine ou un peptide d'intérêt mutant(e) plus soluble ou mieux exprimé(e) que la protéine d'intérêt de départ.

20

8) Procédé selon la revendication 7, caractérisé en ce que les mutants sont générés à l'étape ii) par mutagenèse massive.

9) Procédé selon l'une quelconque des revendications 1 à 8, caractérisé en ce  
25 que le vecteur d'expression est constitué de telle sorte que le gène codant pour la protéine ou le peptide d'intérêt soit situé en 5' du gène de résistance à la kanamycine.

10) Procédé selon l'une quelconque des revendications 1 à 9, caractérisé en ce  
30 que le gène codant pour la protéine ou le peptide d'intérêt et le gène de résistance à la kanamycine sont séparés par un espaceur ne changeant pas la phase de lecture.

11) Procédé selon l'une quelconque des revendications 1 à 10, caractérisé en ce que les bactéries sont transformées par électroporation ou par choc thermique.

12) Procédé selon l'une quelconque des revendications 1 à 11, caractérisé en ce que le milieu de culture contenant de la kanamycine est solide ou liquide.

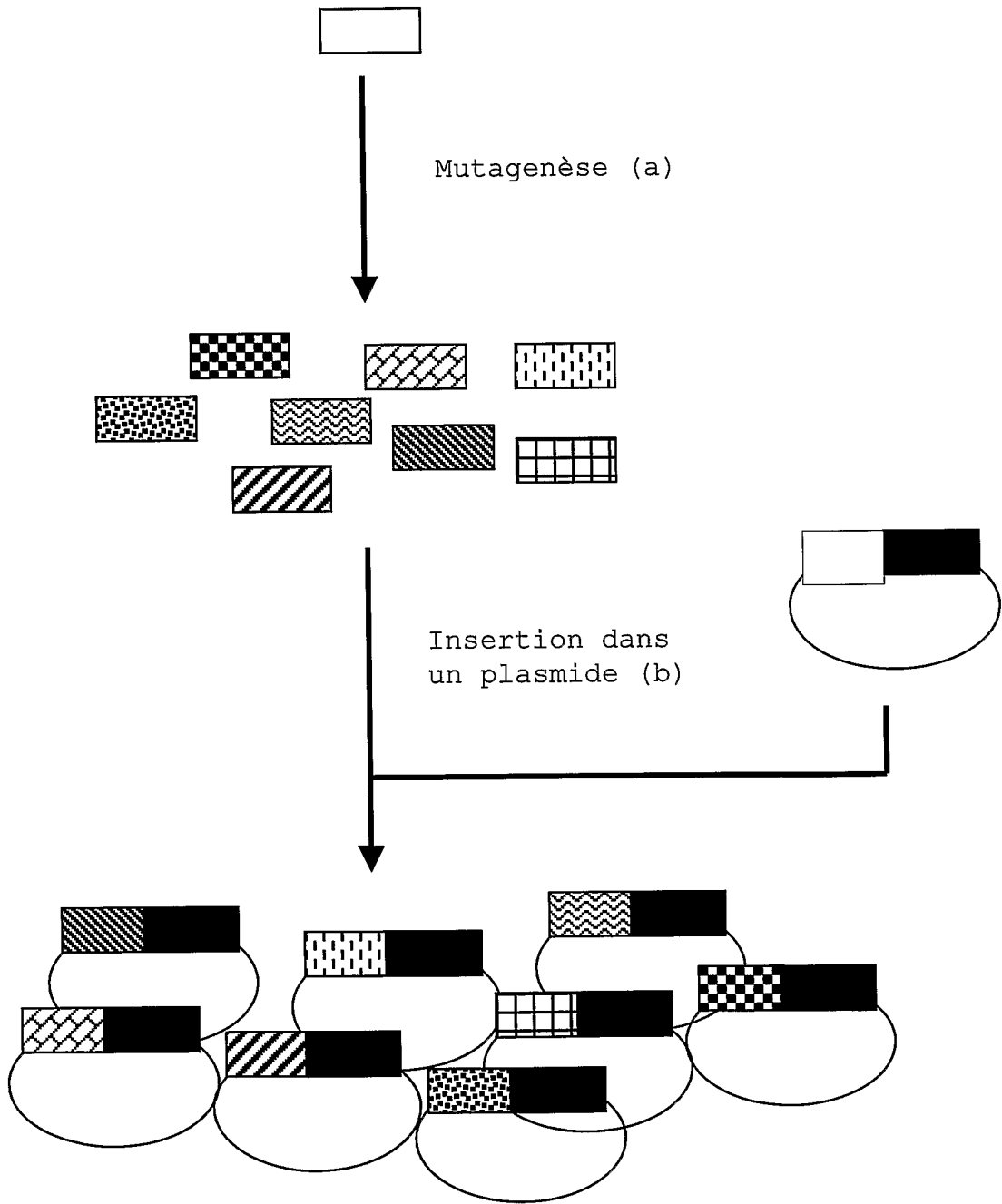
13) Procédé selon l'une quelconque des revendications 1 à 12, caractérisé en ce  
5 que le milieu de culture contient une quantité de kanamycine stringente.

14) Utilisation d'un vecteur d'expression codant une protéine de fusion entre un  
mutant d'une protéine ou d'un peptide d'intérêt et le polynucléotide codant une  
kanamycine phosphotransférase pour sélectionner un variant de cette protéine ou de ce  
10 peptide présentant une solubilité améliorée ou une expression augmentée.

15) Utilisation selon la revendication 14, caractérisée en ce que la kanamycine  
phosphotransférase est la kanamycine kinase II, un fragment ou un variant fonctionnel  
de celle-ci.  
15

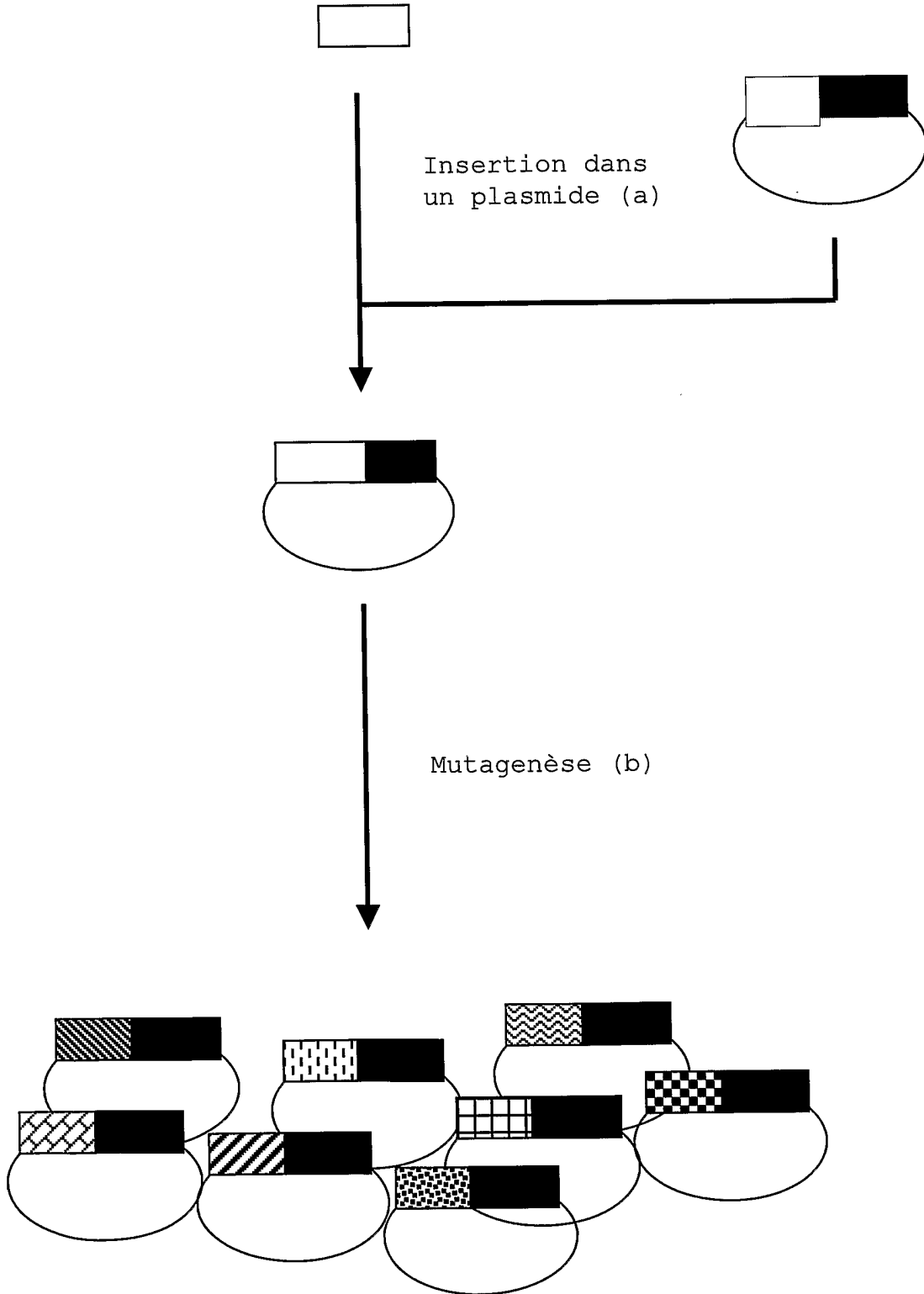
16) Utilisation selon la revendication 14 ou 15, caractérisée en ce que la  
kanamycine phosphotransférase est une kanamycine kinase II améliorée, un fragment  
ou un variant fonctionnel de celle-ci comprenant au moins une substitution sélectionnée  
parmi Q155W et S114H.  
20

1/10  
Figure 1



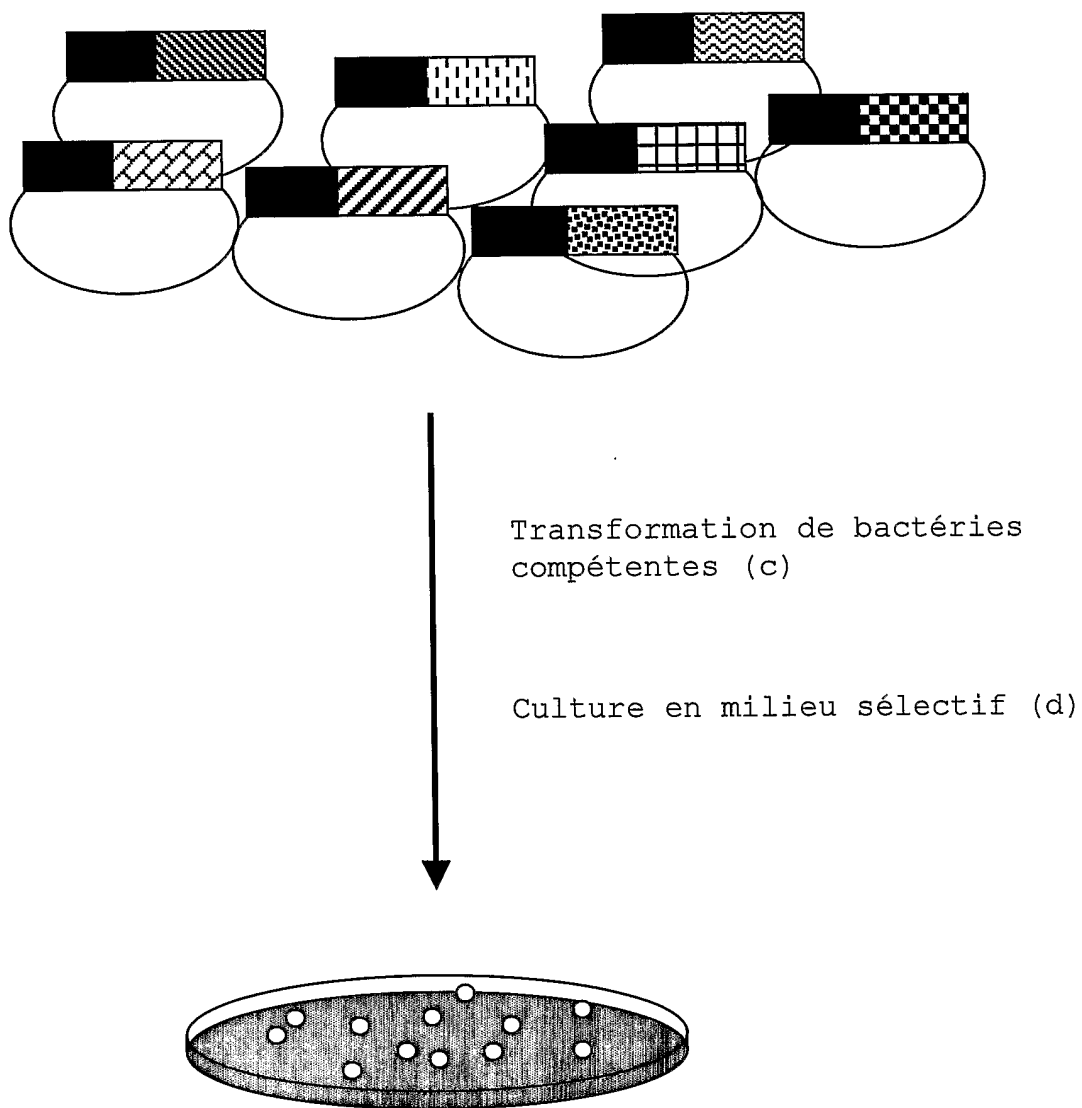
2/10

Figure 2



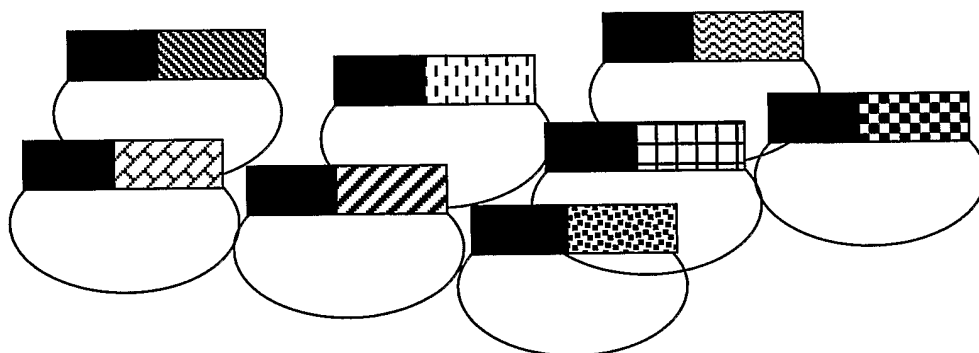
3/10

Figure 3



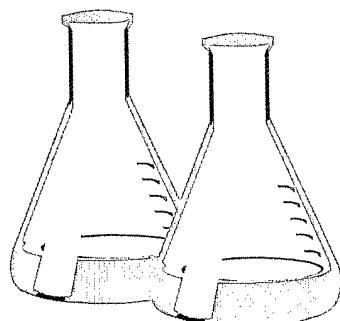
4/10

Figure 4



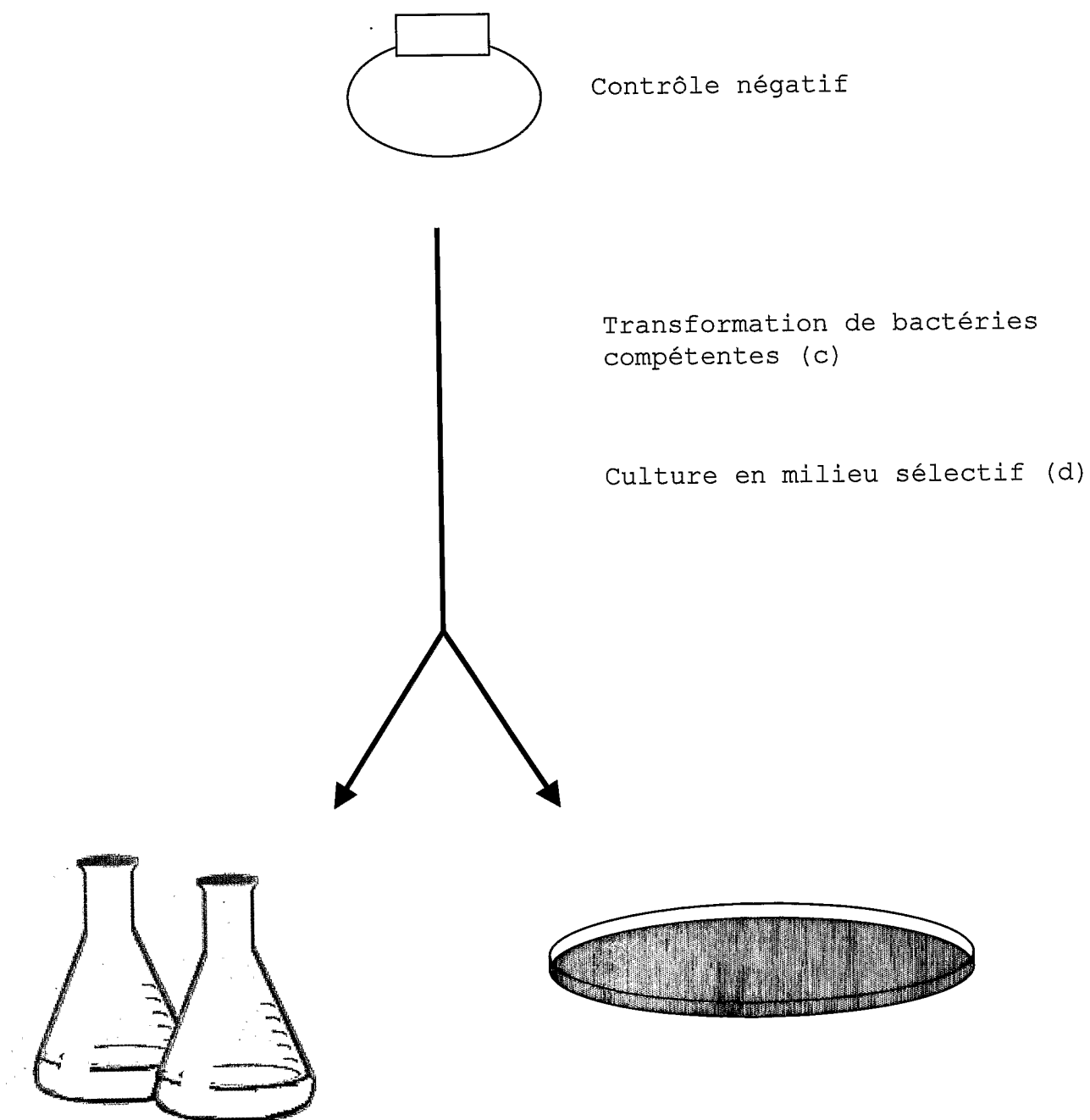
Transformation de bactéries  
compétentes (c)

Culture en milieu sélectif (d)



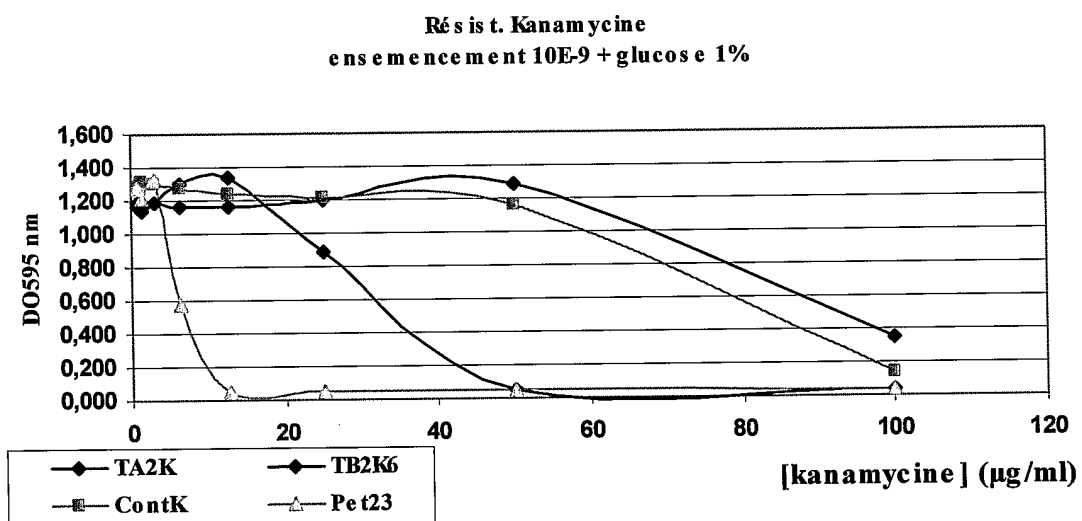
5/10

Figure 5



6/10

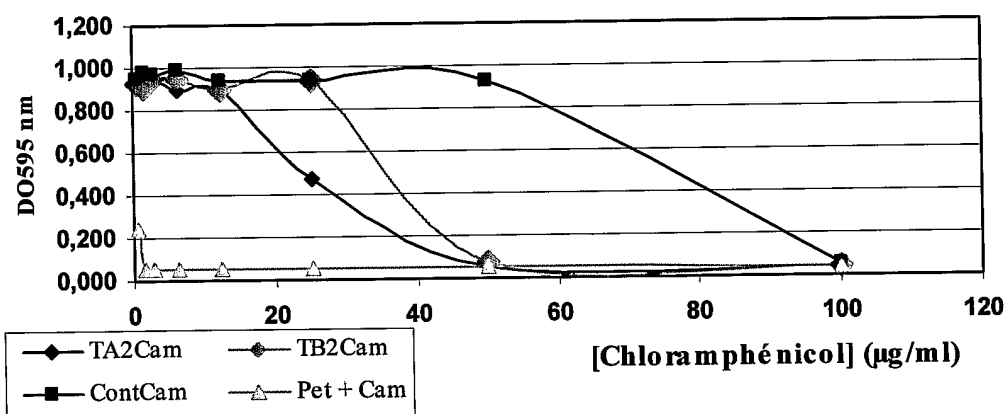
Figure 6



7/10

Figure 7

**Résist. Chloramphénicol ensemencement 10E-4 +  
glucose 1%**



8/10

Figure 8

	Kanamycine 50
TA2K (insoluble)	4
TB2K (soluble)	17000
TB2K (50%) + TA2K (50%)	15000
TB2K (10%) + TA2K (90%)	<b>4000</b>
TB2K (1%) + TA2K (99%)	<b>340</b>
TB2K (0,1%) + TA2K (99,9%)	<b>27</b>

9/10

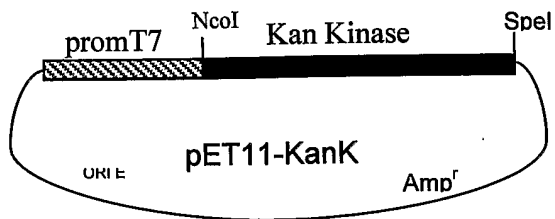


FIGURE 9

Croissance (Dilution 10<sup>-4</sup>) / concentration en kanamycine

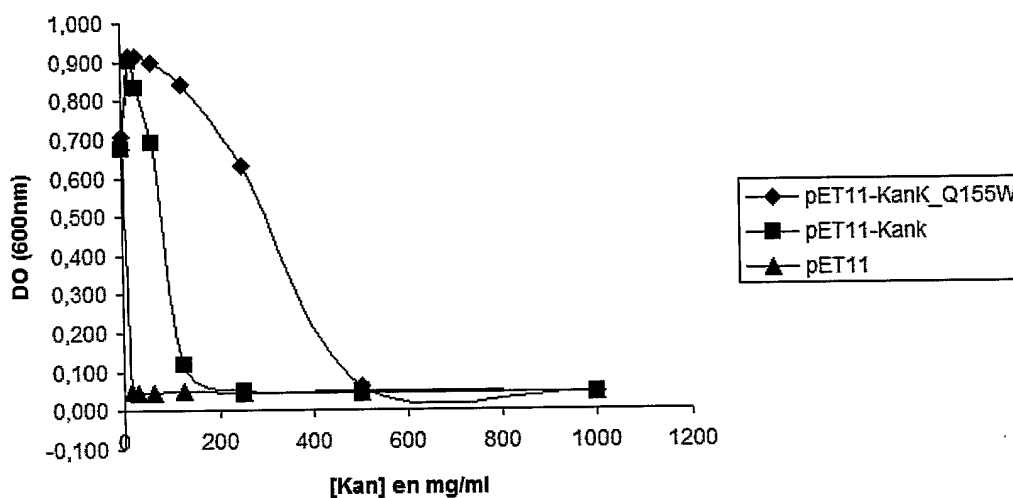


FIGURE 10

Croissance des différents clones (dilution 10<sup>-6</sup>) / concentration en kanamycine

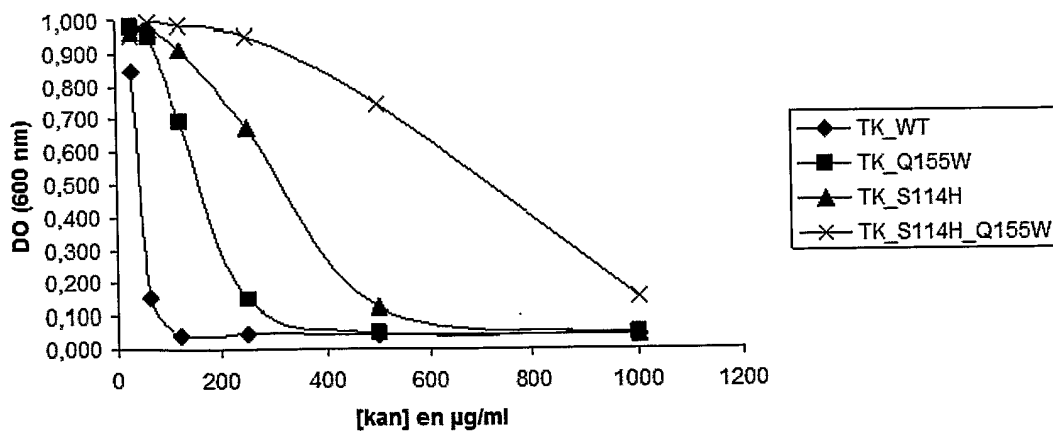


FIGURE 11

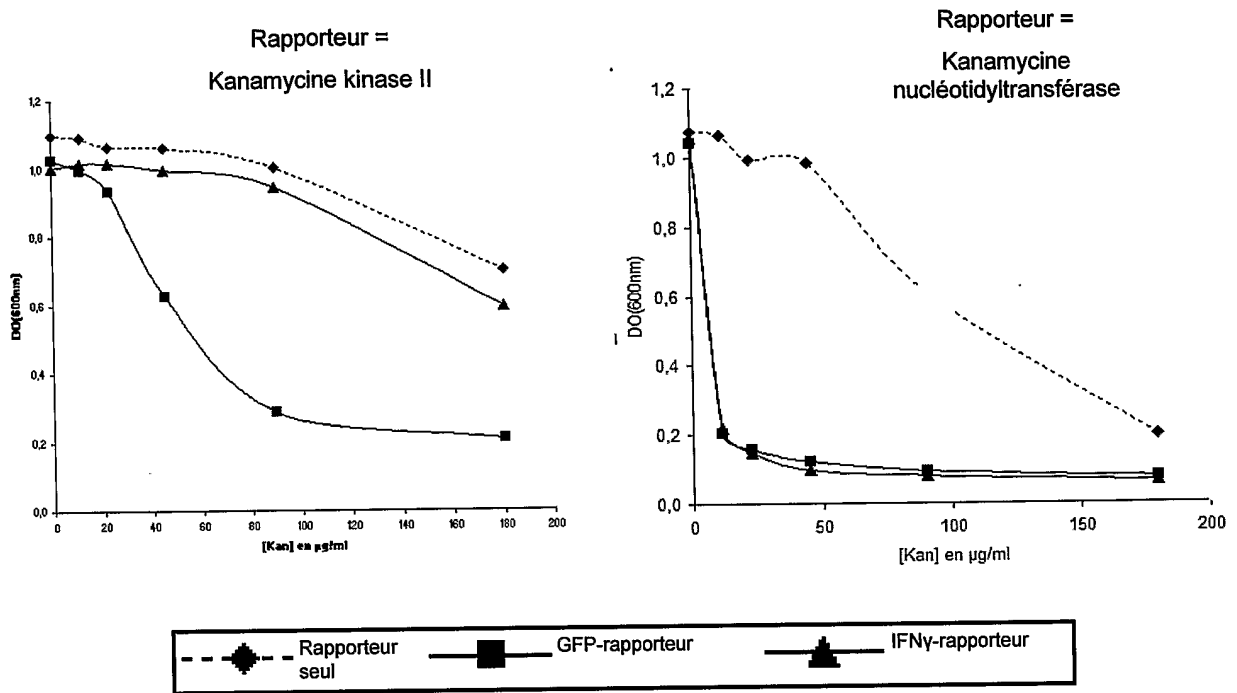


FIGURE 12

**Croissance (dilution 10<sup>-6</sup>) des clones exprimant les protéine de fusion TA/TB- kanK:Q155W**

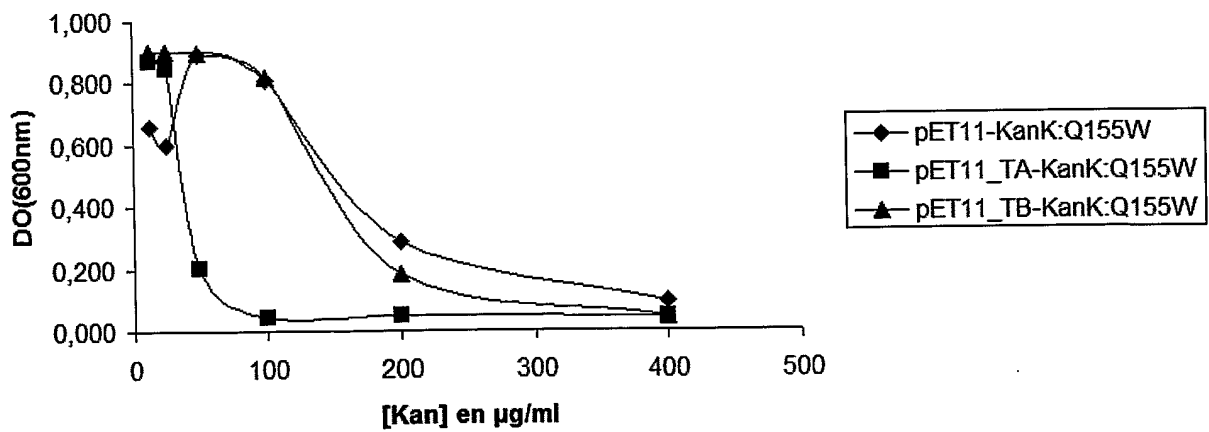


FIGURE 13

1  
SEQUENCE LISTING

<110> BIOMETHODES

<120> Méthode d'obtention de variants solubles ou mieux exprimés d'une protéine d'intérêt

<130> B0361WO

<150> FR0503366

<151> 2005-04-05

<160> 31

<170> PatentIn version 3.1

<210> 1

<211> 792

<212> DNA

<213> Klebsiella pneumoniae

<220>

<221> CDS

<222> (1)..(792)

<223>

<400> 1	
att gaa caa gat gga ttg cac gca ggt tct ccg gcc gct tgg gtg gag	48
Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu	
1 5 10 15	
agg cta ttc ggc tat gac tgg gca caa cag aca atc ggc tgc tct gat	96
Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp	
20 25 30	
gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg gtt ctt ttt gtc	144
Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val	
35 40 45	
aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag gac gag gca gcg	192
Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala	
50 55 60	
cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc gca gct gtg ctc	240

Arg Leu Ser Trp Leu Ala Thr Thr Gly Val<sup>2</sup> Pro Cys Ala Ala Val Leu  
 65 70 75 80

gac gtt gtc act gaa gcg gga agg gac tgg ctg cta ttg ggc gaa gtg 288  
 Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val  
 85

ccg ggg cag gat ctc ctg tca tct cac ctt gct cct gcc gag aaa gta 336  
 Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val  
 100

tcc atc atg gct gat gca atg cgg cgg ctg cat acg ctt gat ccg gct 384  
 Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala  
 115

acc tgc cca ttc gac cac caa gcg aaa cat cgc atc gag cga gca cgt 432  
 Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala Arg  
 130

act cgg atg gaa gcc ggt ctt gtc gat cag gat gat ctg gac gaa gag 480  
 Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu  
 145 150 155 160

cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg ctc aag gcg cgc 528  
 His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg  
 165 170 175

atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc gat gcc tgc ttg 576  
 Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys Leu  
 180 185 190

ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga ttc atc gac tgt 624  
 Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys  
 195 200 205

ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata gcg ttg gct acc 672  
 Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr  
 210 215 220

cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct gac cgc ttc ctc 720  
 Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu  
 225 230 235 240

gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc atc gcc ttc tat 768  
 Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr  
 245 250 255

cgc ctt ctt gac gag ttc ttc tga 792  
 Arg Leu Leu Asp Glu Phe Phe  
 260

<210> 2

<211> 263

<212> PRT

<213> Klebsiella pneumoniae

<400> 2

Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu  
 1 5 10 15

Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp





5

185		190		195		200	
gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc gca							678
Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala							
		205		210		215	
gct gtg ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta ttg							726
Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu							
		220		225		230	
ggc gaa gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct gcc							774
Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala							
		235		240		245	
gag aaa gta tcc atc atg gct gat gca atg cgg cgg ctg cat acg ctt							822
Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu							
		250		255		260	
gat ccg gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc gag							870
Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu							
		265		270		275	280
cga gca cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat ctg							918
Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu							
		285		290		295	
gac gaa gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg ctc							966
Asp Glu Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu							
		300		305		310	
aag gcg cgc atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc gat							1014
Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp							
		315		320		325	
gcc tgc ttg ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga ttc							1062
Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe							
		330		335		340	
atc gac tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata gcg							1110
Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala							
		345		350		355	360
ttg gct acc cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct gac							1158
Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp							
		365		370		375	
cgc ttc ctc gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc atc							1206
Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile							
		380		385		390	
gcc ttc tat cgc ctt ctt gac gag ttc ttc tga tgtacaagtg aagatc							1255
Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe							
		395		400			

<210> 4

<211> 402

<212> PRT

<213> artificial sequence

<220>

<223> Vecteur TA2Kan

6

&lt;400&gt; 4

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

His Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala  
 275 280 285

Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala  
 290 295 300

Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu  
 305 310 315 320

Asp Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val  
 325 330 335

Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val  
 340 345 350

Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu  
 355 360 365

Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile  
 370 375 380

Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu  
 385 390 395 400

Phe Phe

<210> 5

<211> 1255

<212> DNA

<213> artificial sequence

<220>

<223> vecteur TB2Kan

<220>

<221> CDS

<222> (31)..(1239)

<223>

<400> 5  
 ttttgtttaa ctttaagaag gagaggatcc atg gca gcg cct aag aat gag ttg 54  
 Met Ala Ala Pro Lys Asn Glu Leu  
 1 5

gtc cag aag ttc caa gtc tat tac ctg ggg aat gta cct gtt gct aaa 102  
 Val Gln Lys Phe Gln Val Tyr Tyr Leu Gly Asn Val Pro Val Ala Lys  
 10 15 20

cct gtt ggg gta gat gtg att aat ggg gcc ctc gag tca gtc ctg tcc Pro Val Gly Val Asp Val Ile Asn Gly Ala Leu Glu Ser Val Leu Ser 25 30 35 40	150
tcc agc agc cgt gaa caa tgg acc cca agt cat gtc agt gtg gcc cct Ser Ser Ser Arg Glu Gln Trp Thr Pro Ser His Val Ser Val Ala Pro 45 50 55	198
gct acc ctc acc atc ttg cac cag cag aca gag gca gtg ctg gga gag Ala Thr Leu Thr Ile Leu His Gln Gln Thr Glu Ala Val Leu Gly Glu 60 65 70	246
tgt cgg gtg cgt tcc ctc tcc ttc ctg gcc gtg ggc aga gat gtc cac Cys Arg Val Arg Ser Leu Ser Phe Leu Ala Val Gly Arg Asp Val His 75 80 85	294
acg ttt gca ttc atc atg gct gcc ggc cca gcc tcc ttc tgc tgc cac Thr Phe Ala Phe Ile Met Ala Ala Gly Pro Ala Ser Phe Cys Cys His 90 95 100	342
atg ttc tgg tgc gag ccc aat gct gcc agc ctc tca gag gct gtg cag Met Phe Trp Cys Glu Pro Asn Ala Ala Ser Leu Ser Glu Ala Val Gln 105 110 115 120	390
gct gcg tgc atg ctt cgc tac cag aag tgt gcg gcc gca gga agc tct Ala Ala Cys Met Leu Arg Tyr Gln Lys Cys Ala Ala Ala Gly Ser Ser 125 130 135	438
ggc tcc atc att gaa caa gat gga ttg cac gca ggt tct ccg gcc gct Gly Ser Ile Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala 140 145 150	486
tgg gtg gag agg cta ttc ggc tat gac tgg gca caa cag aca atc ggc Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly 155 160 165	534
tgc tct gat gcc tcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg gtt Cys Ser Asp Ala Ser Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val 170 175 180	582
ctt ttt gtc aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag gac Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp 185 190 200	630
gag gca gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc gca Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala 205 210 215	678
gct gtg ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta ttg Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu 220 225 230	726
ggc gaa gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct gcc Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala 235 240 245	774
gag aaa gta tcc atc atg gct gat gca atg cgg cgg ctg cat acg ctt Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu 250 255 260	822
gat ccg gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc gag Asp Pro Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu 265 270 275 280	870
cga gca cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat ctg Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu 285 290 295	918

9

gac gaa gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg ctc	966
Asp Glu Glu His 300 Gln Gly Leu Ala Pro 305 Ala Glu Leu Phe Ala Arg Leu 310	
aag gcg cgc atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc gat	1014
Lys Ala Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp 315 320 325	
gcc tgc ttg ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga ttc	1062
Ala Cys Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe 330 335 340	
atc gac tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata gcg	1110
Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala 345 350 355 360	
ttg gct acc cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct gac	1158
Leu Ala Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp 365 370 375	
cgc ttc ctc gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc atc	1206
Arg Phe Leu Val 380 Leu Tyr Gly Ile Ala 385 Ala Pro Asp Ser Gln Arg Ile 390	
gcc ttc tat cgc ctt ctt gac gag ttc ttc tga tgtacaagtg aagatc	1255
Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe 395 400	

<210> 6

<211> 402

<212> PRT

<213> artificial sequence

<220>

<223> vecteur TB2Kan

<400> 6

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

Gly Pro Ala Ser Phe Cys Cys His Met Phe Trp Cys Glu Pro Asn Ala  
 100 105 110

Ala Ser Leu Ser Glu Ala Val Gln Ala Ala Cys Met Leu Arg Tyr Gln  
 115 120 125

Lys Cys Ala Ala Ala Gly Ser Ser Gly Ser Ile Ile Glu Gln Asp Gly  
 130 135 140

Leu His Ala Gly Ser Pro Ala Ala Trp Val Glu Arg Leu Phe Gly Tyr  
 145 150 155 160 165

Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser Asp Ala Ser Val Phe Arg  
 165 170 175

Leu Ser Ala Gln Gly Arg Pro Val Leu Phe Val Lys Thr Asp Leu Ser  
 180 185 190

Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala Ala Arg Leu Ser Trp Leu  
 195 200 205

Ala Thr Thr Gly Val Pro Cys Ala Ala Val Leu Asp Val Val Thr Glu  
 210 215 220

Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu Val Pro Gly Gln Asp Leu  
 225 230 235 240

Leu Ser Ser His Leu Ala Pro Ala Glu Lys Val Ser Ile Met Ala Asp  
 245 250 255

Ala Met Arg Arg Leu His Thr Leu Asp Pro Ala Thr Cys Pro Phe Asp  
 260 265 270

His Gln Ala Lys His Arg Ile Glu Arg Ala Arg Thr Arg Met Glu Ala  
 275 280 285

Gly Leu Val Asp Gln Asp Asp Leu Asp Glu Glu His Gln Gly Leu Ala  
 290 295 300

Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala Arg Met Pro Asp Gly Glu  
 305 310 315 320

Asp Leu Val Val Thr His Gly Asp Ala Cys Leu Pro Asn Ile Met Val  
 325 330 335

Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp Cys Gly Arg Leu Gly Val  
 340 345 350

Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala Thr Arg Asp Ile Ala Glu  
 355 360 365

Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe Leu Val Leu Tyr Gly Ile

370

375

380

Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe Tyr Arg Leu Leu Asp Glu  
 385 390 395 400

Phe Phe

<210> 7

<211> 660

<212> DNA

<213> artificial sequence

<220>

<223> Gène de résistance au chloramphénicol

<220>

<221> CDS

<222> (1)..(660)

<223>

<400> 7

atg gag aaa aaa atc act gga tat acc acc gtt gat ata tcc caa tgg	48
Met Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp	
1 5 10 15	
cat cgt aaa gaa cat ttt gag gca ttt cag tca gtt gct caa tgt acc	96
His Arg Lys Glu His Phe Glu Ala Phe Gln Ser Val Ala Gln Cys Thr	
20 25 30	
tat aac cag acc gtt cag ctg gat att acg gcc ttt tta aag acc gta	144
Tyr Asn Gln Thr Val Gln Leu Asp Ile Thr Ala Phe Leu Lys Thr Val	
35 40 45	
aag aaa aat aag cac aag ttt tat ccg gcc ttt att cac att ctt gcc	192
Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile His Ile Leu Ala	
50 55 60	
cgc ctg atg aat gct cat ccg gaa ttt cgt atg gca atg aaa gac ggt	240
Arg Leu Met Asn Ala His Pro Glu Phe Arg Met Ala Met Lys Asp Gly	
65 70 75 80	
gag ctg gtg ata tgg gat agt gtt cac cct tgt tac acc gtt ttc cat	288
Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His	
85 90 95	
gag caa act gaa acg ttt tca tcg ctc tgg agt gaa tac cac gac gat	336
Glu Gln Thr Glu Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp	
100 105 110	
ttc cgg cag ttt cta cac ata tat tcg caa gat gtg gcg tgt tac ggt	384
Phe Arg Gln Phe Leu His Ile Tyr Ser Gln Asp Val Ala Cys Tyr Gly	
115 120 125	
gaa aac ctg gcc tat ttc cct aaa ggg ttt att gag aat atg ttt ttc	432

										12									
Glu	Asn	Leu	Ala	Tyr	Phe	Pro	Lys	Gly	Phe	Ile	Glu	Asn	Met	Phe	Phe				
	130					135					140								
gtc	tca	gcc	aat	ccc	tgg	gtg	agt	ttc	acc	agt	ttt	gat	tta	aac	gtg				480
Val	Ser	Ala	Asn	Pro	Trp	Val	Ser	Phe	Thr	Ser	Phe	Asp	Leu	Asn	Val				160
	145				150					155									
gcc	aat	atg	gac	aac	ttc	ttc	gcc	ccc	gtt	ttc	acc	atg	ggc	aaa	tat				528
Ala	Asn	Met	Asp	Asn	Phe	Phe	Ala	Pro	Val	Phe	Thr	Met	Gly	Lys	Tyr				175
				165					170										
tat	acg	caa	ggc	gac	aag	gtg	ctg	atg	ccg	ctg	gcg	att	cag	gtt	cat				576
Tyr	Thr	Gln	Gly	Asp	Lys	Val	Leu	Met	Pro	Leu	Ala	Ile	Gln	Val	His				190
			180					185											
cat	gcc	gtt	tgt	gat	ggc	ttc	cat	gtc	ggc	aga	atg	ctt	aat	gaa	tta				624
His	Ala	Val	Cys	Asp	Gly	Phe	His	Val	Gly	Arg	Met	Leu	Asn	Glu	Leu				205
		195					200												
caa	cag	tac	tgc	gat	gag	tgg	cag	ggc	ggg	gcg	taa								660
Gln	Gln	Tyr	Cys	Asp	Glu	Trp	Gln	Gly	Gly	Ala									210
	210					215													

<210> 8

<211> 219

<212> PRT

<213> artificial sequence

<220>

<223> Gène de résistance au chloramphénicol

<400> 8

Met Glu Lys Lys Ile Thr Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp  
1 5 10 15

His Arg Lys Glu His Phe Glu Ala Phe Gln Ser Val Ala Gln Cys Thr  
20 25 30

Tyr Asn Gln Thr Val Gln Leu Asp Ile Thr Ala Phe Leu Lys Thr Val  
35 40 45

Lys Lys Asn Lys His Lys Phe Tyr Pro Ala Phe Ile His Ile Leu Ala  
50 55 60

Arg Leu Met Asn Ala His Pro Glu Phe Arg Met Ala Met Lys Asp Gly  
65 70 75 80

Glu Leu Val Ile Trp Asp Ser Val His Pro Cys Tyr Thr Val Phe His  
85 90 95

Glu Gln Thr Glu Thr Phe Ser Ser Leu Trp Ser Glu Tyr His Asp Asp  
100 105 110

Phe Arg Gln Phe Leu His Ile Tyr Ser Gln Asp Val Ala Cys Tyr Gly

13

	115		120		125										
Glu	Asn 130	Leu	Ala	Tyr	Phe	Pro 135	Lys	Gly	Phe	Ile	Glu 140	Asn	Met	Phe	Phe
Val 145	Ser	Ala	Asn	Pro	Trp 150	Val	Ser	Phe	Thr	Ser 155	Phe	Asp	Leu	Asn	Val 160
Ala	Asn	Met	Asp	Asn 165	Phe	Phe	Ala	Pro	Val 170	Phe	Thr	Met	Gly	Lys 175	Tyr
Tyr	Thr	Gln	Gly 180	Asp	Lys	Val	Leu	Met 185	Pro	Leu	Ala	Ile	Gln 190	Val	His
His	Ala	Val 195	Cys	Asp	Gly	Phe	His 200	Val	Gly	Arg	Met	Leu 205	Asn	Glu	Leu
Gln	Gln 210	Tyr	Cys	Asp	Glu	Trp 215	Gln	Gly	Gly	Ala					

<210> 9

<211> 1120

<212> DNA

<213> artificial sequence

<220>

<223> Vecteur TA2Cam

<220>

<221> CDS

<222> (31)..(1104)

<223>

<400> 9

	ttttgtttaa ctttaagaag gagaggatcc atg gca gcg cct aag aat gag ttg <span style="margin-left: 100px;">Met Ala Ala Pro Lys Asn Glu Leu</span>	54
	<span style="margin-left: 100px;">1</span>	
	gtc cag aag ttc caa gtc tat tac ctg ggg aat gta cct gtt gct aaa Val Gln Lys Phe Gln Val Tyr Tyr Leu Gly Asn Val Pro Val Ala Lys	102
	<span style="margin-left: 10px;">10</span> <span style="margin-left: 100px;">15</span> <span style="margin-left: 100px;">20</span>	
	cct gtt ggg gta gat gtg att aat ggg gcc ctc gag tca gtc ctg tcc Pro Val Gly Val Asp Val Ile Asn Gly Ala Leu Glu Ser Val Leu Ser	150
	<span style="margin-left: 10px;">25</span> <span style="margin-left: 100px;">30</span> <span style="margin-left: 100px;">35</span> <span style="margin-left: 100px;">40</span>	
	tcc agc agc cgt gaa caa tgg acc cca agt cat gtc agt gtg gcc cct Ser Ser Ser Arg Glu Gln Trp Thr Pro Ser His Val Ser Val Ala Pro	198
	<span style="margin-left: 10px;">45</span> <span style="margin-left: 100px;">50</span> <span style="margin-left: 100px;">55</span>	
	gct acc ctc acc atc ttg cac cag cag aca gag gca gtg ctg gga gag	246

Ala	Thr	Leu	Thr 60	Ile	Leu	His	Gln	Gln 65	Thr	Glu	Ala	Val	Leu 70	Gly	Glu		
tgt	cgg	gtg	cgt	ttc	ctc	tcc	ttc	ctg	gcc	gtg	ggc	aga	gat	gtc	cac		294
Cys	Arg	Val 75	Arg	Phe	Leu	Ser	Phe 80	Leu	Ala	Val	Gly	Arg 85	Asp	Val	His		
acg	ttt	gca	ttc	atc	atg	gct	gcc	ggc	cca	gcc	tcc	ttc	tgc	tgc	cac		342
Thr	Phe 90	Ala	Phe	Ile	Met	Ala 95	Ala	Gly	Pro	Ala	Ser 100	Phe	Cys	Cys	His		
atg	ttc	tgg	tgc	gag	ccc	aat	gct	gcc	agc	ctc	tca	gag	gct	gtg	cag		390
Met	Phe	Trp	Cys	Glu	Pro 110	Asn	Ala	Ala	Ser	Leu 115	Ser	Glu	Ala	Val	Gln 120		
gct	gcg	tgc	atg	ctt	cgc	tac	cag	aag	tgt	gcg	gcc	gca	gga	agc	tct		438
Ala	Ala	Cys	Met	Leu 125	Arg	Tyr	Gln	Lys	Cys 130	Ala	Ala	Ala	Gly	Ser 135	Ser		
ggt	tcc	atc	gag	aaa	aaa	atc	act	gga	tat	acc	acc	gtt	gat	ata	tcc		486
Gly	Ser	Ile	Glu 140	Lys	Lys	Ile	Thr	Gly 145	Tyr	Thr	Thr	Val	Asp 150	Ile	Ser		
caa	tgg	cat	cgt	aaa	gaa	cat	ttt	gag	gca	ttt	cag	tca	gtt	gct	caa		534
Gln	Trp	His 155	Arg	Lys	Glu	His	Phe 160	Glu	Ala	Phe	Gln	Ser 165	Val	Ala	Gln		
tgt	acc	tat	aac	cag	acc	gtt	cag	ctg	gat	att	acg	gcc	ttt	tta	aag		582
Cys	Thr 170	Tyr	Asn	Gln	Thr	Val 175	Gln	Leu	Asp	Ile	Thr 180	Ala	Phe	Leu	Lys		
acc	gta	aag	aaa	aat	aag	cac	aag	ttt	tat	ccg	gcc	ttt	att	cac	att		630
Thr	Val	Lys	Lys	Asn	Lys 190	His	Lys	Phe	Tyr	Pro 195	Ala	Phe	Ile	His	Ile 200		
ctt	gcc	cgc	ctg	atg	aat	gct	cat	ccg	gaa	ttt	cg	atg	gca	atg	aaa		678
Leu	Ala	Arg	Leu	Met 205	Asn	Ala	His	Pro	Glu 210	Phe	Arg	Met	Ala	Met 215	Lys		
gac	ggt	gag	ctg	gtg	ata	tgg	gat	agt	gtt	cac	cct	tgt	tac	acc	gtt		726
Asp	Gly	Glu	Leu 220	Val	Ile	Trp	Asp	Ser 225	Val	His	Pro	Cys	Tyr 230	Thr	Val		
ttc	cat	gag	caa	act	gaa	acg	ttt	tca	tcg	ctc	tgg	agt	gaa	tac	cac		774
Phe	His	Glu 235	Gln	Thr	Glu	Thr	Phe 240	Ser	Ser	Leu	Trp	Ser 245	Glu	Tyr	His		
gac	gat	ttc	cgg	cag	ttt	cta	cac	ata	tat	tcg	caa	gat	gtg	gcg	tgt		822
Asp	Asp 250	Phe	Arg	Gln	Phe	Leu 255	His	Ile	Tyr	Ser	Gln 260	Asp	Val	Ala	Cys		
tac	ggt	gaa	aac	ctg	gcc	tat	ttc	cct	aaa	ggg	ttt	att	gag	aat	atg		870
Tyr	Gly	Glu	Asn	Leu 270	Ala	Tyr	Phe	Pro	Lys	Gly 275	Phe	Ile	Glu	Asn	Met 280		
ttt	ttc	gtc	tca	gcc	aat	ccc	tgg	gtg	agt	ttc	acc	agt	ttt	gat	tta		918
Phe	Phe	Val	Ser	Ala 285	Asn	Pro	Trp	Val	Ser 290	Phe	Thr	Ser	Phe	Asp 295	Leu		
aac	gtg	gcc	aat	atg	gac	aac	ttc	ttc	gcc	ccc	gtt	ttc	acc	atg	ggc		966
Asn	Val	Ala	Asn 300	Met	Asp	Asn	Phe	Phe	Ala	Pro	Val	Phe	Thr 310	Met	Gly		
aaa	tat	tat	acg	caa	ggc	gac	aag	gtg	ctg	atg	ccg	ctg	gcg	att	cag		1014
Lys	Tyr	Tyr 315	Thr	Gln	Gly	Asp	Lys 320	Val	Leu	Met	Pro	Leu 325	Ala	Ile	Gln		
gtt	cat	cat	gcc	gtt	tgt	gat	ggc	ttc	cat	gtc	ggc	aga	atg	ctt	aat		1062
Val	His	His	Ala	Val	Cys	Asp	Gly	Phe	His	Val	Gly	Arg	Met	Leu	Asn		

330 335 15 340  
 gaa tta caa cag tac tgc gat gag tgg cag ggc ggg gcg taa 1104  
 Glu Leu Gln Gln Tyr Cys Asp Glu Trp Gln Gly Gly Ala  
 345 350 355

tgtacaagtg aagatc 1120

<210> 10

<211> 357

<212> PRT

<213> artificial sequence

<220>

<223> Vecteur TA2Cam

<400> 10

Met Ala Ala Pro Lys Asn Glu Leu Val Gln Lys Phe Gln Val Tyr Tyr  
 1 5 10 15

Leu Gly Asn Val Pro Val Ala Lys Pro Val Gly Val Asp Val Ile Asn  
 20 25 30

Gly Ala Leu Glu Ser Val Leu Ser Ser Ser Ser Arg Glu Gln Trp Thr  
 35 40 45

Pro Ser His Val Ser Val Ala Pro Ala Thr Leu Thr Ile Leu His Gln  
 50 55 60

Gln Thr Glu Ala Val Leu Gly Glu Cys Arg Val Arg Phe Leu Ser Phe  
 65 70 75 80

Leu Ala Val Gly Arg Asp Val His Thr Phe Ala Phe Ile Met Ala Ala  
 85 90 95

Gly Pro Ala Ser Phe Cys Cys His Met Phe Trp Cys Glu Pro Asn Ala  
 100 105 110

Ala Ser Leu Ser Glu Ala Val Gln Ala Ala Cys Met Leu Arg Tyr Gln  
 115 120 125

Lys Cys Ala Ala Ala Gly Ser Ser Gly Ser Ile Glu Lys Lys Ile Thr  
 130 135 140

Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu His Phe  
 145 150 155 160

Glu Ala Phe Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln  
 165 170 175

Leu Asp Ile Thr Ala Phe Leu Lys Thr Val<sup>16</sup> Lys Lys Asn Lys His Lys  
 180 185 190  
 Phe Tyr Pro Ala Phe Ile His Ile Leu Ala Arg Leu Met Asn Ala His  
 195 200 205  
 Pro Glu Phe Arg Met Ala Met Lys Asp Gly Glu Leu Val Ile Trp Asp  
 210 215 220  
 Ser Val His Pro Cys Tyr Thr Val Phe His Glu Gln Thr Glu Thr Phe  
 225 230 235  
 Ser Ser Leu Trp Ser Glu Tyr His Asp Asp Phe Arg Gln Phe Leu His  
 245 250 255  
 Ile Tyr Ser Gln Asp Val Ala Cys Tyr Gly Glu Asn Leu Ala Tyr Phe  
 260 265 270  
 Pro Lys Gly Phe Ile Glu Asn Met Phe Phe Val Ser Ala Asn Pro Trp  
 275 280 285  
 Val Ser Phe Thr Ser Phe Asp Leu Asn Val Ala Asn Met Asp Asn Phe  
 290 295 300  
 Phe Ala Pro Val Phe Thr Met Gly Lys Tyr Tyr Thr Gln Gly Asp Lys  
 305 310 315  
 Val Leu Met Pro Leu Ala Ile Gln Val His His Ala Val Cys Asp Gly  
 325 330 335  
 Phe His Val Gly Arg Met Leu Asn Glu Leu Gln Gln Tyr Cys Asp Glu  
 340 345 350  
 Trp Gln Gly Gly Ala  
 355

<210> 11

<211> 1120

<212> DNA

<213> artificial sequence

<220>

<223> vecteur TB2Cam

<220>

<221> CDS

<222> (31)..(1104)

<223>



18

Asp Asp Phe Arg Gln Phe Leu His Ile Tyr Ser Gln Asp Val Ala Cys  
 250 255 260

tac ggt gaa aac ctg gcc tat ttc cct aaa ggg ttt att gag aat atg 870  
 Tyr Gly Glu Asn Leu Ala Tyr Phe Pro Lys Gly Phe Ile Glu Asn Met 280  
 265 270

ttt ttc gtc tca gcc aat ccc tgg gtg agt ttc acc agt ttt gat tta 918  
 Phe Phe Val Ser Ala Asn Pro Trp Val Ser Phe Thr Ser Phe Asp Leu 295  
 285

aac gtg gcc aat atg gac aac ttc ttc gcc ccc gtt ttc acc atg ggc 966  
 Asn Val Ala Asn Met Asp Asn Phe Phe Ala Pro Val Phe Thr Met Gly 310  
 300 305

aaa tat tat acg caa ggc gac aag gtg ctg atg ccg ctg gcg att cag 1014  
 Lys Tyr Tyr Thr Gln Gly Asp Lys Val Leu Met Pro Leu Ala Ile Gln 325  
 315

gtt cat cat gcc gtt tgt gat ggc ttc cat gtc ggc aga atg ctt aat 1062  
 Val His His Ala Val Cys Asp Gly Phe His Val Gly Arg Met Leu Asn 340  
 330 335

gaa tta caa cag tac tgc gat gag tgg cag ggc ggg gcg taa 1104  
 Glu Leu Gln Gln Tyr Cys Asp Glu Trp Gln Gly Gly Ala 355  
 345

tgtacaagtg aagatc 1120

<210> 12

<211> 357

<212> PRT

<213> artificial sequence

<220>

<223> vecteur TB2Cam

<400> 12

Met Ala Ala Pro Lys Asn Glu Leu Val Gln Lys Phe Gln Val Tyr Tyr  
 1 5 10 15

Leu Gly Asn Val Pro Val Ala Lys Pro Val Gly Val Asp Val Ile Asn  
 20 25 30

Gly Ala Leu Glu Ser Val Leu Ser Ser Ser Ser Arg Glu Gln Trp Thr  
 35 40 45

Pro Ser His Val Ser Val Ala Pro Ala Thr Leu Thr Ile Leu His Gln  
 50 55 60

Gln Thr Glu Ala Val Leu Gly Glu Cys Arg Val Arg Ser Leu Ser Phe  
 65 70 75 80

Leu Ala Val Gly Arg Asp Val His Thr Phe Ala Phe Ile Met Ala Ala  
 85 90 95

19

Gly Pro Ala Ser Phe Cys Cys His Met Phe Trp Cys Glu Pro Asn Ala  
 100 105 110

Ala Ser Leu Ser Glu Ala Val Gln Ala Ala Cys Met Leu Arg Tyr Gln  
 115 120 125

Lys Cys Ala Ala Ala Gly Ser Ser Gly Ser Ile Glu Lys Lys Ile Thr  
 130 135 140

Gly Tyr Thr Thr Val Asp Ile Ser Gln Trp His Arg Lys Glu His Phe  
 145 150 155 160

Glu Ala Phe Gln Ser Val Ala Gln Cys Thr Tyr Asn Gln Thr Val Gln  
 165 170 175

Leu Asp Ile Thr Ala Phe Leu Lys Thr Val Lys Lys Asn Lys His Lys  
 180 185 190

Phe Tyr Pro Ala Phe Ile His Ile Leu Ala Arg Leu Met Asn Ala His  
 195 200 205

Pro Glu Phe Arg Met Ala Met Lys Asp Gly Glu Leu Val Ile Trp Asp  
 210 215 220

Ser Val His Pro Cys Tyr Thr Val Phe His Glu Gln Thr Glu Thr Phe  
 225 230 235 240

Ser Ser Leu Trp Ser Glu Tyr His Asp Asp Phe Arg Gln Phe Leu His  
 245 250 255

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

Pro Lys Gly Phe Ile Glu Asn Met Phe Phe Val Ser Ala Asn Pro Trp  
 275 280 285

Val Ser Phe Thr Ser Phe Asp Leu Asn Val Ala Asn Met Asp Asn Phe  
 290 295 300

Phe Ala Pro Val Phe Thr Met Gly Lys Tyr Tyr Thr Gln Gly Asp Lys  
 305 310 315 320

Val Leu Met Pro Leu Ala Ile Gln Val His His Ala Val Cys Asp Gly  
 325 330 335

Phe His Val Gly Arg Met Leu Asn Glu Leu Gln Gln Tyr Cys Asp Glu  
 340 345 350

Trp Gln Gly Gly Ala  
 355

<210> 13

<211> 795

<212> DNA

<213> *Klebsiella pneumoniae*

<220>

<221> CDS

<222> (1)..(795)

<223>

<400> 13

atg att gaa caa gat gga ttg cac gca ggt tct ccg gcc gct tgg gtg	48
Met Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val	
1 5 10 15	
gag agg cta ttc ggc tat gac tgg gca caa cag aca atc ggc tgc tct	96
Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser	
20 25 30	
gat gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg gtt ctt ttt	144
Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe	
35 40 45	
gtc aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag gac gag gca	192
Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala	
50 55 60	
gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc gca gct gtg	240
Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val	
65 70 75 80	
ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta ttg ggc gaa	288
Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu	
85 90 95	
gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct gcc gag aaa	336
Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys	
100 105 110	
gta tcc atc atg gct gat gca atg cgg cgg ctg cat acg ctt gat ccg	384
Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro	
115 120 125	
gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc gag cga gca	432
Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala	
130 135 140	
cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat ctg gac gaa	480
Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu	
145 150 155 160	
gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg ctc aag gcg	528
Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala	
165 170 175	
cgc atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc gat gcc tgc	576
Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys	
180 185 190	

21

ttg	ccg	aat	atc	atg	gtg	gaa	aat	ggc	cgc	ttt	tct	gga	ttc	atc	gac	624
Leu	Pro	Asn	Ile	Met	Val	Glu	Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	
		195					200					205				
tgt	ggc	cgg	ctg	ggt	gtg	gcg	gac	cgc	tat	cag	gac	ata	gcg	ttg	gct	672
Cys	Gly	Arg	Leu	Gly	Val	Ala	Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	
	210					215					220					
acc	cgt	gat	att	gct	gaa	gag	ctt	ggc	ggc	gaa	tgg	gct	gac	cgc	ttc	720
Thr	Arg	Asp	Ile	Ala	Glu	Glu	Leu	Gly	Gly	Glu	Trp	Ala	Asp	Arg	Phe	
	225				230					235					240	
ctc	gtg	ctt	tac	ggt	atc	gcc	gct	ccc	gat	tcg	cag	cgc	atc	gcc	ttc	768
Leu	Val	Leu	Tyr	Gly	Ile	Ala	Ala	Pro	Asp	Ser	Gln	Arg	Ile	Ala	Phe	
				245					250					255		
tat	cgc	ctt	ctt	gac	gag	ttc	ttc	tga								795
Tyr	Arg	Leu	Leu	Asp	Glu	Phe	Phe									
			260													

<210> 14

<211> 264

<212> PRT

<213> *Klebsiella pneumoniae*

<400> 14

Met	Ile	Glu	Gln	Asp	Gly	Leu	His	Ala	Gly	Ser	Pro	Ala	Ala	Trp	Val
1			5						10					15	
Glu	Arg	Leu	Phe	Gly	Tyr	Asp	Trp	Ala	Gln	Gln	Thr	Ile	Gly	Cys	Ser
			20					25					30		
Asp	Ala	Ala	Val	Phe	Arg	Leu	Ser	Ala	Gln	Gly	Arg	Pro	Val	Leu	Phe
		35					40					45			
Val	Lys	Thr	Asp	Leu	Ser	Gly	Ala	Leu	Asn	Glu	Leu	Gln	Asp	Glu	Ala
	50					55					60				
Ala	Arg	Leu	Ser	Trp	Leu	Ala	Thr	Thr	Gly	Val	Pro	Cys	Ala	Ala	Val
	65				70					75					80
Leu	Asp	Val	Val	Thr	Glu	Ala	Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu
				85					90					95	
Val	Pro	Gly	Gln	Asp	Leu	Leu	Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys
			100					105					110		
Val	Ser	Ile	Met	Ala	Asp	Ala	Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro
		115					120					125			
Ala	Thr	Cys	Pro	Phe	Asp	His	Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala
	130					135					140				

Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu  
 145 150 22 155 160  
 Glu His Gln Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg Leu Lys Ala  
 165 170  
 Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys  
 180 185 190  
 Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp  
 195 200 205  
 Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala  
 210 215 220  
 Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe  
 225 230 235 240  
 Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe  
 245 250 255  
 Tyr Arg Leu Leu Asp Glu Phe Phe  
 260

<210> 15

<211> 795

<212> DNA

<213> artificial sequence

<220>

<223> kanamycine kinase II Q155W

<220>

<221> CDS

<222> (1)..(795)

<223>

<400> 15

atg att gaa caa gat gga ttg cac gca ggt tct ccg gcc gct tgg gtg	48
Met Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val	
1 5 10 15	
gag agg cta ttc ggc tat gac tgg gca caa cag aca atc ggc tgc tct	96
Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser	
20 25 30	
gat gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg gtt ctt ttt	144
Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe	
35 40 45	

															23	
gtc	aag	acc	gac	ctg	tcc	ggt	gcc	ctg	aat	gaa	ctg	cag	gac	gag	gca	192
Val	Lys	Thr	Asp	Leu	Ser	Gly	Ala	Leu	Asn	Glu	Leu	Gln	Asp	Glu	Ala	
	50					55				60						
gcg	cgg	cta	tcg	tgg	ctg	gcc	acg	acg	ggc	gtt	cct	tgc	gca	gct	gtg	240
Ala	Arg	Leu	Ser	Trp	Leu	Ala	Thr	Thr	Gly	Val	Pro	Cys	Ala	Ala	Val	
65				70					75						80	
ctc	gac	gtt	gtc	act	gaa	gcg	gga	agg	gac	tgg	ctg	cta	ttg	ggc	gaa	288
Leu	Asp	Val	Val	Thr	Glu	Ala	Gly	Arg	Asp	Trp	Leu	Leu	Leu	Gly	Glu	
				85					90					95		
gtg	ccg	ggg	cag	gat	ctc	ctg	tca	tct	cac	ctt	gct	cct	gcc	gag	aaa	336
Val	Pro	Gly	Gln	Asp	Leu	Leu	Ser	Ser	His	Leu	Ala	Pro	Ala	Glu	Lys	
			100					105					110			
gta	tcc	atc	atg	gct	gat	gca	atg	cgg	cgg	ctg	cat	acg	ctt	gat	ccg	384
Val	Ser	Ile	Met	Ala	Asp	Ala	Met	Arg	Arg	Leu	His	Thr	Leu	Asp	Pro	
		115					120					125				
gct	acc	tgc	cca	ttc	gac	cac	caa	gcg	aaa	cat	cgc	atc	gag	cga	gca	432
Ala	Thr	Cys	Pro	Phe	Asp	His	Gln	Ala	Lys	His	Arg	Ile	Glu	Arg	Ala	
	130					135					140					
cgt	act	cgg	atg	gaa	gcc	ggt	ctt	gtc	gat	tgg	gat	gat	ctg	gac	gaa	480
Arg	Thr	Arg	Met	Glu	Ala	Gly	Leu	Val	Asp	Trp	Asp	Asp	Leu	Asp	Glu	
145				150					155						160	
gag	cat	cag	ggg	ctc	gcg	cca	gcc	gaa	ctg	ttc	gcc	agg	ctc	aag	gcg	528
Glu	His	Gln	Gly	Leu	Ala	Pro	Ala	Glu	Leu	Phe	Ala	Arg	Leu	Lys	Ala	
				165					170					175		
cgc	atg	ccc	gac	ggc	gag	gat	ctc	gtc	gtg	acc	cat	ggc	gat	gcc	tgc	576
Arg	Met	Pro	Asp	Gly	Glu	Asp	Leu	Val	Val	Thr	His	Gly	Asp	Ala	Cys	
			180					185					190			
ttg	ccg	aat	atc	atg	gtg	gaa	aat	ggc	cgc	ttt	tct	gga	ttc	atc	gac	624
Leu	Pro	Asn	Ile	Met	Val	Glu	Asn	Gly	Arg	Phe	Ser	Gly	Phe	Ile	Asp	
		195					200					205				
tgt	ggc	cgg	ctg	ggt	gtg	gcg	gac	cgc	tat	cag	gac	ata	gcg	ttg	gct	672
Cys	Gly	Arg	Leu	Gly	Val	Ala	Asp	Arg	Tyr	Gln	Asp	Ile	Ala	Leu	Ala	
	210					215					220					
acc	cgt	gat	att	gct	gaa	gag	ctt	ggc	ggc	gaa	tgg	gct	gac	cgc	ttc	720
Thr	Arg	Asp	Ile	Ala	Glu	Glu	Leu	Gly	Gly	Glu	Trp	Ala	Asp	Arg	Phe	
				225		230				235					240	
ctc	gtg	ctt	tac	ggt	atc	gcc	gct	ccc	gat	tcg	cag	cgc	atc	gcc	ttc	768
Leu	Val	Leu	Tyr	Gly	Ile	Ala	Ala	Pro	Asp	Ser	Gln	Arg	Ile	Ala	Phe	
				245					250					255		
tat	cgc	ctt	ctt	gac	gag	ttc	ttc	tga								795
Tyr	Arg	Leu	Leu	Asp	Glu	Phe	Phe									
			260													

<210> 16

<211> 264

<212> PRT

<213> artificial sequence

<220>

24

&lt;223&gt; kanamycine kinase II Q155W

&lt;400&gt; 16

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

Tyr Arg Leu Leu Asp Glu Phe Phe

260

- <210> 17
- <211> 795
- <212> DNA
- <213> artificial sequence

- <220>
- <223> kanamycine kinase II S114H
- <220>
- <221> CDS
- <222> (1)..(795)
- <223>

```

<400> 17
atg att gaa caa gat gga ttg cac gca ggt tct ccg gcc gct tgg gtg      48
Met Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala Ala Trp Val
1           5           10           15

gag agg cta ttc ggc tat gac tgg gca caa cag aca atc ggc tgc tct      96
Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile Gly Cys Ser
           20           25           30

gat gcc gcc gtg ttc cgg ctg tca gcg cag ggg cgc ccg gtt ctt ttt      144
Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro Val Leu Phe
           35           40           45

gtc aag acc gac ctg tcc ggt gcc ctg aat gaa ctg cag gac gag gca      192
Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln Asp Glu Ala
           50           55           60

gcg cgg cta tcg tgg ctg gcc acg acg ggc gtt cct tgc gca gct gtg      240
Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys Ala Ala Val
           65           70           75           80

ctc gac gtt gtc act gaa gcg gga agg gac tgg ctg cta ttg ggc gaa      288
Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu Leu Gly Glu
           85           90           95

gtg ccg ggg cag gat ctc ctg tca tct cac ctt gct cct gcc gag aaa      336
Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro Ala Glu Lys
           100           105           110

gta cac atc atg gct gat gca atg cgg cgg ctg cat acg ctt gat ccg      384
Val His Ile Met Ala Asp Ala Met Arg Arg Leu His Thr Leu Asp Pro
           115           120           125

gct acc tgc cca ttc gac cac caa gcg aaa cat cgc atc gag cga gca      432
Ala Thr Cys Pro Phe Asp His Gln Ala Lys His Arg Ile Glu Arg Ala
           130           135           140

cgt act cgg atg gaa gcc ggt ctt gtc gat cag gat gat ctg gac gaa      480
Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gln Asp Asp Leu Asp Glu
           145           150           155           160

gag cat cag ggg ctc gcg cca gcc gaa ctg ttc gcc agg ctc aag gcg      528

```

Glu His Gln Gly Leu Ala Pro Ala Glu <sup>26</sup> Leu Phe Ala Arg Leu Lys Ala  
 165 170 175  
 cgc atg ccc gac ggc gag gat ctc gtc gtg acc cat ggc gat gcc tgc 576  
 Arg Met Pro Asp Gly Glu Asp Leu Val Val Thr His Gly Asp Ala Cys  
 180 185 190  
 ttg ccg aat atc atg gtg gaa aat ggc cgc ttt tct gga ttc atc gac 624  
 Leu Pro Asn Ile Met Val Glu Asn Gly Arg Phe Ser Gly Phe Ile Asp  
 195 200 205  
 tgt ggc cgg ctg ggt gtg gcg gac cgc tat cag gac ata gcg ttg gct 672  
 Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile Ala Leu Ala  
 210 215 220  
 acc cgt gat att gct gaa gag ctt ggc ggc gaa tgg gct gac cgc ttc 720  
 Thr Arg Asp Ile Ala Glu Glu Leu Gly Gly Glu Trp Ala Asp Arg Phe  
 225 230 235 240  
 ctc gtg ctt tac ggt atc gcc gct ccc gat tcg cag cgc atc gcc ttc 768  
 Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg Ile Ala Phe  
 245 250 255  
 tat cgc ctt ctt gac gag ttc ttc tga 795  
 Tyr Arg Leu Leu Asp Glu Phe Phe  
 260

<210> 18

<211> 264

<212> PRT

<213> artificial sequence

<220>

<223> kanamycine kinase II S114H

<400> 18

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



<220>

<223> amorce

<400> 20

gaatgagttg gtcgagaagt tccaagtc

28

<210> 21

<211> 29

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 21

tattacctgg gggtagtacc tgttgctaa

29

<210> 22

<211> 27

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 22

tagatgtgat tgatggggcc ctcgagt

27

<210> 23

<211> 26

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 23

agcagccgtg aagaatggac cccaag

26

<210> 24

<211> 29

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 24

accatcttgc acgagcagac agaggcagt

29

<210> 25

<211> 29

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 25

accatcttgc accaggagac agaggcagt

29

<210> 26

<211> 27

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 26

ttctggtgcg agcccgatgc tgccagc

27

<210> 27

<211> 26

<212> DNA

<213> artificial sequence

<220>

<223> amorce

<400> 27

tcagaggctg tggaggctgc gtgcat

26

<210> 28

30

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;223&gt; amorce

&lt;400&gt; 28

atgcttcgct acgagaagtg tgcgg

25

&lt;210&gt; 29

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;223&gt; amorce

&lt;400&gt; 29

tcagaggctg tggaggctgc gtgcat

26

&lt;210&gt; 30

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;223&gt; amorce

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (16)..(17)

&lt;223&gt; n = a t g or c

&lt;400&gt; 30

agtgtcgggt gcgtnnctc tccttcctgg

30

&lt;210&gt; 31

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

31

&lt;220&gt;

&lt;223&gt; amorce

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (16)..(17)

&lt;223&gt; n = a c t o r g

<400> 31  
atgtcagtgt ggcccngct accctcacca

30

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/FR2006/000257

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C12N15/62 C07K14/00 C12N15/10 G01N33/50		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) C07K C12N C12Q G01N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, BIOSIS, CHEM ABS Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUSSMAN J K ET AL: "VECTORS FOR CONSTRUCTING KAN GENE FUSIONS DIRECT SELECTION OF MUTATIONS AFFECTING IS10 GENE EXPRESSION" GENE (AMSTERDAM), vol. 90, no. 1, 1990, pages 135-140, XP002357616 ISSN: 0378-1119 the whole document	14, 15
Y	US 2003/138843 A1 (WALDO GEOFFREY S) 24 July 2003 (2003-07-24) cited in the application abstract; claims 4,5,10,16,17 ----- -/--	1, 2, 4-13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.		
<input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents :		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed		
*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family		
Date of the actual completion of the international search  31 May 2006		Date of mailing of the international search report  21/06/2006
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  Espen, J

## INTERNATIONAL SEARCH REPORT

International application No

PCT/FR2006/000257

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KROMER WOLFGANG J ET AL: "Expression of the membrane protein glycophorin A as a fusion with the antibiotic resistance protein neomycin phosphotransferase II" BIOTECHNOLOGY AND BIOENGINEERING, vol. 57, no. 2, 20 January 1998 (1998-01-20), pages 238-244, XP002383206 ISSN: 0006-3592 the whole document	1,2,4-13
Y	REISS B ET AL: "PROTEIN FUSIONS WITH THE KANAMYCIN RESISTANCE GENE FROM TRANSPOSON TN-5" EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 3, no. 13, 1984, pages 3317-3322, XP009067205 ISSN: 0261-4189 the whole document	1,2,4-13
Y	NAKAYAMA M ET AL: "A system using convertible vectors for screening soluble recombinant proteins produced in Escherichia coli from randomly fragmented cDNAs" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 312, no. 3, 19 December 2003 (2003-12-19), pages 825-830, XP004474921 ISSN: 0006-291X the whole document	1,2,4-13
Y	GB 2 400 102 A (* SENSE PROTEOMIC LIMITED) 6 October 2004 (2004-10-06) page 7 - page 12; figure 1	1,2,4-13
Y	MAXWELL ET AL: "A simple in vivo assay for increased protein solubility" PROTEIN SCIENCE, CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE, GB, vol. 8, September 1999 (1999-09), pages 1908-1911, XP002146236 ISSN: 0961-8368 the whole document	1,2,4-13
	----- -/--	

## INTERNATIONAL SEARCH REPORT

International application No

PCT/FR2006/000257

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>FANG J ET AL: "Studies on interaction between hTNF-alpha and its two receptors with expressed hsTR55-preS1/hsTR75-preS1 fusion soluble receptors" JOURNAL OF IMMUNOLOGICAL METHODS, ELSEVIER SCIENCE PUBLISHERS B.V., AMSTERDAM, NL, vol. 274, no. 1-2, 1 March 2003 (2003-03-01), pages 199-207, XP004411945 ISSN: 0022-1759 abstract</p> <p style="text-align: center;">-----</p>	1, 2, 4-13

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/FR2006/000257

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2003138843	A1	24-07-2003	NONE
GB 2400102	A	06-10-2004	NONE

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/FR2006/000257

**A. CLASSEMENT DE L'OBJET DE LA DEMANDE**

INV. C12N15/62 C07K14/00 C12N15/10 G01N33/50

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

**B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE**

Documentation minimale consultée (système de classification suivi des symboles de classement)

C07K C12N C12Q G01N

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

EPO-Internal, BIOSIS, CHEM ABS Data

**C. DOCUMENTS CONSIDERES COMME PERTINENTS**

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	SUSSMAN J K ET AL: "VECTORS FOR CONSTRUCTING KAN GENE FUSIONS DIRECT SELECTION OF MUTATIONS AFFECTING IS10 GENE EXPRESSION" GENE (AMSTERDAM), vol. 90, no. 1, 1990, pages 135-140, XP002357616 ISSN: 0378-1119 le document en entier	14, 15
Y	US 2003/138843 A1 (WALDO GEOFFREY S) 24 juillet 2003 (2003-07-24) cité dans la demande abrégé; revendications 4,5,10,16,17 ----- -/--	1,2,4-13

 Voir la suite du cadre C pour la fin de la liste des documents

 Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

\*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

\*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

\*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

\*&amp;\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

31 mai 2006

Date d'expédition du présent rapport de recherche internationale

21/06/2006

Nom et adresse postale de l'administration chargée de la recherche internationale

 Office Européen des Brevets, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Espen, J

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°  
PCT/FR2006/000257

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	<p>KROMER WOLFGANG J ET AL: "Expression of the membrane protein glycophorin A as a fusion with the antibiotic resistance protein neomycin phosphotransferase II" BIOTECHNOLOGY AND BIOENGINEERING, vol. 57, no. 2, 20 janvier 1998 (1998-01-20), pages 238-244, XP002383206 ISSN: 0006-3592 le document en entier</p>	1,2,4-13
Y	<p>REISS B ET AL: "PROTEIN FUSIONS WITH THE KANAMYCIN RESISTANCE GENE FROM TRANSPOSON TN-5" EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 3, no. 13, 1984, pages 3317-3322, XP009067205 ISSN: 0261-4189 le document en entier</p>	1,2,4-13
Y	<p>NAKAYAMA M ET AL: "A system using convertible vectors for screening soluble recombinant proteins produced in Escherichia coli from randomly fragmented cDNAs" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 312, no. 3, 19 décembre 2003 (2003-12-19), pages 825-830, XP004474921 ISSN: 0006-291X le document en entier</p>	1,2,4-13
Y	<p>GB 2 400 102 A (* SENSE PROTEOMIC LIMITED) 6 octobre 2004 (2004-10-06) page 7 - page 12; figure 1</p>	1,2,4-13
Y	<p>MAXWELL ET AL: "A simple in vivo assay for increased protein solubility" PROTEIN SCIENCE, CAMBRIDGE UNIVERSITY PRESS, CAMBRIDGE, GB, vol. 8, septembre 1999 (1999-09), pages 1908-1911, XP002146236 ISSN: 0961-8368 le document en entier</p>	1,2,4-13
	-/--	

**RAPPORT DE RECHERCHE INTERNATIONALE**

Demande internationale n°  
PCT/FR2006/000257

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
Y	<p>FANG J ET AL: "Studies on interaction between hTNF-alpha and its two receptors with expressed hsTR55-preS1/hsTR75-preS1 fusion soluble receptors"                      JOURNAL OF IMMUNOLOGICAL METHODS, ELSEVIER                      SCIENCE PUBLISHERS B.V.,AMSTERDAM, NL,                      vol. 274, no. 1-2,                      1 mars 2003 (2003-03-01), pages 199-207,                      XP004411945                      ISSN: 0022-1759                      abrégé</p> <p style="text-align: center;">-----</p>	1,2,4-13

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°

PCT/FR2006/000257

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2003138843	A1	24-07-2003	AUCUN
GB 2400102	A	06-10-2004	AUCUN