



(86) Date de dépôt PCT/PCT Filing Date: 2005/06/07  
(87) Date publication PCT/PCT Publication Date: 2006/01/12  
(45) Date de délivrance/Issue Date: 2013/06/04  
(85) Entrée phase nationale/National Entry: 2006/12/01  
(86) N° demande PCT/PCT Application No.: FR 2005/001392  
(87) N° publication PCT/PCT Publication No.: 2006/003298  
(30) Priorité/Priority: 2004/06/07 (FR0406120)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01),  
*C12N 15/00* (2006.01), *G06F 19/22* (2011.01),  
*C40B 30/02* (2006.01)

(72) Inventeurs/Inventors:  
CHODORGE, MATTHIEU, FR;  
FOURAGE, LAURENT, FR;  
LEFEVRE, FABRICE, FR;  
MASSON, JEAN-MICHEL, FR

(73) Propriétaire/Owner:  
PROTEUS, FR

(74) Agent: SMART & BIGGAR

(54) Titre : PROCEDE DE DETERMINATION DE LA CHARGE MUTATIONNELLE D'UNE BANQUE DE GENES  
OBTENUE PAR MUTAGENESE ALEATOIRE D'UN GENE D'INTERET ET LES MOYENS POUR SA MISE EN  
OEUVRE

(54) Title: METHOD OF DETERMINING THE MUTATIONAL LOAD OF A GENE LIBRARY OBTAINED BY RANDOM  
MUTAGENESIS OF A PARTICULAR GENE AND MEANS FOR IMPLEMENTING SAME

(57) **Abrégé/Abstract:**

La présente invention se rapporte à un procédé de détermination de la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt comprenant : a) la préparation d'un abaque liant la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction des gènes modèle mutés observée dans ladite banque ; b) la mutagenèse aléatoire du gène modèle ayant servi pour la préparation de l'abaque de l'étape (a) et du gène d'intérêt pour obtenir les banques de gènes mutés correspondantes ; c) la détermination de la charge mutationnelle (CM) de la banque de gènes obtenue à partir du gène modèle à l'étape (b) à partir de l'abaque établi à l'étape (a) ; d) l'application d'un facteur de correction (CF) à la charge mutationnelle (CM) de la banque de gènes modèle mutés déterminée à l'étape (c) pour déterminer la charge mutationnelle (CM) de la banque de gènes d'intérêt mutés de l'étape (b).

(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 janvier 2006 (12.01.2006)

PCT

(10) Numéro de publication internationale  
**WO 2006/003298 A3**(51) Classification internationale des brevets :  
*C12Q 1/68* (2006.01) *G06F 1/00* (2006.01)(21) Numéro de la demande internationale :  
PCT/FR2005/001392

(22) Date de dépôt international : 7 juin 2005 (07.06.2005)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :  
0406120 7 juin 2004 (07.06.2004) FR(71) Déposant (*pour tous les États désignés sauf US*) : **PRO-  
TEUS** [FR/FR]; Parc Scientifique Georges Besse, 70, allée  
Graham Bell, F-30000 Nîmes (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (*pour US seulement*) :  
**CHODORGE, Matthieu** [FR/FR]; 7 rue Germaine,  
F-92160 Antony (FR). **FOURAGE, Laurent** [FR/FR];  
195 rue du 19 Mars 1962, F-30870 Clarensac (FR).  
**LEFEVRE, Fabrice** [FR/FR]; Au village, F-32120  
Bajonnette (FR). **MASSON, Jean-Michel** [FR/FR]; 48,  
chemin Flou de Rious, F-31400 Toulouse (FR).(74) Mandataire : **BREESE DERAMBURE MAJEROW-  
ICZ**; 38, avenue de l'Opéra, F-75002 Paris (FR).(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, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, 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.(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, 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).

Publiée :

- avec rapport de recherche internationale
- avant l'expiration du délai prévu pour la modification des  
revendications, sera republiée si des modifications sont re-  
çues

(88) Date de publication du rapport de recherche  
internationale: 26 mai 2006*En ce qui concerne les codes à deux lettres et autres abrégia-  
tions, se référer aux "Notes explicatives relatives aux codes et  
abréviations" figurant au début de chaque numéro ordinaire de  
la Gazette du PCT.*(54) Title: METHOD OF DETERMINING THE MUTATIONAL LOAD OF A GENE LIBRARY OBTAINED BY RANDOM  
MUTAGENESIS OF A PARTICULAR GENE AND MEANS FOR IMPLEMENTING SAME(54) Titre : PROCEDE DE DETERMINATION DE LA CHARGE MUTATIONNELLE D'UNE BANQUE DE GENES OBTENUE  
PAR MUTAGENESE ALEATOIRE D'UN GENE D'INTERET ET LES MOYENS POUR SA MISE EN OEUVRE

(57) Abstract: The invention relates to a method of determining the mutational load of a gene library obtained by random mutagenesis of a particular gene of interest and means for implementing same, comprising the following steps: a) preparation of an abacus that links the mutational load (CM) of a gene library obtained by random mutagenesis of a model gene to the fraction of mutated model genes observed in said library; b) random mutagenesis of the model gene that was used to prepare the abacus in step (a) and of the gene of interest, in order to produce the corresponding mutated gene libraries; c) determination of the mutational load (CM) of the gene library obtained from the model gene in step (b) using the abacus established in step (a); and d) application of a correction factor (CF) to the mutational load (CM) of the mutated model gene library determined in step (c), in order to determine the mutational load (CM) of the library of mutated genes of interest from step (b).

(57) Abrégé : La présente invention se rapporte à un procédé de détermination de la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt comprenant : a) la préparation d'un abaque liant la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction des gènes modèle mutés observée dans ladite banque ; b) la mutagenèse aléatoire du gène modèle ayant servi pour la préparation de l'abaque de l'étape (a) et du gène d'intérêt pour obtenir les banques de gènes mutés correspondantes ; c) la détermination de la charge mutationnelle (CM) de la banque de gènes obtenue à partir du gène modèle à l'étape (b) à partir de l'abaque établi à l'étape (a) ; d) l'application d'un facteur de correction (CF) à la charge mutationnelle (CM) de la banque de gènes modèle mutés déterminée à l'étape (c) pour déterminer la charge mutationnelle (CM) de la banque de gènes d'intérêt mutés de l'étape (b).

WO 2006/003298 A3

PROCEDE DE DETERMINATION DE LA CHARGE  
MUTATIONNELLE D'UNE BANQUE DE GENES OBTENUE PAR  
MUTAGENESE ALEATOIRE D'UN GENE D'INTERET ET LES MOYENS  
POUR SA MISE EN OEUVRE

La présente invention se rapporte à un procédé permettant de déterminer la charge mutationnelle introduite dans une banque de gènes à l'issue d'une étape de mutagenèse aléatoire d'un gène d'intérêt. L'invention se rapporte également aux moyens pour la mise en oeuvre de ce procédé comme des kits et des supports destinés à être exploités par des ordinateurs.

La mutagenèse aléatoire des acides nucléiques permet d'introduire des modifications dans les séquences d'acides aminés des protéines. Il s'agit d'une stratégie alternative performante à la mutagenèse dirigée et au design rationnel des protéines. En effet, en l'absence de toutes données précises sur les caractéristiques d'une protéine comme sa structure, c'est à dire dans la majorité des cas, seule la mutagenèse aléatoire permet d'identifier rapidement des mutants ou des variants d'un gène ayant acquis de nouvelles propriétés.

De manière générale, bien qu'il ait été possible d'isoler des variants améliorés dans des banques hyper-mutées (Zaccolo M. et al. (1999) Journal of Molecular Biology. 285 :775-783), il est admis que l'introduction d'un trop grand nombre de mutations conduit à une majorité de protéines inactives et limite sévèrement l'identification de variants améliorés alors qu'un taux de mutation trop faible diminue la probabilité d'obtenir une mutation bénéfique. Ainsi

plusieurs exemples d'évolution dirigée tendraient à montrer que le taux de mutation le plus efficace pour identifier des variants améliorés serait de 1-5 acides aminés par protéine, soit 2-7 nucléotides par gène  
5 (Cherry J .R. et al. (1999) Nature Biotechnology. 17 :379-384. / Shafikhani S. et al. (1997) BioTechniquers 23 :304-310. / Moore J.C. et al. (1997) Journal of Molecular Biology. 272 :336-347.)

Il est donc nécessaire de pouvoir contrôler  
10 le nombre de mutations introduites dans une banque de gènes par mutagenèse aléatoire. La charge mutationnelle (CM) est définie comme le nombre moyen de mutations par gène au sein d'une banque de gènes mutés.

La méthode de référence de l'art antérieur  
15 pour déterminer précisément le taux de mutation dans une banque de gènes mutés reste aujourd'hui l'observation statistique par séquençage d'une fraction de gènes de la banque. Cette méthode reste encore lourde à mettre en oeuvre et surtout est très coûteuse.

Comme alternative au séquençage, d'autres  
20 méthodes pour quantifier le taux de mutation ont été développées comme par exemple le « Single Strand Conformation Polymorphism » (SSCP) ou le « Denaturing Gradient Gel Electrophoresis » (DGGE) qui permettent de  
25 détecter des différences entre les séquences jusqu'à une base (Cariello et al. (1993) Mutational Research 288 :103-112 / Brail et al, (1993) Mutational Research 303 :171-175). Néanmoins ces techniques sont plutôt adaptées à une analyse qualitative comme la détection  
30 de mutation et la détermination de faibles taux d'erreur qu'à une quantification de charge mutationnelle.

Une autre méthode de l'art antérieur fait appel à l'analyse des changements dans le profil de

restriction de séquences ayant intégrées ou perdues des sites de restriction particuliers. Cette méthode s'applique donc uniquement à des séquences particulières avec une modification de sites de restriction représentatifs.

Il a aussi été proposé une méthode de suivi de la fréquence de mutation basée sur des tests phénotypiques où un gène rapporteur est soumis à une mutagenèse aléatoire avant clonage. Les fréquences de mutations sont alors simplement analysées en enregistrant la perte de fonctionnalité du produit du gène rapporteur due à l'apparition de mutations dans celui-ci. Cette dernière méthode est cependant limitée à des gènes pour lesquels l'introduction de mutations peut être observée facilement par un test phénotypique. Il est également décrit dans l'art antérieur des tables permettant d'associer à des conditions expérimentales une charge mutationnelle d'une banque de gènes mutés pour un gène d'intérêt donné en fonction de sa longueur. Ces tables sont construites sur la base d'un gène modèle soumis à différentes conditions expérimentales puis par le séquençage d'un nombre statistiquement représentatif de gènes de la banque de gènes mutés. Cependant, ces techniques ne tiennent compte que de la longueur dudit gène d'intérêt.

De plus, aucune des méthodes de l'art antérieur n'utilisent de standard interne permettant de tenir compte des variabilités expérimentales susceptibles d'influencer la charge mutationnelle d'une banque. Ce qui signifie que ces méthodes ne sont pas suffisamment fiables et reproductibles pour qu'il soit possible de s'affranchir du séquençage afin de déterminer la charge mutationnelle obtenue dans une banque de gènes mutés.

La présente invention a précisément pour but d'offrir dans le domaine de la mutagenèse aléatoire des méthodes fiables pour la prédiction de la charge mutationnelle de banques de gènes mutés.

Ce but est atteint selon l'invention grâce à un procédé de détermination de la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt, caractérisé en ce qu'il comprend les étapes suivantes :

a) la préparation d'un abaque liant la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction des gènes modèles mutés observés dans ladite banque ;

b) la mutagenèse aléatoire du gène modèle ayant servi pour la préparation de l'abaque de l'étape (a) et du gène d'intérêt pour obtenir les banques de gènes mutés correspondantes ;

c) la détermination de la charge mutationnelle (CM) de la banque de gènes obtenue à partir du gène modèle à l'étape (b) à partir de l'abaque établi à l'étape (a) ;

d) l'application d'un facteur de correction (CF) à la charge mutationnelle (CM) de la banque de gènes modèles mutés déterminée à l'étape (c) pour déterminer la charge mutationnelle (CM) de la banque de gènes d'intérêt mutés de l'étape (b).

Selon une forme toute préférée du procédé de l'invention, la technique de mutagenèse mise en oeuvre à l'étape (b) est identique pour le gène modèle et pour le gène d'intérêt et est également similaire à celle de l'étape (a) pour construire l'abaque à partir du gène modèle. Avantageusement, on entend par

technique identique ou similaire, le fait que non seulement la technique est la même mais aussi les conditions de mise en oeuvre. La similarité ou l'identité de la technique de mutagenèse et de ses conditions n'excluent pas le fait que l'homme de métier puisse y apporter des aménagements qui ne soient pas de nature à modifier substantiellement les performances de la technique de mutagenèse. A titre d'exemple, les polymérases utilisées aux étapes (a) et (b) pour mettre en oeuvre une technique de PCR error-prone peuvent ne pas être identiques.

De même, l'homme de métier est à même de déterminer un gène modèle pour l'étape b) différent de celui de l'étape a), mais présentant des propriétés substantiellement équivalentes pour la mise en oeuvre desdites étapes.

Le procédé de l'invention est remarquable en ce qu'il offre une précision suffisante pour constituer une alternative efficace au séquençage qui est très coûteux. Ainsi les étapes (c) et (d) peuvent s'affranchir du séquençage.

Dans la suite de la description de l'invention qui suit, on utilisera indifféremment :

« banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle » ou « banque de gènes modèles mutés » ; et

« banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt » ou « banque de gènes d'intérêt mutés »

85387-79

5a

L'invention a pour objet un procédé de détermination de la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt, comprenant les étapes suivantes :

a) préparation d'un abaque liant la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction des gènes modèles mutés observée dans ladite banque, comprenant les étapes suivantes :

1) sélection du gène modèle,

2) sélection de la méthode de mutagenèse aléatoire, et

10 3) obtention dudit abaque ;

b) mutagenèse aléatoire du gène modèle ayant servi pour la préparation de l'abaque de l'étape (a) et du gène d'intérêt pour obtenir les banques de gènes mutés correspondantes ;

c) détermination de la charge mutationnelle (CM) de la banque de gènes obtenue à partir du gène modèle à l'étape (b) à partir de l'abaque établi à l'étape (a) ; et

d) application d'un facteur de correction (CF) à la charge mutationnelle (CM) de la banque de gènes modèles mutés déterminée à l'étape (c) pour déterminer la charge mutationnelle (CM) de la  
20 banque de gènes d'intérêt mutés de l'étape (b), où ledit facteur de correction (C.F.) se décompose en plusieurs sous-facteurs définis suivant des paramètres sélectionnés comme influençant la charge mutationnelle d'une banque de gènes d'intérêt mutés par rapport à une banque de gènes modèles mutés

où les étapes (c) et (d) ne comprennent pas de séquençage.

On entend par gène au sens de la présente invention, toute séquence d'acide nucléotidique, codante ou non.

5 Une caractéristique essentielle de la présente invention réside dans le fait que les modifications de la séquence du gène modèle peuvent être observées après mutagenèse.

10 On entend par gènes modèles mutés les gènes modèles dont la séquence nucléotidique est modifiée par rapport à celle du gène modèle avant mutagenèse.

On entend par gène modèle muté observé un gène modèle dont la modification de séquence est observable.

15 On entend par observable le fait que l'introduction aléatoire de mutation n'est pas analysée par séquençage mais par l'analyse directe ou indirecte de la modification d'une ou plusieurs propriétés des gènes modèles mutés par rapport au gène modèle avant  
20 mutagenèse.

On entend par analyse indirecte le fait que la modification d'une ou plusieurs propriétés du gène muté est observée par une réaction intermédiaire.

25 A titre d'exemple la modification d'un gène modèle peut être observée par la modification d'une ou plusieurs propriétés de la protéine codée par le gène modèle. La modification d'une ou plusieurs propriétés  
30 de la protéine peut consister en la perte ou le gain d'activité(s) caractérisant la protéine.

Dans une forme préférée de réalisation de l'invention, l'observation des gènes modèles mutés est réalisée par un test phénotypique. On entend par test

phénotypique, un test permettant de détecter la fonction de la protéine codée par le gène modèle, muté ou non.

5 A titre d'exemple de gène dont la fonction peut être détectée par un test phénotypique, on peut citer un gène :

- impliqué dans la résistances aux antibiotiques,
- impliqué dans l'auxotrophie de la souche
- 10 hôte,
- codant pour une protéine présentant une activité aisément détectable (beta-galactosidase / xylanase / Green Fluorescent Protein (GFP...),
- codant pour une protéine impliquée dans
- 15 une réaction immunologique de type antigène-anticorps.

Le gène modèle est utilisé dans la méthode de l'invention pour :

- établir un abaque tel que défini dans
- 20 l'étape (a), et
- être utilisé en parallèle du gène d'intérêt à l'étape (b) pour tenir compte des variabilités que peuvent avoir les conditions expérimentales sur les résultats de mutagenèse
- 25 aléatoire, et ainsi servir de standard interne.

Contrairement à un témoin classiquement utilisé en biologie, le standard interne défini dans le procédé de l'invention par le gène modèle, n'indique pas uniquement le bon déroulement ou le degré

30 d'avancement d'une expérience, il donne en plus une mesure nécessaire à la réalisation du procédé.

Dans une forme particulière de réalisation de l'invention, le gène modèle est un gène modèle dit

universel, c'est à dire, qu'il possède des caractéristiques partagées par un grand nombre de gènes. A titre d'exemple, il présente une longueur comprise entre 500pb et 10kpb et un pourcentage en GC compris entre 20 et 80. Préférentiellement, le gène modèle universel présente un pourcentage en GC compris entre 40 et 60. A titre d'exemple d'un gène modèle universel, on peut citer le gène codant pour la protéine GFP, dont l'introduction de mutations peut être réalisée par un test phénotypique selon lequel la fluorescence des colonies sous ultraviolet reflète un phénotype GFP actif.

Dans une autre forme préférée de réalisation de l'invention, le gène modèle présente des caractéristiques proches de celles du gène d'intérêt. Ces caractéristiques correspondent avantageusement à des propriétés intrinsèques. Il est entendu par propriété intrinsèque d'un gène des propriétés inhérentes à la séquence de ce gène.

Parmi les propriétés intrinsèques d'un gène envisagées ci-dessus, on peut citer à titre d'exemple et de manière non limitative :

- sa longueur,
- sa composition en base, la composition en base peut être avantageusement analysée sur un seul brin.
- sa fréquence de bases répétées dans la séquence,
- la présence de séquences symétriques dans le gène.

Une fois le gène modèle sélectionné, une technique de mutagenèse aléatoire est choisie pour

mettre en oeuvre la méthode de l'invention. Il est  
entendu par technique de mutagenèse aléatoire toute  
technique connue de l'homme de métier. Elle peut  
correspondre également à une combinaison de plusieurs  
5 techniques de mutagenèse. Elle peut être *in vivo* ou *in vitro*.

Dans une forme de mise en oeuvre préférée,  
la méthode de l'invention se rapporte à des techniques  
de mutagenèse *in vitro* basée préférentiellement sur  
10 l'utilisation d'une polymérase ou sur la mise en oeuvre  
d'une technique d'amplification. A titre d'exemple, la  
technique de mutagenèse correspond à la PCR error-prone  
(Cadwell R.C et *al.* (1992) PCR Methods Appl. 2 :28-33.  
/ Leung D.W. et *al.* (1989) Techniques 1 :11-15.)

15 Une fois le gène modèle sélectionné et la  
technique de mutagenèse choisie, un abaque permettant  
de lier la charge mutationnelle d'une banque de gènes  
obtenue par mutagenèse aléatoire d'un gène modèle à la  
fraction de gènes modèles mutés observés dans ladite  
20 banque est préparé (étape a).

Il est entendu par abaque au sens de  
l'invention tout graphique, toute table ou toute  
équation permettant de lier de manière biunivoque la  
25 charge mutationnelle (CM) d'une banque de gènes obtenue  
par mutagenèse aléatoire d'un gène modèle à la fraction  
de gènes modèles mutés observés dans ladite banque.

Un abaque est obtenu pour une technique de  
mutagenèse et un gène modèle donné.

30 Ainsi, dans la forme préférée de  
l'invention où la modification de la séquence du gène  
modèle est observable par un test phénotypique,  
l'abaque obtenu à l'étape (a) permet de lier la charge  
mutationnelle d'une banque de gènes obtenue par

mutagenèse aléatoire du gène modèle à la fraction de gènes présentant un phénotype mutant dans ladite banque.

5 L'abaque peut avoir été réalisé au préalable par l'homme de l'art ou peut être obtenu de la manière suivante :

i) Une technique de mutagenèse donnée est appliquée à un gène modèle préalablement sélectionné pour obtenir une banque de gènes modèles mutés.

10 ii) La charge mutationnelle (CM) dans la banque de gènes modèles mutés est ensuite déterminée par séquençage. Il est observé en parallèle la fraction de gènes modèles mutés dans ladite banque présentant une modification d'une ou plusieurs propriétés par rapport au gène sauvage.

15

Dans une forme préférée de l'invention, un nombre de clones statistiquement représentatif est analysé et le gène modèle muté correspondant est séquencé. Dans une forme préférée de l'invention, entre 20 10 et 100 clones sont séquencés. Avantageusement, le séquençage est effectué sur 25 clones tandis que l'analyse de la fraction de clones présentant une modification de séquence est effectuée sur au moins 500 clones.

25

Une liaison biunivoque peut ainsi être établie entre la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle et la fraction de gènes modèles mutés observés dans ladite banque.

30

Disposant de cet abaque, la technique de mutagenèse aléatoire utilisée pour construire l'abaque est appliquée en parallèle au gène modèle et à un gène d'intérêt à l'étape (b), le gène modèle de l'étape (b)

étant de préférence le même que celui ayant servi à construire l'abaque à l'étape (a).

A l'étape (b), la même technique de mutagenèse est appliquée simultanément au gène modèle et au gène d'intérêt.

Le fait d'utiliser le gène modèle avec le gène d'intérêt à l'étape (b) revient à utiliser le gène modèle comme standard interne. Le fait d'utiliser le gène modèle comme standard interne présente l'avantage :

- de tenir compte des variations expérimentales qui peuvent avoir lieu lors d'une expérience effectuée à l'étape (b) sur le gène modèle par rapport à celles effectuées antérieurement pour obtenir la relation définie à l'étape (a) ;

- de minimiser les variations expérimentales d'un échantillon à un autre, c'est à dire du traitement du gène modèle par rapport au gène d'intérêt lors de l'étape (b).

Ces variations expérimentales sont non négligeables sur la charge mutationnelle observée. En effet, il a été démontré que des modifications même minimales des conditions expérimentales peuvent avoir des conséquences importantes sur la charge mutationnelle introduite dans une banque. Le tableau 1 ci-après présente les moyennes, les écart-types et les coefficients de variations C.V. (C.V.=écart-type/moyenne en %) des fractions de clones exprimant une GFP fonctionnelle à l'issue de six expériences répétées de PCR error-prone effectuées avec différentes concentrations de  $MgCl_2$  et  $MnCl_2$  (paramètres de contrôle du niveau de mutagenèse). A partir d'une expérience répétée à l'identique, le résultat de la

mutagenèse error-prone est fortement variable avec des coefficients de variations sur les fractions de clones exprimant une GFP fonctionnelle, et donc du taux de mutation compris entre 10 et 40%.

5

Tableau 1

| Conditions réactionelles |                         | Fraction population clones <i>gfp</i> active |            |        |
|--------------------------|-------------------------|----------------------------------------------|------------|--------|
| [MgCl <sub>2</sub> ] mM  | [MnCl <sub>2</sub> ] mM | Moyenne                                      | Ecart Type | CV (%) |
| 5                        | 0                       | 0,722                                        | 0,071      | 9,9    |
| 7                        | 0,2                     | 0,411                                        | 0,122      | 29,7   |
| 7                        | 0,4                     | 0,091                                        | 0,040      | 43,8   |
| 7                        | 0,5                     | 0,036                                        | 0,013      | 35,5   |

Le tableau 1 ci-dessus indique la variabilité de la fraction de clones exprimant une GFP fonctionnelle dans des banques de gènes *gfp* mutés aléatoirement par PCR error-prone lors d'expériences répétées 6 fois à l'identiques.

L'étape (b) est donc l'étape clé de la méthode de l'invention. Elle permet de tenir compte des variations expérimentales pouvant avoir lieu lors de la mise en oeuvre d'une technique de mutagenèse aléatoire sur un gène d'intérêt, lesdites variations pouvant influencer la définition de la charge mutationnelle de la banque de gènes ainsi obtenue. De ce fait, la méthode de l'invention permet de s'affranchir du séquençage au niveau des étapes (c) et (d) pour déterminer la charge mutationnelle d'une banque de gènes d'intérêt mutés.

Dans une forme toute préférée de l'invention, le même mélange réactionnel est utilisé pour mettre en oeuvre en parallèle la technique de mutagenèse sur le gène modèle et sur le gène d'intérêt pour minimiser encore les variations expérimentales d'un échantillon à un autre.

La banque de gènes modèles mutés est ensuite analysée pour déterminer la charge mutationnelle de la banque correspondante (étape c). A l'issue de l'étape (b), la fraction de gènes modèles mutés observés dans la banque obtenue après mutagenèse du gène modèle est déterminée. Grâce à l'abaque défini à l'étape (a), la fraction de gènes modèles mutés observés dans ladite banque permet de définir la charge mutationnelle de cette banque.

A l'étape (d), la charge mutationnelle déterminée à l'étape (c) est corrigée en lui appliquant un facteur de correction (CF) pour déterminer la charge mutationnelle introduite dans la banque de gènes obtenue par mutagenèse aléatoire du gène d'intérêt.

Ceci est réalisé selon l'équation suivante :

$$CM (\text{gène d'intérêt}) = CM (\text{gène modèle}) \times CF$$

dans laquelle, CF représente le facteur de correction, et CM représente la charge mutationnelle.

Le facteur de correction (CF) peut se décomposer en plusieurs sous-facteurs de correction suivant les paramètres sélectionnés comme influençant la charge mutationnelle d'une banque de gènes d'intérêt mutés par rapport à une banque de gènes modèle mutés.

Un sous-facteur de correction permet de normaliser l'effet d'un paramètre sélectionné comme capable d'influencer la charge mutationnelle d'une banque par rapport à une autre.

La construction du facteur de correction CF s'effectue par la multiplication des différents sous-facteurs que l'on souhaite prendre en compte pour la détermination de la charge mutationnelle de la banque

de gènes d'intérêt mutés dans la méthode de l'invention.

Les paramètres sélectionnés comme pouvant influencer la charge mutationnelle peuvent correspondre à titre d'exemple à un ou plusieurs des paramètres suivants :

- paramètres intrinsèques
- paramètres expérimentaux.

Il est entendu par paramètres intrinsèques d'un gène ceux qui découlent de la séquence de ce gène comme indiqué précédemment.

Il est entendu par paramètre expérimental tout paramètre capable d'engendrer des différences dans l'introduction de mutations dans le gène modèle par rapport au gène d'intérêt lors de la mise en oeuvre de la technique de mutagenèse.

La nature de ces paramètres expérimentaux sont dépendants de la technique de mutagenèse aléatoire utilisée. Ils ne peuvent être calculés qu'après avoir mis en oeuvre la technique de mutagenèse sur le gène modèle et sur le gène d'intérêt à l'étape (b).

La considération de ce type de paramètre pour effectuer la correction à l'étape (d) s'ajoute à l'avantage d'utiliser un standard interne afin de tenir compte au mieux des variations expérimentales sur la détermination de la charge mutationnelle d'une banque de gènes.

30

Lors de techniques de mutagenèse basées sur l'usage de la PCR (dont la PCR error-prone), les paramètres expérimentaux peuvent correspondre, de manière non limitative :

- au nombre de multiplication du gène durant l'étape de mutagenèse aléatoire,
- à la quantité de matrice initiale,
- au nombre de cycle d'amplification,
- 5       - à la nature et à la quantité de polymérase utilisée
- aux conditions expérimentales utilisées comme la concentration en cations divalents, la concentration et le ratio de
- 10       concentration des dNTPs.

La définition des différents sous-facteurs de correction doit s'effectuer suivant des lois mathématiques prenant en compte la nature des évènements biologiques qui seront normalisés.

Par exemple, il est convenu dans l'art antérieur que le nombre de mutations introduites dans un gène est lié linéairement à la longueur du gène (paramètre intrinsèque).

Ainsi  $CF[L]$ , défini comme le sous-facteur de correction lié à la longueur des gènes, est explicité comme suit :

$$CF[L] = L(\text{gène d'intérêt}) / L(\text{gène modèle})$$

où L correspond à la longueur des gènes.

De même, il est connu de l'art antérieur que les différentes méthodes de mutagenèse n'affectent pas de manière équivalente les différents nucléotides A, T, G ou C. Pour chaque méthode de mutagenèse on peut définir un sous-facteur de correction  $CF[\text{base}]$  qui tient compte des différences de composition en bases entre le gène modèle et le gène d'intérêt vis à vis du biais d'introduction de mutation de la technique de mutagenèse.

Pour établir un tel sous-facteur de correction il est nécessaire d'introduire la notion de matrice cible des mutations.

En effet, la matrice cible des mutations du gène modèle ( $M_{\text{target}}[X]$  où  $X$  représente le gène modèle) décomposent sur les quatre bases la probabilité de toucher un type de base particulier dans le gène modèle et permet de normaliser la composition en base par rapport au gène modèle utilisé.

10

La matrice cible des mutations du gène modèle est déduite à partir de séquençages des gènes modèles mutés par une technique définie de mutagenèse aléatoire, et identification des mutations apparues.

15

L'analyse des mutations apparues dans les séquences des gènes modèles mutés permet d'établir la matrice des mutations  $Mut_x$ , où  $Mut_x$  répertorie les fréquences des 12 types possibles de mutations, selon la matrice suivante :

20

$$Mut_x = \begin{pmatrix} & f_{A \rightarrow T} & f_{A \rightarrow C} & f_{A \rightarrow G} \\ f_{T \rightarrow A} & & f_{T \rightarrow C} & f_{T \rightarrow G} \\ f_{C \rightarrow A} & f_{C \rightarrow T} & & f_{C \rightarrow G} \\ f_{G \rightarrow A} & f_{G \rightarrow T} & f_{G \rightarrow C} & \end{pmatrix}$$

dans laquelle,  $f_{x \rightarrow y}$  représente la probabilité de muter la base  $X$  en base  $Y$ .

25

La matrice cible de mutation du gène modèle  $X$ ,  $M_{\text{target}}[X]$  est alors déduite de la matrice  $Mut_x$  en sommant les trois fréquences affectant un type de base selon la matrice suivante :

$$M_{\text{target}}[X] = \begin{pmatrix} A_{\text{Mtarget}[X]} \\ T_{\text{Mtarget}[X]} \\ C_{\text{Mtarget}[X]} \\ G_{\text{Mtarget}[X]} \end{pmatrix}$$

dans laquelle,  $Z_{\text{Mtarget}[X]} = \sum_{Y=(A,T,C,G)-Z} f_{Z \rightarrow Y}$ , où Z est A, T, C ou G.

5 La matrice cible des mutations est dépendante de la nature du gène modèle. Plus un gène modèle sera riche dans un type de base particulier plus la probabilité d'affecter ce type de base avec une méthode définie de mutagenèse aléatoire sera élevée.

10 La matrice cible des mutations est dépendante de la technique de mutagenèse utilisée car cette dernière est fonction du biais de la méthode de mutagenèse aléatoire.

15 Ainsi CF[base], défini comme le sous-facteur de correction lié à la composition en base, est explicité selon l'équation suivante :

$$\text{C.F.}[\text{base}] = \left( \frac{A_{\text{Mtarget}[X]} \times A_{(\text{g.interêt})}}{A_{(X)}} + \frac{T_{\text{Mtarget}[X]} \times T_{(\text{g.interêt})}}{T_{(X)}} + \frac{C_{\text{Mtarget}[X]} \times C_{(\text{g.interêt})}}{C_{(X)}} + \frac{G_{\text{Mtarget}[X]} \times G_{(\text{g.interêt})}}{G_{(X)}} \right)$$

20 dans laquelle,  $A_{(\text{g.interêt})}$ ,  $A_{(X)}$ ,  $T_{(\text{g.interêt})}$ ,  $T_{(X)}$ ,  $C_{(\text{g.interêt})}$ ,  $C_{(X)}$ ,  $G_{(\text{g.interêt})}$  et  $G_{(X)}$  correspondent respectivement à la composition en base A, T, C et G du gène modèle X et du gène d'intérêt.

25 De plus, l'impact de certains mutagènes sur un nucléotide donné peut varier significativement selon le type de nucléotide qui lui est adjacent dans la séquence. C'est pourquoi le facteur correctif pourra intégrer, lorsque cela sera pertinent, un terme de comparaison entre les fréquences des dinucléotides dans le gène d'intérêt par rapport au gène modèle.

La fréquence relative des trinuécléotides entre gène modèle et gène d'intérêt pourra de même être prise en compte dans le facteur correctif.

5                   A titre d'exemple également, dans le cas d'une mutagenèse «*in vitro*» basée sur la PCR, le taux de multiplication (paramètre expérimental) subit par la matrice va directement influencer le taux de mutation. En effet, l'efficacité de réplication est dépendante de  
10 la matrice ADN et du programme PCR utilisé. Dues à des différences de longueur, de composition en base, de la présence de structure secondaire, certains gènes sont plus difficiles à amplifier que d'autres. Ainsi, les éventuelles différences de rendement PCR entre la  
15 mutagenèse du gène modèle et celle du gène d'intérêt pourront générer des différences dans le taux de multiplication et ceci même en partant d'une quantité identique de matrice et en réalisant un même nombre de cycle PCR. Le taux de multiplication (d) est calculé  
20 d'après l'équation suivante :

$$d = \ln(Q_f/Q_i)/\ln 2$$

dans laquelle,  $Q_i$  et  $Q_f$  représentent respectivement les quantités molaires de matrice initiale et de produit à l'issue de la PCR.

25                   Les quantités initiale de matrice et finale de produit peuvent être déterminées par de nombreux moyens connus de l'homme de métier. Citons par exemple le dosage spectrophotométrique aux UV ou le calibrage  
30 par rapport à une quantité connue d'ADN.

Les paramètres expérimentaux peuvent également correspondre au nombre de générations d'une culture cellulaire lors d'une technique de mutagenèse

« *in vivo* ». A titre d'exemple, dans le cas d'une mutagenèse «*in vivo*», le nombre de mutations introduites dans les gènes va dépendre du nombre de générations de la culture. L'expérimentateur doit alors  
5 prendre soin de l'analyser, par exemple en mesurant la turbidité des cultures.

Différentes formes de réalisation de l'invention peuvent être envisagées, où :

10 - au moins un paramètre intrinsèque de chacun des gènes modèle et d'intérêt est considéré pour calculer le facteur de correction,

- au moins deux paramètres intrinsèques du gène modèle et du gène d'intérêt sont considérés pour  
15 calculer le facteur de correction selon l'équation suivante :

$$CF = CF[\text{paramètre intrinsèque 1}] \times CF[\text{paramètre intrinsèque 2}]$$

A titre d'exemple, ces paramètres  
20 intrinsèques correspondent à la longueur et à la composition en base des gènes, selon l'équation suivante :

$$CF = CF[L] \times CF[\text{base}]$$

dans laquelle :

25 -  $CF[L]$  , le sous-facteur de correction est lié à la longueur des gènes,

-  $CF[\text{base}]$ , le sous-facteur de correction est lié à la composition en base des gènes,

- au moins un paramètre intrinsèque et au  
30 moins un paramètre expérimental sont considérés pour établir le facteur de correction, avec C.F. déterminé selon l'équation suivante :

$$CF = CF[\text{paramètre intrinsèque}] \times CF[\text{paramètre expérimental}]$$

5                   - au moins deux paramètres intrinsèques et au moins un paramètre expérimental sont considérés pour établir le facteur de correction, avec CF déterminé selon l'équation suivante :

$$CF = CF[\text{paramètre intrinsèque 1}] \times CF[\text{paramètre intrinsèque 2}] \times CF[\text{paramètre expérimental}]$$

10                   Dans le cas particulier où la technique de mutagenèse correspond à la PCR error-prone et le gène modèle correspond à la GFP (séquence de la Figure 1), différents sous-facteurs de correction correspondent pour :

15                   - paramètre intrinsèque 1 représente la longueur des gènes selon l'équation suivante :

$$CF[L] = L(\text{gène d'intérêt})/717$$

où L est en nombre de paire de base,

20                   - paramètre intrinsèque 2 représente la composition en base sur les brins d'ADN selon l'équation suivante :

$$C.F.[\text{base}] = \left( \frac{0,506 \times A_{(g.\text{intérêt})}}{0,34} + \frac{0,346 \times T_{(g.\text{intérêt})}}{0,26} + \frac{0,063 \times C_{(g.\text{intérêt})}}{0,197} + \frac{0,085 \times G_{(g.\text{intérêt})}}{0,203} \right)$$

où  $A_{(g.\text{intérêt})}$ ,  $T_{(g.\text{intérêt})}$ ,  $C_{(g.\text{intérêt})}$  et  $G_{(g.\text{intérêt})}$  sont en fraction,

25                   - paramètre expérimental représente le taux de multiplication selon l'équation suivante :

$$CF[d] = d(\text{gène d'intérêt}) / d(\text{gfp})$$

où d est le taux de multiplication calculé selon l'équation suivante :

30                   
$$d = \ln(Qf/Qi)/\ln 2$$

dans laquelle, Qi et Qf représentent respectivement les quantités molaires de la matrice initiale et des produits à l'issue de la PCR.

L'invention concerne aussi les utilisations du procédé préalablement défini :

1) Le procédé de l'invention peut être  
5 utilisé pour obtenir et contrôler la charge  
mutationnelle désirée dans une banque de gènes  
d'intérêt mutés. L'expérimentateur, sur la base d'un  
abaque lui permettant de lier la charge mutationnelle  
d'une banque de gènes modèles mutés aux conditions  
10 expérimentales de mutagenèse aléatoire, va sélectionner  
les conditions expérimentales ou une gamme de  
conditions expérimentales lui permettant de se  
rapprocher d'une charge mutationnelle désirée pour sa  
banque obtenue après mutagenèse aléatoire d'un gène  
15 d'intérêt. L'expérimentateur va faire cette sélection  
en tenant compte au moins d'un paramètre intrinsèque  
des gènes modèles et d'intérêt. L'expérimentateur met  
alors en œuvre les étapes (a) à (d) du procédé de  
l'invention préalablement décrite en utilisant à  
20 l'étape (b) les conditions expérimentales préalablement  
sélectionnées. Il utilise alors la méthode de  
l'invention pour vérifier que la charge mutationnelle  
introduite est bien celle désirée. Ainsi, l'utilisation  
de la méthode de l'invention pour obtenir et contrôler  
25 une charge mutationnelle désirée dans une banque de  
gènes d'intérêt mutés est caractérisée par la mise en  
œuvre d'une étape supplémentaire de sélection des  
conditions expérimentales de mutagenèse aléatoire qui  
seront utilisées pour mettre en œuvre l'étape (b). A  
30 titre d'exemple cette utilisation permet de contrôler  
la charge mutationnelle dans une banque de gènes  
d'intérêt mutés en préalable à une étude structure-  
fonction. L'expérimentateur réalise premièrement une  
banque de gènes d'intérêt faiblement muté puis il

vérifie, par la méthode de l'invention, que la charge mutationnelle introduite est bien de 1-2 mutations par gène, le nombre de mutation optimale pour ce type d'application (Vartanian J.P., Henry M., Wain-Hobson S. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Res. (1996). 24 :2627-2631). Le criblage des variants issus de la banque dans une fonction donnée et le séquençage des clones identifiés positifs permettent ensuite de lier sans ambiguïté la séquence à la fonction étudiée.

2) Le procédé de l'invention peut être utilisé pour contrôler la charge mutationnelle dans une banque de gènes d'intérêt mutés en préalable à une recherche de variants améliorés. L'expérimentateur réalise premièrement une banque de gène d'intérêt muté puis il vérifie, par la méthode de l'invention, que la charge mutationnelle introduite est bien de 3-7 mutations par gène, le nombre de mutations optimale pour ce type d'application (Wan L. et al. (1998) Proc. Natl. Acad. Sci. USA 95 :12825-12831. / You L et al. (1996) Protein Eng. 9 :77-83). L'application d'un nombre optimal de mutation permet d'augmenter la probabilité d'identifier des gènes codant pour des protéines améliorées.

3) Le procédé de l'invention peut être utilisé pour étudier la plasticité protéique c'est à dire la capacité à incorporer des mutations tout en conservant une activité présentée par la protéine sauvage (Shafikhani S. et al. (1997) BioTechniques 23 :304-310). L'expérimentateur réalise premièrement des banques de gènes d'intérêt plus ou moins mutés puis il détermine, par la méthode de l'invention, les charges mutationnelles associées aux différentes

banques. L'analyse des fractions de clones produisant un variant protéique fonctionnel dans les différentes banques permet l'étude de la plasticité du produit du gène d'intérêt en liant ces dernières aux charges mutationnelles précédemment déterminées.

4) Le procédé de l'invention peut être utilisé pour étudier le taux d'évolution d'une protéine d'intérêt, c'est à dire la capacité à acquérir une nouvelle propriété avec une charge mutationnelle donnée (Miura T. et *al.* (2001) J. theor. Biol. 209 :497-502). L'expérimentateur réalise premièrement des banques de gènes d'intérêt plus ou moins mutés puis il détermine, par la méthode de l'invention, les charges mutationnelles associées aux différentes banques. L'analyse des fractions de clones produisant un variant présentant une propriété originale par rapport à la protéine sauvage dans les différentes banques permet d'étudier le taux d'évolution en fonction de la charge mutationnelle.

5) L'invention peut être utilisée pour déterminer la charge mutationnelle optimale dans des banques de gènes d'intérêt mutés pour générer un lot de gènes parentaux avant la mise en oeuvre de méthodes de recombinaison moléculaire. L'expérimentateur réalise premièrement des banques de gènes d'intérêt plus ou moins mutés puis il détermine, par la méthode de l'invention, les charges mutationnelles associées aux différentes banques. Deuxièmement la réalisation de criblages adéquats sur les banques de gènes d'intérêt mutés permettent d'identifier des clones présentant une/des propriété(s) améliorée(s) par rapport à celles présentées par la protéine sauvage. Chacune des banques de gènes d'intérêt mutés génèrent ainsi un lot de gènes optimisés, qui servent alors de gènes parentaux dans

une étape de recombinaison. Les différentes banques de gènes recombinés sont criblées pour identifier des protéines encore plus améliorées que celles produites par les gènes parentaux ou par le gène sauvage.

5 L'analyse et la comparaison de ces banques, par exemple sur la fréquence d'apparition de clones améliorés ou sur la quantification des améliorations obtenues, permettent de déterminer la charge mutationnelle optimale à introduire initialement dans une banque de

10 gène d'intérêt en préambule à une étape de recombinaison. Toutes les méthodes de recombinaison *in vitro* : DNA Shuffling (Stemmer W. (1994) Nature 370 :389-391), L-Shuffling (WO0009679), StEP (Zhao H. et *al.* (1998) Nature Biotechnology 16 :258-261),

15 RACHiTT (Coco W. et *al.* (2001) Nature Biotechnology 19 :354-359) et *in vivo* comme la recombinaison dans la levure (Pompon D. et *al.* (1989) Gene 83:15-24), ainsi que la chimerisation / réparation *in vivo* (Volkov A. et *al.* (1999) Nucleic Acids Research 27 :e18) sont

20 concernées par l'invention.

6) Le procédé de l'invention peut être utilisé pour quantifier ou comparer la fidélité d'une (de) méthode(s) de polymérisation d'acides nucléiques. L'expérimentateur réalise une/des banque(s) de gènes

25 modèles à partir des produits issus de polymérisation(s) réalisée(s) dans des conditions expérimentales données, composition en tampon, type de polymerase etc...

Dans un deuxième temps il détermine, par la

30 méthode de l'invention, les charges mutationnelles associées aux banques de gène modèle. Ces dernières, après la prise en compte éventuelle du nombre de cycles de répllication, permettent d'estimer la fidélité de la

méthode de polymérisation, dans les conditions expérimentales utilisées.

La présente invention concerne également les kits pour la mise en oeuvre du procédé et de ses différentes applications ou utilisations.

Un tel kit comprend :

- au moins un gène modèle par exemple sous la forme d'un plasmide ;
- 10 - au moins un abaque permettant de lier la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire du gène modèle à la fraction de gènes modèles mutés observés dans ladite banque ;
- les réactifs nécessaires à la mise en oeuvre de la technique de mutagenèse aléatoire comme par exemple une ADN polymérase thermostable, dNTP, magnésium, manganèse, tampon de PCR dans le cas où la technique de mutagenèse correspond à la PCR error-prone ;
- 15 - les réactifs, si nécessaires, pour observer des gènes modèles mutants ;
- une notice indiquant les sous-facteurs de correction définis pour un paramètre sélectionné comme influençant la charge mutationnelle.

25 Il y a autant d'abaques que de gènes modèles dans ledit kit.

Dans une forme de mise en oeuvre où la technique de mutagenèse utilise une étape d'amplification, le kit peut posséder les amorces pour l'amplification du gène modèle.

30

Lorsque le kit est mis en œuvre pour obtenir et contrôler une charge mutationnelle désirée dans une banque de gènes mutés, il contient en outre au

moins un abaque lui permettant de lier la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèses aléatoire d'un gène modèle aux conditions expérimentales de mutagenèses aléatoire.

5

Les sous-facteurs de correction fournis seront complétés par une ou plusieurs des données suivantes, et de façon non limitative, telles que :

- la longueur du gène d'intérêt,
- 10       - la composition en base A du gène d'intérêt,
- la composition en base C du gène d'intérêt,
- la composition en base G du gène d'intérêt,
- 15       - la composition en base T du gène d'intérêt,
- le nombre de multiplication de la matrice du gène modèle
- 20       - le nombre de multiplication de la matrice du gène d'intérêt,
- dont les valeurs correspondantes sont intégrées par l'expérimentateur, en fonction du gène d'intérêt. Le choix de ces données se fera selon que
- 25       l'on souhaite tenir compte d'un ou plusieurs paramètres intrinsèques et/ou expérimentaux.

Dans le cas où le gène modèle correspond à la *gfp* et la technique de mutagenèse correspond à la PCR error-prone, les différents sous-facteurs de

30       correction sont ceux indiqués précédemment.

Dans une forme préférée, le kit possède plusieurs gènes modèles et plusieurs abaques pour

chaque gène modèle de manière à choisir le gène modèle qui ait les propriétés intrinsèques les plus proches du gène d'intérêt.

5 Selon une forme particulière de mise en oeuvre du kit de l'invention, les sous-facteurs de correction sont enregistrés dans un fichier numérique sauvegardé sur un support approprié (CD-Rom, disquette, etc.).

10 De même, dans un kit selon l'invention, l'abaque peut être enregistré sous forme d'une base de données numériques sauvegardées sur un support approprié.

15 L'invention se rapporte enfin à un procédé dans lequel l'une au moins des étapes (a), (c) et (d) est mise en oeuvre par des moyens informatiques.

Lesdits moyens informatiques peuvent :

- 20 - lier la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction de gènes modèles mutés observés dans ladite banque.
- 25 - être incrémentés par chaque nouvel abaque créé. Ce procédé intègre la mise en oeuvre d'une base de données de différents abaques, chacun défini pour une technique de mutagenèse et un gène modèle donné.
- 30 - posséder les données pour établir au moins un facteur de correction pour la détermination de la charge mutationnelle d'une banque de gènes d'intérêt mutés.

L'expérimentateur devra alors préciser, en fonction de ses travaux, des données numériques relatives à l'une au moins des caractéristiques suivantes, que lesdits moyens informatiques exploitent:

5                   - le gène modèle et ses propriétés intrinsèques

                  - la technique de mutagenèse

                  - le ou les paramètres sélectionnés comme capables d'influencer la charge mutationnelle d'une  
10 banque de gènes par rapport à une autre pour définir les sous-facteurs de correction.

Lesdits moyens informatiques peuvent être présentés sous la forme d'un logiciel.

15                   Ledit logiciel peut être utilisé sur tout type d'appareil multimédia, tel par exemple, une calculatrice préprogrammée, ou un micro-ordinateur, et comporte une interface simple permettant à l'expérimentateur de préciser les différentes données  
20 numériques exploitées par lesdits moyens informatiques.

Ledit logiciel permet aussi d'intégrer les données liées aux abaques du ou des gènes modèles, et aux sous-facteurs de correction.

25                   Lesdits abaques et sous-facteurs de correction peuvent être soit directement enregistrés par l'expérimentateur dans ledit logiciel, soit être enregistrés dans un fichier numérique sauvegardé sur un support approprié (CD-Rom, disquette, etc.), ledit support étant introduit dans l'appareil multimédia  
30 (ordinateur, calculatrice, etc.) par l'expérimentateur au moment souhaité.

A titre d'exemple, l'expérimentateur entrera une ou plusieurs des données déjà indiquées précédemment.

Ainsi, l'invention se rapporte aussi :

5                   - d'une part à un appareil multimédia comme une calculatrice préprogrammée ou un micro-ordinateur pour la mise du procédé ci-dessus, et

10                   - d'autre part à un dispositif susceptible d'être mis en oeuvre dans un procédé pour déterminer la charge mutationnelle d'une banque obtenue à l'issue de la mutagenèse aléatoire d'un gène d'intérêt comprenant un kit défini précédemment et cet appareil multimédia.

15                   D'autres avantages et caractéristiques de l'invention apparaîtront dans les exemples qui suivent, où il sera fait référence aux dessins en annexe dans lesquels :

La figure 1 représente la séquence du gène de la *gfp*.

20                   La figure 2 représente la relation entre la fraction de clones exprimant une GFP fonctionnelle et la charge mutationnelle dans la banque de gènes *gfp* mutés.

25                   La figure 3 représente la matrice des mutations  $Mut_{GFP}$  pour le gène modèle *gfp* et pour la technique de mutagenèse aléatoire par PCR error-prone.

La figure 4 représente la matrice cible de mutation  $M_{target}[gfp]$  sur le gène *gfp* pour la technique de mutagenèse aléatoire par PCR error-prone.

30                   La figure 5 représente la plasticité de TEM-1  $\beta$ -lactamase, soit le taux de survie des protéines mutées en fonction de la charge mutationnelle des banques de gènes *TEM-1* mutés.

La figure 6 représente l'évolution de TEM-1  $\beta$ -lactamase, soit la fréquence d'apparition de clones

résistant à la cefotaxime en fonction de la charge mutationnelle des banques de gènes *TEM-1* mutés.

5                   Exemple 1 : Détermination de la charge mutationnelle introduite dans une banque de gènes de xylanase de *Thermotoga maritima* (*xynA*) à l'issue d'une étape de mutagenèse aléatoire grâce à la mise en oeuvre des étapes suivantes :

10                   1) Sélection du gène modèle

Dans l'exemple développé, le gène de la Green Fluorescent Protein (*gfp*) a été choisi comme gène modèle grâce à la facile mise en oeuvre d'un test phénotypique basé sur l'expression du produit du gène *gfp* : les colonies exprimant une protéine correspondant à la « Green Fluorescent Protein » ou GFP fonctionnelle présentent une couleur verte aux ultraviolets.

20                   2) Sélection de la méthode de mutagenèse aléatoire

Du fait de leur facilité de mise en oeuvre, les mutagenèses aléatoires ont été menées par PCR error-prone comme décrit par Cadwell (Cadwell R.C., Joyce G.F. (1992) PCR Methods Appl. 2 :28-33).

25                   3) Obtention d'un abaque permettant de lier la charge mutationnelle d'une banque de gènes modèle *gfp* mutés aléatoirement à la fraction de gènes modèle mutés observés dans ladite banque

30                   Pour déterminer la correspondance entre la charge mutationnelle dans la banque de gènes modèle *gfp* mutés et la fraction de clones exprimant une GFP fonctionnelle, trois expériences de mutagenèse

aléatoire (A,B et C) par PCR error-prone ont été effectuées.

Les réactions de PCR error-prone ont été réalisées dans des volumes de 100 $\mu$ L contenant 10mM de Tris-HCl pH9, 50mM de KCl, 0.1% de Triton X-100, 0.2mg/mL de BSA, 20pmol d'amorces universelles pET5' et pET3', 2,5U de Taq polymerase (QBiogen). Les concentrations en cations divalents sont de 4mM MgCl<sub>2</sub> pour l'expérience A, 7mM MgCl<sub>2</sub> plus 0,3mM MnCl<sub>2</sub> pour l'expérience B et 7mM de MgCl<sub>2</sub> plus 0,5mM MnCl<sub>2</sub> pour l'expérience C. Les concentrations en dNTP utilisées sont 0,2mM de dATP, 0,2mM de dGTP, 1mM de dTTP et 1mM de dCTP tandis que 1fmol de vecteur a servi comme matrice ADN. Un programme PCR de 94°C pendant 5min; (91°C pendant 30s, 60°C pendant 30s, 72°C pendant 1min.) 25 fois puis 72°C pendant 10 min en guise d'étape d'élongation finale, a été appliqué dans un thermocycler MJ Research PTC-200.

Les produits PCR ont alors été purifiés en utilisant le kit de purification sur colonne QIAquick (Qiagen), vérifiés sur gel et les quantités d'ADN obtenues ont été quantifiées par le kit PicoGreen (Molecular Probes) comme décrit par le fournisseur.

Les produits de PCR error-prone (1 $\mu$ g) ont été digérés par Eco RI et Nde I (New England Biolabs) pendant 1h à 37°C dans le tampon de réaction adéquat comme décrit par le fournisseur puis purifiés sur colonne QIAquick et insérés dans le vecteur pET26b+ (Stratagene) préalablement digéré avec Eco RI et Nde I. Les cellules *E. coli* MC1061(DE3) ont été transformées par choc thermique comme décrit par Maniatis (Sambrook J. et al. (1989) Molecular cloning: a laboratory manual Cold Spring Harbor Laboratory Press -New York) avec 1 $\mu$ L des réactions de ligation. Les transformants ont été

sélectionnés sur boîtes LB agar complémentées avec 0,6mg/mL de kanamycine.

5 Après clonage et transformation, les trois banques sont analysées pour la fraction de clones exprimant une GFP fonctionnelle mais également par séquençage pour déterminer les charges mutationnelles dans les trois banques de gènes *gfp* mutés à l'issue des expériences de mutagenèse A, B et C.

10 Le criblage des clones exprimant une protéine GFP fonctionnelle a été réalisé en identifiant les colonies présentant une fluorescence verte sous banc UV (longueur d'onde d'excitation à 355nm). Parallèlement, le rendement d'insertion des gènes *gfp* mutés dans les différentes banques a été analysé par  
15 PCR sur des colonies choisies aléatoirement. Par banque, le criblage pour la recherche de clones présentant une GFP fonctionnelle, a été réalisé sur environ 1000 clones et le rendement d'insertion (nombre de clones présentant un insert par nombre total de  
20 clones) a été estimé à partir de 96 PCR sur colonies, ceci de façon à être statistiquement pertinent.

Les banques de gènes *gfp* mutés sont ainsi analysées sur le nombre total de colonie (*cfu\_total*), le nombre de colonies exprimant une GFP fonctionnelle  
25 (*cfu\_fluo*) et le rendement d'insertion des gènes *gfp* mutés dans les banques (*Ri*). La fraction de clones exprimant une GFP fonctionnelle dans une banque donnée (*Fa[GFP]*) est alors déduite de l'équation suivante:

$$Fa[GFP] = \frac{cfu\_fluo}{cfu\_total \times Ri}$$

30 Pour déterminer les charges mutationnelles dans les banques A, B ou C de gènes *gfp* mutés, 24 clones par banque contenant un insert *gfp* ont été

sélectionnés pour une amplification PCR avec les amorces pET5'/pET3'. Les produits de digestion ont été purifiés sur colonne QIAquick puis séquencés en totalité. Toutes les mutations (modification de base, insertion et délétion) détectées dans les séquences par rapport à celle présentée par le gène *gfp* sauvage ont été vérifiées sur les chromatogrammes.

Pour chacune des trois banques, le séquençage a porté sur 13623 à 15774 pb. Le séquençage d'un nombre suffisamment important de bases permet d'obtenir une mesure statistiquement fiable des charges mutationnelles, même dans le cas de la banque contenant les gènes les moins mutés (banque A). Il est à noter que le ratio entre les clones exprimant une GFP fonctionnelle et ceux exprimant une GFP inactive a été conservé pour la sélection des 24 clones dédiés au séquençage. En effet, il a été décrit qu'à l'issue d'une étape de mutagenèse aléatoire les gènes codant pour une protéine active pourraient avoir une charge mutationnelle plus faible que des gènes non sélectionnés (Daugherty P et al. (2000) Proc. Natl. Acad. Sci. USA. 97 :2029-2034). Ainsi, la sélection aléatoire pour le séquençage d'un nombre statistiquement faible de clones aurait pu biaiser l'observation de la charge mutationnelle.

Le tableau 2 ci-après représente l'analyse de la fraction de clone exprimant une GFP fonctionnelle en fonction de la charge mutationnelle des banques de *gfp* mutés. Il résume les résultats obtenus sur l'analyse du test phénotypique GFP, tandis que la figure 2 présente le logarithme de la fraction de GFP fonctionnelle en fonction de la charge mutationnelle dans les banques de *gfp* mutés.

Tableau 2

| Expérience<br>PCR-ep | Conditions réactionelles |                      | Nombre de<br>pb séquencées | Nombre de<br>mutations | Charge mutationnelle<br>(mutations/gfp) | Fraction GFP<br>fonctionnelle (%) |
|----------------------|--------------------------|----------------------|----------------------------|------------------------|-----------------------------------------|-----------------------------------|
|                      | [MgCl <sub>2</sub> ] mM  | [MnCl <sub>2</sub> ] |                            |                        |                                         |                                   |
| A                    | 4                        | 0                    | 15057                      | 33                     | 1,57                                    | 75,1                              |
| B                    | 7                        | 0,3                  | 15774                      | 212                    | 9,64                                    | 16,5                              |
| C                    | 7                        | 0,5                  | 13623                      | 285                    | 15                                      | 2,9                               |

A partir des données générées, il est possible  
5 d'établir une courbe de régression polynomiale d'ordre  
2, représentant l'abaque selon l'équation suivante :

$$\ln(\text{Fa}[\text{GFP}]) = -0,0083 \times \text{CM}[\text{gfp}]^2 - 0,1111 \times \text{CM}[\text{gfp}]$$

liant la charge mutationnelle d'une banque  
de gènes obtenue par PCR error-prone du gène modèle *gfp*  
10 (C.M.[*gfp*]) à la fraction de gènes *gfp* mutés observés  
dans ladite banque lors du test phénotypique (Fa[GFP]).

La valeur du coefficient de régression ( $R^2$ )  
étant de 0,999 (cf figure 2), l'abaque précédemment  
déterminé permet d'estimer de façon précise la charge  
15 mutationnelle dans une banque de gènes *gfp* mutés à  
l'issue d'une étape de mutagenèse aléatoire par PCR  
error-prone à partir de la fraction de clone exprimant  
une GFP fonctionnelle.

L'analyse des séquençages sur les gènes *gfp*  
20 mutés a permis également d'analyser la matrice de  
mutation Mut<sub>GFP</sub> (figure 3) et d'en déduire la matrice  
cible de mutation sur le gène *gfp*, M<sub>target</sub>[*gfp*] (figure  
4).

#### 25 4) Application en parallèle de la mutagenèse aléatoire au gène modèle *gfp* et au gène d'intérêt *xynA* de *Thermotoga maritima*

Le gène *xynA* de xylanase issue de  
*Thermotoga maritima* (numéro d'accension EMBL Z46264) a  
30 été soumis à mutagenèse aléatoire par PCR error-prone  
avec 7mM MgCl<sub>2</sub> et 0,3mM MnCl<sub>2</sub> (conditions B). Le  
protocole expérimental est identique à celui  
précédemment décrit, excepté que l'étape d'élongation

PCR a été fixée à 3 minutes au lieu de 1 min. Il est à noter que le même mélange réactionnel a servi pour la PCR error-prone sur le gène *gfp* et pour la PCR error-prone sur le gène *xynA*. L'utilisation du gène *gfp* comme standard interne dans le cadre de l'invention doit permettre de s'affranchir des variations expérimentales sur les charges mutationnelles dans les banques de gènes modèle *gfp* mutés.

Les rendements des expériences de PCR error-prone ont été déterminés par comparaison avec une quantité connue de standard ADN après migration de 1µL de réaction sur gel agarose 1%. La quantité initiale de vecteur étant de 1fmole dans chacune des deux expériences menées, le nombre de multiplication (d) a pu être déterminé à 15 et 12 pour les PCR error-prone *gfp* et *xynA* respectivement.

5) Détermination de la charge mutationnelle de la banque obtenue à partir du gène modèle *gfp*

A l'issue de la PCR error-prone réalisée dans les conditions B, la fraction de gènes *gfp* codant pour une protéine fonctionnelle dans la banque a été déterminée à 0.165.

En utilisant l'abaque donné par l'équation suivante (équation 4) :

$$\ln(Fa[GFP]) = -0,0083 \times CM[gfp]^2 - 0,1111 \times CM[gfp]$$

liant la charge mutationnelle d'une banque de gènes modèle *gfp* mutés à la fraction de gènes modèle mutés observés, la charge mutationnelle dans la banque de *gfp* mutés est simulée à 9,5 mutations par gène.

Il est à noter que la charge mutationnelle dans cette même banque a été déterminée par séquençage à 9,64 mutations par gène *gfp* soit une erreur de 1,5%

entre la charge mutationnelle donnée par l'abaque et celle déterminée par séquençage.

5 6) Détermination de la charge mutationnelle introduite dans la banque de gènes *xynA* mutés obtenue à partir de la charge mutationnelle dans la banque de gènes modèle *gfp* mutés

10 La charge mutationnelle introduite dans la banque de gènes *xynA* mutés à l'issue de l'expérience de PCR error-prone est déterminée en corrigeant la charge mutationnelle calculée dans la banque de gènes modèle *gfp* mutés avec le facteur de correction C.F. définie selon l'équation suivante :

15 C.M. (gène d'intérêt) = C.M. (gène modèle) x C.F.  
dans laquelle, C.F. représente le facteur de correction, et C.M. représente la charge mutationnelle.

a) En tenant compte d'au moins un paramètre intrinsèque

20 Dans cette analyse, le paramètre intrinsèque retenu est la longueur des gènes, ce dernier étant reconnu par l'homme de métier comme ayant une forte influence sur la charge mutationnelle des banques de gènes mutés.

25 Le gène modèle correspondant à la *gfp* et la technique de mutagenèse correspondant à la PCR error-prone, l'unique sous-facteur de correction correspond pour :

30 paramètre intrinsèque 1 = longueur des gènes

Avec  $C.F.[L] = L(\text{gène d'intérêt})/717$  où L est en nombre de paire de base.

Le tableau 3 ci-après reprend les principaux paramètres intrinsèques, notamment la longueur du gène modèle *gfp* et du gène d'intérêt *xynA*.

Tableau 3

| Gène        | L (pb) | Composition en base (%) |    |      |      |
|-------------|--------|-------------------------|----|------|------|
|             |        | A                       | T  | C    | G    |
| <i>gfp</i>  | 717    | 34                      | 26 | 19,7 | 20,3 |
| <i>xynA</i> | 3120   | 32                      | 19 | 23   | 26   |

5

Ainsi la résolution donne :

$$C.F. = C.F.[L] = 3120/717 = 4,35$$

10 A l'issue de la simulation avec la longueur comme unique paramètre intrinsèque, la charge mutationnelle calculée dans la banque de gène *xynA* est de :

$$C.M.[xynA]_{sim1} = 4,35 \times 9,5 = 41,3 \text{ mutations/gène } xynA$$

15 b) En tenant compte d'au moins deux paramètres intrinsèques et d'un paramètre expérimental

Dans cette analyse, plus poussée que la précédente, les paramètres intrinsèques retenus sont la longueur des gènes et la composition en base. Le paramètre expérimental retenu est le nombre de multiplication de la matrice ADN lors de la PCR error-prone. Ces paramètres sont connus de l'homme de métier comme pouvant influencer la charge mutationnelle dans les banques de gènes à l'issue d'une mutagenèse par PCR error-prone.

25 Le gène modèle correspondant à la *gfp* et la technique de mutagenèse correspondant à la PCR error-prone, différents sous-facteurs de correction correspondent pour :

30 - paramètre intrinsèque 1 représente la longueur des gènes selon l'équation suivante :

$$C.F.[L] = L(\text{gène d'intérêt})/717$$

où L est en nombre de paire de base.

- paramètre intrinsèque 2 représente la composition en base sur les brins codant d'ADN selon l'équation suivante :

$$\text{Avec C.F.}[\text{base}] = \left( \frac{0,506 \times A_{g.\text{intérêt}}}{0,34} + \frac{0,346 \times T_{g.\text{intérêt}}}{0,26} + \frac{0,063 \times C_{g.\text{intérêt}}}{0,197} + \frac{0,085 \times G_{g.\text{intérêt}}}{0,203} \right)$$

5

où  $A_{g.\text{intérêt}}$ ,  $T_{g.\text{intérêt}}$ ,  $C_{g.\text{intérêt}}$  et  $G_{g.\text{intérêt}}$  sont en fractions.

- paramètre expérimental représente le taux de multiplication selon l'équation suivante :

10

$$\text{C.F.}[\text{d}] = d(\text{gène d'intérêt}) / d(\text{gfp})$$

Le tableau 3 reprend les principaux paramètres intrinsèques, longueur et composition en base, du gène modèle *gfp* et du gène d'intérêt *xynA*. Par ailleurs, à l'issue des PCR error-prone réalisées sur le gène modèle *gfp* et le gène d'intérêt *xynA*, les taux de multiplication sont de 15 et 12 respectivement pour *gfp* et *xynA*.

15

La résolution donne :

$$\text{C.F.}[\text{L}] = 3120/717 = 4,35$$

20

$$\text{C.F.}[\text{d}] = 12/15 = 0,80$$

$$\text{C.F.}[\text{base}] = \left( \frac{0,506 \times 0,32}{0,34} + \frac{0,346 \times 0,19}{0,26} + \frac{0,063 \times 0,23}{0,197} + \frac{0,085 \times 0,26}{0,203} \right) = 0,91$$

25

La construction du facteur de correction C.F. s'effectuant par la multiplication des différents sous-facteurs pris en compte pour la détermination de la charge mutationnelle de la banque de gènes d'intérêt mutés :

$$\text{C.F.} = \text{C.F.}[\text{L}] \times \text{C.F.}[\text{base}] \times \text{C.F.}[\text{d}]$$

Soit :

$$\text{C.F.} = 4,35 \times 0,91 \times 0,8 = 3,17$$

A l'issue de la simulation avec la longueur et la composition en base comme paramètres intrinsèques et le taux de multiplication comme paramètre expérimental, la charge mutationnelle calculée dans la banque de gènes *xynA* mutés est de :

$$C.M.[xynA]_{sim2} = 3,17 \times 9,5 = 30,1$$

mutations/gène *xynA*

La charge mutationnelle dans la banque de gènes *xynA* mutés a également été déterminée par séquençage pour valider les méthodes de détermination de charge mutationnelle découlant de l'invention.

Douze clones recombinant *xynA* *Thermotoga maritima* mutés ont été sélectionnés pour séquençage. Les gènes xylanases mutés ont été entièrement séquencés comme décrit dans l'exemple 1, (3).

Parmi les 38160pb séquencées, 381 mutations ont été détectées dans les gènes mutés *xynA*. Ainsi la charge mutationnelle observée dans la banque de gènes *xynA* mutés est de 31,8 mutations / gène *xynA*.

L'erreur constatée entre la charge mutationnelle observée et la charge mutationnelle simulée dans la banque de gène *xynA* est de 29,9% pour la simulation 1 (réalisée avec la seule longueur comme paramètre intrinsèque), et de 5,3% pour la simulation 2 (réalisée avec la longueur et la composition en base comme paramètres intrinsèques et le taux de multiplication comme paramètre expérimental).

La prise en compte, dans les simulations, de la composition en base et du taux de multiplication du gène modèle *gfp* et du gène d'intérêt *xynA* permet de diminuer l'erreur sur la charge mutationnelle d'environ un facteur six. Un tel résultat souligne l'intérêt,

présenté par l'invention, d'intégrer dans le facteur de correction C.F. non pas la seule longueur des gènes mais le plus grand nombre de paramètres intrinsèques et expérimentaux identifiés comme susceptibles d'influencer la charge mutationnelle.

Par ailleurs la justesse du résultat de la charge mutationnelle obtenu dans le cadre de la simulation 2 par rapport à celle observée lors d'un séquençage (moins de 6% d'erreur) démontre la pertinence de l'invention à pouvoir remplacer de manière fiable l'étude de la charge mutationnelle dans une banque de gènes mutés par séquençage.

L'ensemble des résultats sur les charges mutationnelles simulées ou observées dans la banque de gènes *xynA* mutés est présenté dans le tableau 4 ci-après.

Ce tableau 4 récapitule les résultats de charge mutationnelle sur une banque de gènes *xynA* mutés, observée par séquençage (C.M.[*xynA*]*obs*), simulée en tenant compte uniquement de la longueur des gènes (simulation A) ou simulée en tenant compte de la longueur des gènes, de la composition en base et du nombre de multiplication de la matrice ADN (simulation B).

Tableau 4

| Simulation                                   | C.M.[ <i>xynA</i> ] <i>sim</i> | Erreur sim./obs. |
|----------------------------------------------|--------------------------------|------------------|
| A                                            | 41,3                           | 29,9%            |
| B                                            | 30,1                           | 5,3%             |
| <u>C.M.[<i>xynA</i>]<i>obs</i>      31,8</u> |                                |                  |

Exemple 2. Analyse de la plasticité protéique de TEM-1  $\beta$ -lactamase

Cette analyse est effectuée selon les étapes suivantes :

- 1) sélection du gène modèle ;
- 2) sélection de la méthode de mutagenèse aléatoire ;

3) obtention d'un abaque permettant de lier la charge mutationnelle d'une banque de gènes modèle *gfp* mutés aléatoirement à la fraction de gènes modèle mutés observés dans ladite banque ;

Ces étapes sont identiques à celles développées dans l'exemple 1. Par contre cet exemple comprend les étapes 4 à 7 décrites ci-après.

4) Application en parallèle de la mutagenèse aléatoire par PCR error-prone au gène modèle *gfp* et au gène d'intérêt *TEM-1*

Le gène modèle *gfp* et le gène d'intérêt *TEM-1* (numéro d'accèsion EMBL AAR25033) ont été soumis à quatre expériences de mutagenèse aléatoire par PCR error-prone avec 4mM MgCl<sub>2</sub> (condition I), 7mM MgCl<sub>2</sub> plus 0,15mM MnCl<sub>2</sub> (condition II), 7mM MgCl<sub>2</sub> plus 0,3mM MnCl<sub>2</sub> (condition III) et 7mM MgCl<sub>2</sub> plus 0,5mM MnCl<sub>2</sub> (condition IV). Le protocole expérimental est identique à celui décrit dans l'exemple 1.

Les rendements des expériences de PCR error-prone ont été déterminés comme décrit dans l'exemple 1. Le nombre de multiplication a ainsi pu être déterminé à 15 et 14 pour les PCR error-prone *gfp* et *TEM-1* respectivement.

30

5) Détermination des charges mutationnelles dans les banques obtenues à partir du gène modèle *gfp*

A l'issue des quatre expériences de PCR error-prone, les fractions des banques de gènes *gfp*

mutés codant pour une protéine GFP fonctionnelle ont été déterminées avec l'équation suivante (équation 3) :

$$Fa[GFP] = \frac{cfu\_fluo}{cfu\_total \times Ri}$$

5 La détermination des charges mutationnelles dans les banques de gènes *gfp* mutés est effectuée en utilisant l'abaque décrit dans l'équation suivante (équation 4) :

$$\ln(Fa[GFP]) = -0,0083 \times CM[gfp]^2 - 0,1111 \times CM[gfp]$$

10 à partir de la fraction de gène *gfp* codant pour une protéine GFP active ( $Fa[GFP]$ ).

Le tableau 5 ci-après reprend les résultats obtenus à l'issue de cette analyse. Il donne les charges mutationnelles dans les banques de gènes *gfp* mutés à partir de l'analyse du test phénotypique mis en place sur la protéine GFP. Ri est le rendement d'insertion des gènes *gfp* mutés dans les différentes banques, déterminé par PCR/colonie; cfu\_totales et cfu\_fluo sont respectivement le nombre de clones analysés et le nombre de clones fluorescents verts aux U.V.;  $Fa[GFP]$  est la fraction fonctionnelle de clones GFP tandis que C.M.[*gfp*] est la charge mutationnelle dans les différentes banques de gènes *gfp* mutés.

**Tableau 5**

| Conditions PCR-ep | cfu_totales | cfu_fluo | Ri    | Fa[GFP] | C.M.[ <i>gfp</i> ] |
|-------------------|-------------|----------|-------|---------|--------------------|
| I                 | 270         | 226      | 0,958 | 0,874   | 1,1                |
| II                | 340         | 179      | 0,958 | 0,550   | 4,1                |
| III               | 775         | 135      | 0,958 | 0,182   | 9,1                |
| IV                | 1160        | 52       | 0,958 | 0,047   | 13,6               |

25

6) Détermination des charges mutationnelles introduites dans les banques de gènes TEM-1 mutés

obtenues à partir des charges mutationnelles dans les banques de gènes modèle *gfp* mutés

Comme précédemment, les charges mutationnelles introduites dans les banques de gènes *TEM-1* mutés à l'issue des expériences de PCR error-prone sont déterminées en corrigeant les charges mutationnelles calculées dans les banques de gènes modèle *gfp* mutés avec le facteur de correction C.F. selon l'équation suivante (équation 1) :

C.M. (gène d'intérêt) = C.M. (gène modèle) x C.F.

où C.F. représente le facteur de correction et C.M. représente la charge mutationnelle.

Pour obtenir un maximum de précision sur la détermination des charges mutationnelles dans les banques de gènes *TEM-1* mutés, le facteur de correction est calculé en tenant compte de deux paramètres intrinsèques (la longueur et la composition en base des gènes) et d'un paramètre expérimental (le taux de multiplication des matrices ADN lors de la PCR error-prone).

Le gène modèle correspondant à la *gfp* et la technique de mutagenèse correspondant à la PCR error-prone, différents sous-facteurs de correction correspondent pour :

- paramètre intrinsèque 1 représente la longueur des gènes selon l'équation suivante :

$$C.F.[L] = L(\text{gène d'intérêt})/717$$

où L est en nombre de paire de bases.

- paramètre intrinsèque 2 représente la composition en bases sur les brins codant d'ADN selon l'équation suivante :

$$\text{Avec C.F.}[\text{base}] = \left( \frac{0,506 \times A_{g.\text{intéret}}}{0,34} + \frac{0,346 \times T_{g.\text{intéret}}}{0,26} + \frac{0,063 \times C_{g.\text{intéret}}}{0,197} + \frac{0,085 \times G_{g.\text{intéret}}}{0,203} \right)$$

où  $A_{g.\text{intéret}}$ ,  $T_{g.\text{intéret}}$ ,  $C_{g.\text{intéret}}$  et  $G_{g.\text{intéret}}$  sont en fraction.

- paramètre expérimental représente le  
5 taux de multiplication selon l'équation suivante :

$$\text{C.F.}[\text{d}] = d(\text{gène d'intéret}) / d(\text{gfp})$$

Le tableau 6 reprend les principaux  
paramètres intrinsèques, longueur et composition en  
10 bases du gène modèle *gfp* et du gène d'intérêt *TEM-1*.  
Par ailleurs, à l'issue des PCR error-prone réalisées  
sur le gène modèle *gfp* et le gène d'intérêt *xynA*, les  
taux de multiplication sont de 15 et 14 pour *gfp* et  
*TEM-1* respectivement.

15

Tableau 6. Paramètres intrinsèques du gène modèle *gfp*  
et du gène d'intérêt *TEM-1*.

| Gène         | L (pb) | Composition en base (%) |      |      |      |
|--------------|--------|-------------------------|------|------|------|
|              |        | A                       | T    | C    | G    |
| <i>gfp</i>   | 717    | 34                      | 26   | 19,7 | 20,3 |
| <i>TEM-1</i> | 861    | 25,9                    | 24,7 | 23,6 | 25,8 |

20

La résolution donne :

$$\text{C.F.}[\text{L}] = 861/717 = 1,2$$

$$\text{C.F.}[\text{base}] = \left( \frac{0,506 \times 0,259}{0,34} + \frac{0,346 \times 0,247}{0,26} + \frac{0,063 \times 0,236}{0,197} + \frac{0,085 \times 0,258}{0,203} \right) = 0,9$$

$$\text{C.F.}[\text{d}] = 14/15 = 0,93$$

25

Soit :

$$\text{C.F.} = 1,2 \times 0,9 \times 0,93 = 1,01$$

Dans la détermination du facteur de correction C.F., la plus grande longueur du gène *TEM-1* est compensée par un plus fort pourcentage en base C et G et une diminution du rendement de PCR error-prone dans le cas du gène *TEM-1* par rapport au gène modèle *gfp*.

A l'issue de la simulation avec deux paramètres intrinsèques et un paramètre expérimental, les charges mutationnelles calculées dans les banques de gène *TEM-1* sont de :

- Pour la condition I :

$$\text{C.M.}[TEM-1]_{sim} = 1,01 \times 1,1 = 1,1$$

mutations/gène *TEM-1*

- Pour la condition II :

$$\text{C.M.}[TEM-1]_{sim} = 1,01 \times 4,1 = 4,1$$

mutations/gène *TEM-1*

- Pour la condition III :

$$\text{C.M.}[TEM-1]_{sim} = 1,01 \times 9,1 = 9,2$$

mutations/gène *TEM-1*

- Pour la condition IV :

$$\text{C.M.}[TEM-1]_{sim} = 1,01 \times 13,6 = 13,7$$

mutations/gène *TEM-1*

#### 6) Analyse de la plasticité protéique de *TEM-1* $\beta$ -lactamase

L'étude de la plasticité protéique nécessite de pouvoir lier le taux de survie des produits du gène d'intérêt muté, sous un crible donné, dans une banque de gènes d'intérêt mutés en fonction de la charge mutationnelle dans ladite banque (Shafikhani S. et al. (1997) BioTechniques 23 :304-310 / Daugherty P. et al. (2000) Proc. Natl. Acad. Sci. USA 97 :2029-2034).

Pour analyser le taux de survie de TEM-1  $\beta$ -lactamase dans les quatres banques de PCR error-prone précédemment générées, les produits de PCR (1 $\mu$ g) ont été digérés par Eco RI et Nde I (New England Biolabs) pendant 1h à 37°C dans le tampon de réaction adéquat comme décrit par le fournisseur puis purifiés sur colonne QIAquick et insérés dans le vecteur pET26b+ (Stratagene) préalablement digéré avec Eco RI et Nde I. Les cellules *E. coli* MC1061(DE3) ont été transformées par choc thermique comme décrit par Maniatis (Sambrook J. et al. (1989) Molecular cloning: a laboratory manual Cold Spring Harbor Laboratory Press -New York) avec 5 $\mu$ L des réactions de ligation. Le nombre de transformants dans les différentes banques a été estimé en étalant une fraction de la banque sur boites LB agar complémentées avec 0,6mg/mL de kanamycine tandis que le nombre de clones exprimant une TEM-1  $\beta$ -lactamase fonctionnelle a été estimé en étalant le reste de la banque sur boites LB agar complémentées avec 0,6mg/mL de kanamycine et 10 $\mu$ g/mL d'ampicilline. Le rendement d'insertion des gènes mutés TEM-1 a été estimé par PCR sur colonie comme précédemment décrit.

Le tableau 7 reprend les résultats obtenus sur les taux de survie des protéines de TEM-1  $\beta$ -lactamase mutées dans les 4 banques d'error-prone. La fraction fonctionnelle de clones TEM-1 (Fa[TEM]), est déterminée à partir du rendement d'insertion des gènes TEM-1 mutés (Ri), du nombre de clones analysés (cfu\_totales) et du nombre de clones résistants sur ampicilline (cfu\_R).

Tableau 7 : Taux de survie de la protéine TEM-1  $\beta$ -lactamase dans les 4 banques de PCR error-prone.

| Conditions | cfu_totales | cfu_R | Ri    | Fa[TEM] |
|------------|-------------|-------|-------|---------|
| I          | 761         | 539   | 0,958 | 0,739   |
| II         | 746         | 265   | 0,958 | 0,371   |
| III        | 1215        | 55    | 0,938 | 0,048   |
| IV         | 1143        | 7     | 0,938 | 0,0065  |

5 La figure 5 présente l'analyse de la plasticité protéique de TEM-1 Beta-lactamase, à savoir le taux de survie des protéines TEM-1 Beta-lactamase mutées en fonction de la charge mutationnelle dans les banques de gènes *TEM-1* mutés.

10 En conclusion, la présente invention permet d'étudier, dans une forme particulière de mise en oeuvre, la plasticité protéique sans recours au séquençage.

15 Exemple 3 : Analyse de l'évolution de TEM-1 Beta-lactamase vers la résistance à l'antibiotique cefotaxime en fonction de la charge mutationnelle des banques de PCR error-prone.

20 Cette analyse comprend les étapes suivantes:

1) sélection du gène modèle.

2) sélection de la méthode de mutagènese aléatoire.

25 3) obtention d'un abaque permettant de lier la charge mutationnelle d'une banque de gènes modèle *gfp* mutés aléatoirement à la fraction de gènes modèle mutés observés dans ladite banque.

Ces étapes sont identiques à celles développées dans l'exemple 1 et 2. Par contre, cette analyse comprend aussi les étapes 4-7 décrites ci-après.

5

4) Application en parallèle de la mutagenèse aléatoire par PCR error-prone au gène modèle *gfp* et au gène d'intérêt *TEM-1*

10

Le gène modèle *gfp* et le gène d'intérêt *TEM-1* ont été soumis à cinq expériences de mutagenèse aléatoire par PCR error-prone avec 5mM MgCl<sub>2</sub> (condition 1), 7mM MgCl<sub>2</sub> plus 0,1mM MnCl<sub>2</sub> (condition 2), 7mM MgCl<sub>2</sub> plus 0,2mM MnCl<sub>2</sub> (condition 3), 7mM MgCl<sub>2</sub> plus 0,3mM MnCl<sub>2</sub> (condition 4) et 7mM MgCl<sub>2</sub> plus 0,5mM MnCl<sub>2</sub> (condition 5). Le protocole expérimental est identique à celui décrit dans l'exemple 2.

15

20

Les rendements des expériences de PCR error-prone ont été déterminés comme décrit dans l'exemple 1 et sont identiques à ceux trouvés dans l'exemple 2. Le taux de multiplication est de 15 et de 14 pour les PCR error-prone *gfp* et *TEM-1* respectivement.

25

5) Détermination des charges mutationnelles dans les banques obtenues à partir du gène modèle *gfp*

A l'issue des cinq expériences de PCR error-prone, les fractions des banques de gènes *gfp* mutés codant pour une protéine fonctionnelle ont été déterminées avec l'équation suivante (équation 3):

$$Fa[GFP] = \frac{cfu\_fluo}{cfu\_total \times Ri}$$

30

La détermination des charges mutationnelles dans les banques de gènes *gfp* mutés est effectuée en utilisant l'abaque décrit dans l'équation suivante :

$$\ln(\text{Fa}[\text{GFP}]) = -0,0083 \times \text{CM}[\text{gfp}]^2 - 0,1111 \times \text{CM}[\text{gfp}]$$

5 à partir de la fraction de gènes *gfp* mutés codant pour une protéine GFP active ( $\text{Fa}[\text{GFP}]$ ).

Le tableau 8 ci-après reprend les résultats obtenus à l'issue de cette analyse. Ce tableau rapporte la détermination des charges mutationnelles dans les banques de gènes *gfp* mutés à partir de l'analyse du test phénotypique GFP. Ri est le rendement d'insertion des gènes *gfp* mutés dans les différentes banques, cfu\_totales et cfu\_fluo sont respectivement le nombre de clones analysés et le nombre de clones fluorescents aux U.V.,  $\text{Fa}[\text{GFP}]$  est la fraction fonctionnelle de clones GFP tandis que  $\text{C.M.}[\text{gfp}]$  est la charge mutationnelle dans les différentes banques de gènes *gfp* mutés.

Tableau 8

| <u>Conditions PCR-ep</u> | <u>cfu_totales</u> | <u>cfu_fluo</u> | <u>Ri</u> | <u>Fa[GFP]</u> | <u>C.M.[gfp]</u> |
|--------------------------|--------------------|-----------------|-----------|----------------|------------------|
| 1                        | 425                | 305             | 0,96      | 0,749          | 2,2              |
| 2                        | 450                | 268             | 0,96      | 0,621          | 3,4              |
| 3                        | 568                | 212             | 0,96      | 0,389          | 5,9              |
| 4                        | 842                | 100             | 0,96      | 0,124          | 10,5             |
| 5                        | 1093               | 26              | 0,75      | 0,032          | 14,8             |

20

6) Détermination des charges mutationnelles introduites dans les banques de gènes *TEM-1* mutés obtenues à partir des charges mutationnelles dans les banques de gènes modèle *gfp* mutés

25

Le facteur de correction est calculé en tenant compte de deux paramètres intrinsèques, la longueur et la composition en bases des gènes, et d'un

paramètre expérimental, le taux de multiplication des matrices ADN lors de la PCR error-prone.

Le gène modèle *gfp* et le gène d'intérêt *TEM-1* étant identique à celui de l'exemple 2 et les taux de multiplication des matrices *gfp* et *TEM-1* étant également inchangés entre les exemple 1 et 2, le facteur de correction C.F. est identique à celui déterminé dans l'exemple 2.

Soit C.F. = 1,01

A l'issue de la simulation avec deux paramètres intrinsèques et un paramètre expérimental, les charges mutationnelles calculées dans les cinq banques de gènes *TEM-1* mutés sont de :

- Pour la condition 1 :

$C.M.[TEM-1]_{sim} = 1,01 \times 2,2 = 2,2$   
mutations/gène *TEM-1*

- Pour la condition 2 :

$C.M.[TEM-1]_{sim} = 1,01 \times 3,4 = 3,4$   
mutations/gène *TEM-1*

- Pour la condition 3 :

$C.M.[TEM-1]_{sim} = 1,01 \times 5,9 = 6,0$   
mutations/gène *TEM-1*

- Pour la condition 4 :

$C.M.[TEM-1]_{sim} = 1,01 \times 10,5 = 10,6$   
mutations/gène *TEM-1*

- Pour la condition 5 :

$C.M.[TEM-1]_{sim} = 1,01 \times 14,8 = 14,9$   
mutations/gène *TEM-1*

7) Analyse de l'acquisition de la résistance à la cefotaxime des clones recombinant *TEM-1*  $\beta$ -lactamase mutés

Le système d'évolution choisi décrit l'influence de la charge mutationnelle de la banque de gène *TEM-1* mutés sur la fréquence d'apparition de clones résistant à la cefotaxime grâce à l'évolution du gène *TEM-1*.

Pour analyser la fréquence d'apparition de clones résistants, les produits PCR sont clonés dans le vecteur pET26b+ comme décrit précédemment pour former 5 banques de gènes *TEM-1* mutés. Les banques de clones sont réalisées par électrotransformation de cellules *E. coli* MC1061(DE3) avec 10×2μL de ligation. Le nombre de transformants dans chacune des cinq banques a été estimé en étalant 1/25<sup>ème</sup> de la banque sur boîtes LB agar complémentées avec 0,6mg/mL de kanamycine tandis que les clones résistant à la cefotaxime ont été isolés en étalant le reste de la banque (24/25<sup>ème</sup>) sur boîtes LB agar complémentées avec 0,6mg/mL de kanamycine et 15ng/mL de cefotaxime. Le rendement d'insertion des gènes *TEM-1* mutés a été estimé par PCR sur colonie.

Le tableau 9 ci-après reprend les résultats obtenus sur la fréquence d'apparition de clones, résistant à la cefotaxime, dans les 5 banques de *TEM*-mutés. La fréquence de clones améliorés dans les différentes banques est calculée à partir du nombre de clones résistant sur cefotaxime (cfu\_cefoR) et du nombre de clones recombinants (Nb\_recomb.), lui même déduit du nombre de clones dans la banque (cfu\_totales) et du rendement d'insertion des gènes *TEM-1* (Ri) mutés.

Tableau 9

| Conditions | cfu_totales | Ri   | Nb_recomb. | cfu_cefoR | Fréq. améliorés ‰ |
|------------|-------------|------|------------|-----------|-------------------|
| 1          | 103200      | 0,96 | 98900      | 51        | 0,516             |
| 2          | 53750       | 0,96 | 51510      | 41        | 0,796             |
| 3          | 189000      | 0,96 | 181125     | 50        | 0,276             |
| 4          | 158000      | 0,94 | 148125     | 18        | 0,122             |
| 5          | 223500      | 0,92 | 204875     | 11        | 0,054             |

La figure 6 montre l'influence de la charge mutationnelle des banques de gènes *TEM-1* mutés sur la fréquence d'apparition d'un variant *TEM-1* amélioré.

La fréquence d'apparition de clones résistants à la cefotaxime en fonction de la charge mutationnelle suit une courbe en cloche passant par une charge mutationnelle optimale.

La méthode de l'invention peut ainsi être utilisée pour analyser les taux d'évolution des banques de gènes mutés mais également pour se placer à la charge mutationnelle optimale et ceci sans recourir au séquençage.

En conclusion, cet exemple montre également l'importance de contrôler précisément la charge mutationnelle dans les banques de gènes mutés puisque la fréquence d'apparition de variants *TEM-1* résistants à la céfotaxime est diminuée de trois fois pour une charge mutationnelle passant de 3,4 à 6 mutations par gène *TEM-1*. La présente invention permet, sans séquençage, un contrôle de la charge mutationnelle beaucoup plus fin que les méthodes proposées par l'art antérieur et répond parfaitement à ce type de besoin.

Exemple 4 : comparaison et estimation de la fidélité d'un jeu de ADN polymerases dans une réaction de PCR donnée.

Cette mise en oeuvre particulière de l'invention comprend les étapes suivantes :

1) sélection du gène modèle.

2) sélection de la méthode de mutagèneses aléatoire.

3) obtention d'un abaque permettant de lier la charge mutationnelle d'une banque de gènes modèle *gfp* mutés aléatoirement à la fraction de gènes modèle mutés observés dans ladite banque.

Ces étapes sont identiques à celles développées dans les exemple 1 à 3. L'analyse comprend en outre les étapes 4-6 décrites ci-après.

4) Application en parallèle de la PCR au gène modèle *gfp* avec un jeux de ADN polymerases

Le gène modèle *gfp* a été soumis à trois expériences de PCR contenant 7mM de  $MgCl_2$  et 0,2mM de  $MnCl_2$ . Le protocole expérimental est identique à celui décrit dans les exemples précédents avec un temps d'élongation de 1 minute, excepté que 2,5U de *Taq* polymerase (Q-Biogen), de *Vent<sub>R</sub>* (New England Biolabs) ou de *Pfu* Turbo (Stratagène) ont été utilisée comme ADN polymérase.

Les rendements des expériences de PCR ont été déterminés comme décrit précédemment. Le nombre de multiplication est de  $d_A$ ,  $d_B$  et  $d_C$  pour les PCR effectuées avec la *Taq* DNA polymerase, la *Vent<sub>R</sub>* DNA polymerase et la *Pfu* DNA polymerase respectivement.

5) Détermination des charges mutationnelles dans les banques obtenues à partir du gène modèle *gfp*

A l'issue des expériences de PCR, les fractions des banques *gfp* mutés codant pour une GFP

fonctionnelle ont été déterminées avec l'équation suivante (équation 3) :

$$Fa[GFP] = \frac{cfu\_fluo}{cfu\_total \times Ri}$$

Les charges mutationnelles dans les banques de gènes *gfp* mutés sont déterminées en utilisant l'abaque décrit dans l'équation suivante (équation 4) :

$$\ln(Fa[GFP]) = -0,0083 \times CM[gfp]^2 - 0,1111 \times CM[gfp]$$

à partir de la fraction de gènes *gfp* mutés codant pour une protéine GFP active ( $Fa[GFP]$ ).

Les charges mutationnelles sont de A, B et C pour les PCR effectuées avec la *Taq* DNA polymerase, la *Vent<sub>R</sub>* DNA polymerase et la *Pfu* DNA polymerase respectivement.

#### 6) Comparaison de la fidélité des trois polymerases utilisées dans les expériences de PCR

Les taux d'erreur des différentes polymerases dans les conditions réactionnelles utilisées sont déduits des charges mutationnelles dans les banques de gènes *gfp* mutés en normalisant par la longueur du gène modèle *gfp* et les taux de multiplication des PCR.

La longueur du gène modèle *gfp* est de 717 paires de bases et les taux de multiplication à l'issue des PCR sont de  $d_A$ ,  $d_B$  et  $d_C$  avec la *Taq* DNA polymerase, la *Vent<sub>R</sub>* DNA polymerase et la *Pfu* DNA polymerase respectivement.

Ainsi :

$$\text{Taux erreur [Taq DNA polymerase]} = A / (717 \times d_A) \text{ erreur /pb/ duplication}$$

Taux erreur [Vent<sub>R</sub> DNA polymerase] =  
B/(717×d<sub>B</sub>) erreur /pb/ duplication

Taux erreur [Pfu DNA polymerase] =  
C/(717×d<sub>C</sub>) erreur /pb/ duplication

5                   En conclusion, la méthode de l'invention  
permet de déterminer précisément les fidélités d'un  
jeux d'ADN polymerase dans un mélange réactionnel donné  
sans recourir au séquençage, et ceci en  
s'affranchissant des variations de conditions  
10                   expérimentales.

## SEQUENCE LISTING

<110> PROTEUS  
 CHODORGE, Matthieu  
 FOURAGE, Laurent  
 LEFEVRE, Fabrice  
 MASSON, Jean-Michel

<120> PROCEDE DE DETERMINATION DE LA CHARGE MUTATIONNELLE D'UNE BANQUE DE  
 GENES OBTENUE PAR MUTAGENESE ALEATOIRE D'UN GENE D'INTER ET LES MOYENS POUR  
 SA MISE EN OEUVRE

<130> 36647/PCT

<140> PCT/FR2005/001392

<141> 2005-07-06

<150> FR 04/06120

<151> 2004-07-06

<160> 1

<170> PatentIn version 3.1

<210> 1

<211> 717

<212> DNA

<213> Artificial sequence

<220>

<223> gfp gene

<400> 1

|             |            |            |             |             |            |     |
|-------------|------------|------------|-------------|-------------|------------|-----|
| tttcaagagt  | gccatgcccg | agggttatgt | acaggaaaga  | actatatattt | tcaaagatga | 60  |
| cgggaactac  | aagacacgtg | ctgaagtcaa | gtttgaagggt | gatacccttg  | ttaatagaat | 120 |
| cgagttaaaa  | ggtgttgatt | ttaaagaaga | tggaacatt   | cttggacaca  | aattggaata | 180 |
| caactataac  | tcacacaatg | tatacatcat | ggcagacaaa  | caaaagaatg  | gaatcaaagc | 240 |
| taacttcaaa  | gttagacaca | acattgaaga | tggaagcggt  | caactggcag  | accattatca | 300 |
| acaaaatact  | ccaattggcg | atggccctgt | ccttttacca  | gacaaccatt  | acctgtccac | 360 |
| acaatctgcc  | ctttcgaaag | atcccaacga | aaagagagac  | cacatgggtcc | ttcttgagtt | 420 |
| tgtaacagct  | gctgggatta | cacatggcat | ggatgaacta  | tacaaataaa  | tgagtaaagg | 480 |
| agaagaactt  | ttcactagag | ttgtcccaat | tcttgttgaa  | ttagacgggtg | atgttaatgg | 540 |
| gcacaaatth  | tctgtcagtg | gagagggaga | aggtgatgca  | acatacggaa  | aacttaccct | 600 |
| taaattttatt | tgcactactg | gaaaactacc | agttccgtgg  | ccaacacttg  | tcactactct | 660 |
| ctcttatggt  | gttcaatgct | tttcgagata | cccagatcac  | atgaaacggc  | atgactt    | 717 |

85387-79

56

## REVENDICATIONS

1. Procédé de détermination de la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène d'intérêt, comprenant les étapes suivantes :

a) préparation d'un abaque liant la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèse aléatoire d'un gène modèle à la fraction des gènes modèles mutés observée dans ladite banque, comprenant les étapes suivantes :

1) sélection du gène modèle,

2) sélection de la méthode de mutagenèse aléatoire, et

3) obtention dudit abaque ;

b) mutagenèse aléatoire du gène modèle ayant servi pour la préparation de l'abaque de l'étape (a) et du gène d'intérêt pour obtenir les banques de gènes mutés correspondantes ;

c) détermination de la charge mutationnelle (CM) de la banque de gènes obtenue à partir du gène modèle à l'étape (b) à partir de l'abaque établi à l'étape (a) ; et

d) application d'un facteur de correction (CF) à la charge mutationnelle (CM) de la banque de gènes modèles mutés déterminée à l'étape (c) pour déterminer la charge mutationnelle (CM) de la banque de gènes d'intérêt mutés de l'étape (b), où ledit facteur de correction (C.F.) se décompose en plusieurs sous-facteurs définis suivant des paramètres sélectionnés comme influençant la charge mutationnelle d'une banque de gènes d'intérêt mutés par rapport à une banque de gènes modèles mutés,

85387-79

57

où les étapes (c) et (d) ne comprennent pas de séquençage.

2. Procédé selon la revendication 1, où la technique de mutagenèse mise en oeuvre à l'étape (b) est identique pour le gène modèle et pour le gène d'intérêt et est également similaire à celle de l'étape (a) pour construire l'abaque à partir du gène modèle.

3. Procédé selon la revendication 1 ou 2, où pour la préparation de l'abaque de l'étape (a) l'observation des gènes modèles mutés dans ladite banque est réalisée en analysant directement ou indirectement une ou plusieurs modifications d'une ou plusieurs propriétés de la séquence nucléotidique des gènes modèles mutés par rapport au gène modèle avant mutagenèse, et où ladite observation consiste à observer la modification d'une ou plusieurs propriétés de la protéine codée par le gène modèle.

4. Procédé selon la revendication 3, où l'observation de la modification d'une ou plusieurs propriétés de la protéine codée par le gène modèle est réalisée par un test phénotypique.

5. Procédé selon l'une quelconque des revendications 1 à 4, où le gène modèle code une protéine Green Fluorescent Protein (GFP).

6. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que le gène modèle possède des caractéristiques proches de celles du gène d'intérêt.

7. Procédé selon la revendication 6, où les caractéristiques du gène modèle comprennent :

- sa longueur,
- sa composition en base,

85387-79

58

- sa fréquence en bases répétées dans la séquence,
- la présence de séquences symétriques dans le gène.

8. Procédé selon l'une quelconque des revendications 1 à 7, où la mutagenèse aléatoire des étapes (a) et (b) met en oeuvre la polymérisation.

9. Procédé selon la revendication 8, où la polymérisation est la PCR error-prone.

10. Procédé selon l'une quelconque des revendications 1 à 9, où à l'étape (b) un même mélange réactionnel est utilisé pour la mutagenèse du gène modèle et du gène d'intérêt.

11. Procédé selon la revendication 10, où les paramètres sélectionnés pouvant influencer la charge mutationnelle sont choisis parmi :

- au moins un paramètre intrinsèque,
- au moins un paramètre expérimental.

12. Procédé selon la revendication 10, où les paramètres sélectionnés pouvant influencer la charge mutationnelle sont choisis parmi :

- au moins deux paramètres intrinsèques,
- au moins un paramètre expérimental.

13. Procédé selon la revendication 10, où ledit au moins un paramètre intrinsèque est choisi parmi les caractéristiques suivantes d'un gène:

85387-79

59

- sa longueur,
- sa composition en base,
- bases répétées dans la séquence,
- la présence de séquences symétriques dans le gène.

14. Procédé selon la revendication 12, où lesdit au moins deux paramètres intrinsèques sont choisis parmi les caractéristiques suivantes d'un gène:

- sa longueur,
- sa composition en base,
- bases répétées dans la séquence,
- la présence de séquences symétriques dans le gène.

15. Procédé selon l'une quelconque des revendications 11 à 13, où, lorsque la mutagenèse est basée sur une amplification en chaîne utilisant une polymérase, les paramètres expérimentaux correspondent à un ou plusieurs des facteurs suivants :

- le taux d'amplification du gène durant l'étape de mutagenèse,
- la quantité de matrice initiale,
- le nombre de cycle d'amplification,
- à la nature et à la quantité de polymérase utilisée

85387-79

60

- aux conditions expérimentales utilisées comme la concentration en cations divalents, la concentration et le ratio de concentration des dNTPs.

16. Procédé selon l'une quelconque des revendications 1 à 15, où au moins un paramètre intrinsèque et au moins un paramètre expérimental sont considérés pour établir le facteur de correction.

17. Procédé selon l'une quelconque des revendications 1 à 16, où lorsque la mutagenèse est réalisée par PCR error-prone, et le gène modèle code pour la Green Fluorescent Protein (GFP), les différents sous-facteurs de correction sont choisis parmi :

- un paramètre intrinsèque 1 représentant la longueur des gènes selon l'équation suivante :

$$C.F.[L] = L(\text{gène d'intérêt})/717$$

où L est en nombre de paire de base,

- un paramètre intrinsèque 2 représentant

$$C.F.[\text{base}] = \left( \frac{0,506 \times A_{(g.\text{intérêt})}}{0,34} + \frac{0,346 \times T_{(g.\text{intérêt})}}{0,26} + \frac{0,063 \times C_{(g.\text{intérêt})}}{0,197} + \frac{0,085 \times G_{(g.\text{intérêt})}}{0,203} \right)$$

la composition en base selon l'équation suivante :

où  $A_{(g.\text{intérêt})}$ ,  $T_{(g.\text{intérêt})}$ ,  $C_{(g.\text{intérêt})}$  et  $G_{(g.\text{intérêt})}$  sont en fraction,

- un paramètre expérimental représentant un taux de multiplication selon l'équation suivante :

$$C.F.[d] = d(\text{gène d'intérêt}) / d(\text{gfp})$$

85387-79

61

où  $d$  est le taux de multiplication calculé selon l'équation suivante :

$$d = \ln(Q_f/Q_i)/\ln 2$$

dans laquelle,  $Q_i$  et  $Q_f$  représentent respectivement les quantités molaires de la matrice et des produits à l'issue de la PCR.

18. Utilisation du procédé selon l'une quelconque des revendications 1 à 17 pour contrôler la charge mutationnelle dans une banque de gènes mutés, où ledit procédé comprend une étape supplémentaire de sélection des conditions expérimentales de mutagenèse aléatoire qui seront utilisées pour mettre en œuvre l'étape (b).

19. Utilisation du procédé selon l'une quelconque des revendications 1 à 17 pour étudier la capacité d'une protéine à incorporer des mutations tout en conservant une activité présentée par la protéine sauvage.

20. Utilisation du procédé selon l'une quelconque des revendications 1 à 17 pour étudier la capacité d'une protéine à acquérir une nouvelle propriété avec une charge mutationnelle donnée.

21. Utilisation du procédé selon l'une quelconque des revendications 1 à 17 pour déterminer la charge mutationnelle optimale dans des banques de gènes d'intérêt mutés pour générer un lot de gènes parentaux avant la mise en oeuvre de méthodes de recombinaison moléculaire.

85387-79

62

22. Utilisation du procédé selon l'une quelconque des revendications 1 à 17 pour quantifier ou comparer la fidélité d'une ou de méthode(s) de polymérisation d'acides nucléiques.

23. Kit destinée à être mis en oeuvre dans le procédé pour déterminer la charge mutationnelle d'une banque obtenue à l'issue de la mutagenèse aléatoire d'un gène d'intérêt selon l'une quelconque des revendications 1 à 17 ou dans une utilisation selon les revendications 18 à 22, où le kit comprend:

- au moins un gène modèle ;
- au moins un abaque permettant de lier la charge mutationnelle d'une banque de gènes obtenue par mutagenèse aléatoire du gène modèle à la fraction de gènes modèles mutés observés dans ladite banque ;
- les réactifs nécessaires à la mise en oeuvre de la technique de mutagenèse aléatoire; et
- une notice d'utilisation indiquant les sous-facteurs de correction définis pour un paramètre sélectionné comme influençant la charge mutationnelle.

24. Kit selon la revendication 23, où le kit comprend des réactifs pour observer des gènes modèle mutants.

25. Kit selon la revendication 23 ou 24, où ledit au moins un gène modèle est sous la forme d'un plasmide.

26. Kit selon l'une quelconque des revendications 23 à 25, où la technique de mutagenèse correspond à la PCR error-prone, et où lesdits réactifs nécessaires à la mise en oeuvre de la technique de

85387-79

63

mutagenèse aléatoire comprennent une ADN polymérase thermostable, dNTP, magnésium, manganèse, et tampon de PCR.

27. Kit destiné à être mis en oeuvre dans une utilisation selon la revendication 18, où le kit comprend en plus des composant du kit selon l'une quelconque des revendications 23 à 26 au moins un abaque permettant de lier la charge mutationnelle (CM) d'une banque de gènes obtenue par mutagenèses aléatoire d'un gène modèle aux conditions expérimentales de mutagenèses aléatoire.

28. Kit selon l'une quelconque des revendications 23 à 27, où la technique de mutagenèse est basée sur une technique d'amplification et où le kit possède des amorces pour l'amplification du gène modèle.

29. Kit selon l'une quelconque des revendications 23 à 28, où les sous-facteurs de correction fournis doivent être complétés par une ou plusieurs des données choisies dans le groupe suivant :

- la longueur du gène d'intérêt,
- la composition en base A du gène d'intérêt,
- la composition en base C du gène d'intérêt,
- la composition en base G du gène d'intérêt,
- la composition en base T du gène d'intérêt,
- le nombre de multiplication de la matrice du gène modèle, et
- le nombre de multiplication de la matrice du gène d'intérêt,

85387-79

64

dont les valeurs correspondantes sont intégrées par un expérimentateur, en fonction du gène d'intérêt et du résultat de mutagenèse aléatoire.

30. Kit selon l'une quelconque des revendications 23 à 29, où le gène modèle est celui qui code pour la Green Fluorescent Protein (GFP).

31. Kit selon la revendication 30, où la mutagenèse est réalisée par PCR error-prone, le gène modèle est celui qui code pour la GFP, et les différents sous-facteurs de correction sont choisis parmi :

- un paramètre intrinsèque 1 représentant la longueur des gènes selon l'équation suivante :

$$C.F.[L] = L(\text{gène d'intérêt}) / 717$$

où L est en nombre de paire de base,

- un paramètre intrinsèque 2 représentant la composition en base sur les brins codant d'ADN selon l'équation suivante :

$$C.F.[\text{base}] = \left( \frac{0,506 \times A_{(g.\text{intérêt})}}{0,34} + \frac{0,346 \times T_{(g.\text{intérêt})}}{0,26} + \frac{0,063 \times C_{(g.\text{intérêt})}}{0,197} + \frac{0,085 \times G_{(g.\text{intérêt})}}{0,203} \right)$$

où  $A_{(g.\text{intérêt})}$ ,  $T_{(g.\text{intérêt})}$ ,  $C_{(g.\text{intérêt})}$  et  $G_{(g.\text{intérêt})}$  sont en fraction,

- un paramètre expérimental représentant un taux de multiplication selon l'équation suivante :

$$C.F.[d] = d(\text{gène d'intérêt}) / d(\text{gfp})$$

85387-79

65

où  $d$  est le taux de multiplication calculé selon l'équation suivante :

$$d = \ln(Q_f/Q_i)/\ln 2$$

dans laquelle,  $Q_i$  et  $Q_f$  représentent respectivement les quantités molaires de la matrice initiale et des produits à l'issue de la PCR.

32. Kit selon l'une quelconque des revendications 23 à 31, possédant plusieurs gènes modèles et plusieurs abaques pour chaque gène modèle de manière à choisir le gène modèle ayant des caractéristiques intrinsèques les plus proches du gène d'intérêt.

33. Kit selon l'une quelconque des revendications 23 à 32 où les sous-facteurs de correction sont enregistrés dans un fichier numérique sauvegardé sur un support.

34. Kit selon la revendication 33, où le support est un CD-Rom ou une disquette.

35. Kit selon l'une quelconque des revendications 23 à 34, où l'abaque est enregistré sous forme d'une base de données numériques sauvegardées sur un support approprié.

36. Procédé selon l'une quelconque des revendications 1 à 17, où l'une au moins des étapes (a), (c) et (d) est mise en oeuvre par des moyens informatiques.

37. Procédé selon la revendication 36, où lesdits moyens informatiques exploitent des données numériques relatives à l'une au moins des caractéristiques suivantes :

- le gène modèle sélectionné,

85387-79

66

- la technique de mutagenèse choisie,

- le ou les paramètres sélectionnés comme capables d'influencer la charge mutationnelle d'une banque de gènes mutés par rapport à une autre pour définir les sous-facteurs de correction.

38. Procédé selon la revendication 36 ou 37, où lesdits moyens informatiques exploitent des données numériques relatives à l'une au moins des caractéristiques suivantes :

- la longueur du gène d'intérêt,
- la composition en base A du gène d'intérêt,
- la composition en base C du gène d'intérêt,
- la composition en base G du gène d'intérêt,
- la composition en base T du gène d'intérêt,
- le nombre de multiplication de la matrice du gène modèle, et
- le nombre de multiplication de la matrice du gène d'intérêt,

dont les valeurs correspondantes sont intégrées par un expérimentateur, en fonction du gène d'intérêt et du résultat de mutagenèse aléatoire.

39. Procédé selon l'une quelconque des revendications 36 à 38, où lesdits moyens informatiques sont utilisés sur une calculatrice préprogrammée ou un micro-ordinateur.

40. Procédé selon l'une quelconque des revendications 36 à 39, où lesdits moyens informatiques permettent d'intégrer les données

85387-79

67

liées à l'abaque du gène modèle, et aux sous-facteurs de correction.

41. Un dispositif destiné à être mis en oeuvre dans le procédé pour déterminer la charge mutationnelle d'une banque obtenue à l'issue de la mutagenèse aléatoire d'un gène d'intérêt selon l'une quelconque des revendications 1 à 17 et 36 à 40, ou dans l'utilisation selon l'une quelconque des revendications 18 à 22, où le dispositif comprend le kit selon l'une quelconque des revendications 23 à 35 et une calculatrice préprogrammée ou un micro-ordinateur.

Figure 1

1    tttcaagagtgccatgcccgaggggttatgtacaggaaagaactatatatttttcaaagatga    60  
61    cgggaactacaagacacgtgctgaagtcaagtttgaagggtgatacccttggttaatagaat    120  
121   cgagttaaaagggtggttgatttttaaagaagatggaaacattcttggacacaaattggaata    180  
181   caactataactcacacaatgtatacatcatggcagacaaacaaaagaatggaatcaaagc    240  
241   taacttcaaagttagacacaacattgaagatggaagcgttcaactggcagaccattatca    300  
301   acaaaataactccaattggcgatggccctgtcctttttaccagacaaccattacctgtccac    360  
361   acaatctgcccttttcgaaagatcccaacgaaaagagagaccacatggtccttcttgagtt    420  
421   tgtaacagctgctgggattacacatggcatggatgaactatacaaataaatgagtaaagg    480  
481   agaagaacttttccactagagttgtcccaattcttggtgaattagacgggtgatgttaatgg    540  
541   gcacaaattttctgtcagtggagaggggagaagggtgatgcaacatacggaaaacttaccct    600  
601   taaattttatattgcactactggaaaactaccagttccgtggccaacacttgtcactactct    660  
661   ctcttatgggtgttcaatgctttttcgagatacccagatcacatgaaacggcatgactt    717

Figure 2

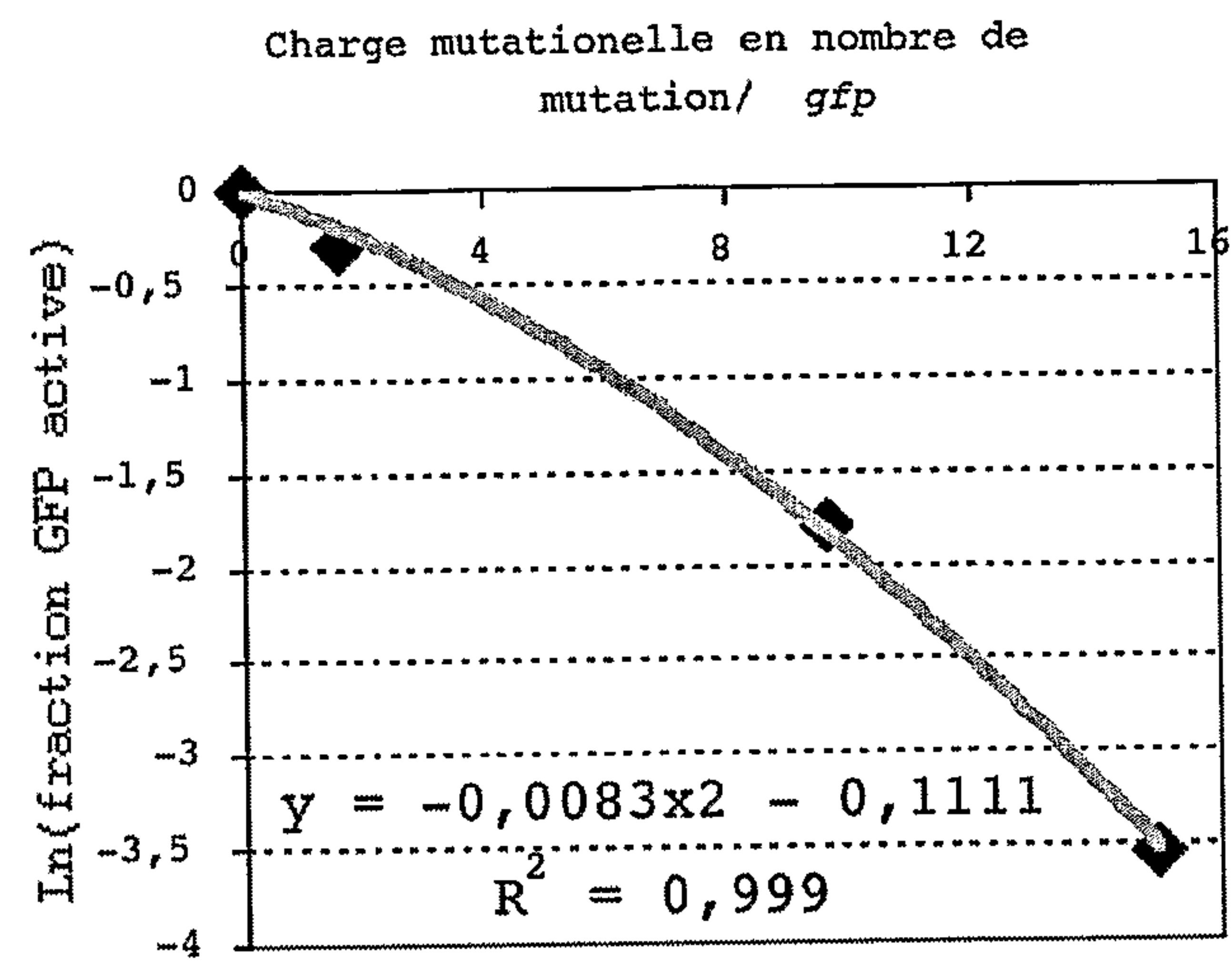
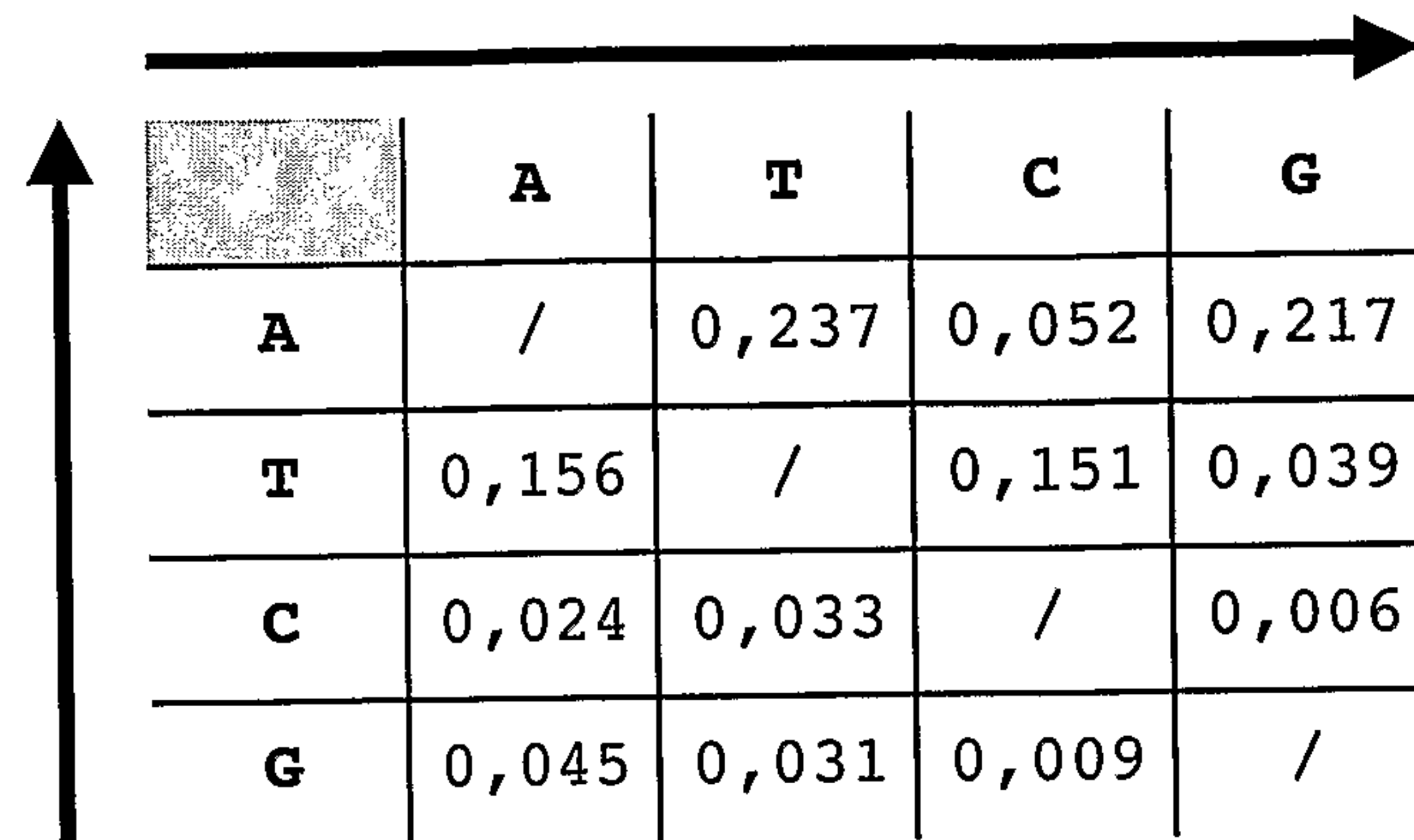


Figure 3



|   | A     | T     | C     | G     |
|---|-------|-------|-------|-------|
| A | /     | 0,237 | 0,052 | 0,217 |
| T | 0,156 | /     | 0,151 | 0,039 |
| C | 0,024 | 0,033 | /     | 0,006 |
| G | 0,045 | 0,031 | 0,009 | /     |

Figure 4

Matrice cible *gfp*  
pour la PCRep

|          |       |
|----------|-------|
| <b>A</b> | 0,506 |
| <b>T</b> | 0,346 |
| <b>C</b> | 0,063 |
| <b>G</b> | 0,085 |

Figure 5

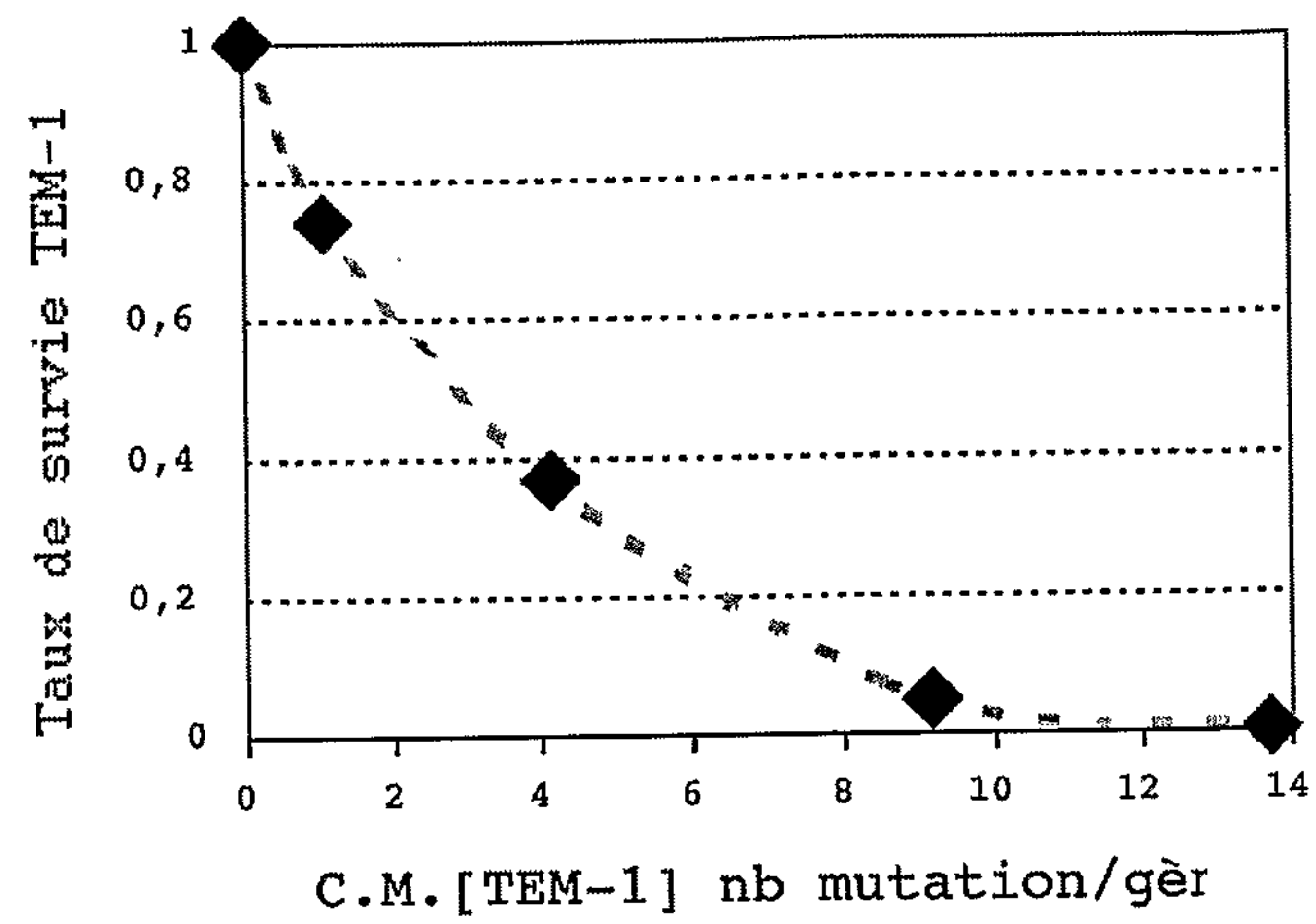


Figure 6

