

(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  
29 décembre 2004 (29.12.2004)

PCT

(10) Numéro de publication internationale  
**WO 2004/113539 A2**

- (51) Classification internationale des brevets<sup>7</sup> : C12N 15/62 (74) Mandataires : BREESE, Pierre etc.; Breesé-Majerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).
- (21) Numéro de la demande internationale : PCT/FR2004/001538 (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, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (22) Date de dépôt international : 18 juin 2004 (18.06.2004)
- (25) Langue de dépôt : français
- (26) Langue de publication : français
- (30) Données relatives à la priorité : 03/07411 19 juin 2003 (19.06.2003) FR (84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, 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).
- (71) Déposants (pour tous les États désignés sauf US) : INSERM [FR/FR]; 101, rue de Tolbiac, F-75654 Paris Cedex 13 (FR). CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE-CNRS [FR/FR]; 3, rue Michel-Ange, F-75794 Paris Cedex 16 (FR).
- (72) Inventeurs; et
- (75) Inventeurs/Déposants (pour US seulement) : MOUIL-LAC, Bernard [FR/FR]; 77, rue des jardins d'Omevades, F-34080 Montpellier (FR). SEN, Tuhnadri [IN/IN]; 711/B Block P, New Aipore 700 053, Kolkata (IN). BANERES, Jean-Louis [FR/FR]; 82, Impasse Bâton Rouge, F-34070 Montpellier (FR).
- Publiée :  
— sans rapport de recherche internationale, sera republiée dès réception de ce rapport
- En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(54) Title: METHOD FOR PRODUCING A RECOMBINANT PROTEIN OF INTEREST AND PROTEIN THUS PRODUCED

(54) Titre : PROCEDE DE PRODUCTION D'UNE PROTEINE RECOMBINANTE D'INTERET ET PROTEINE PRODUITE

(57) Abstract: The invention relates to a novel method for producing large quantities of a protein of interest, that can be directly used for structural analyses. The invention also relates to the recombinant protein obtained, the nucleotide sequence coding for the recombinant protein, vectors containing the nucleotide sequence, and cells containing said sequence or said vector.

(57) Abrégé : L'invention est relative à un nouveau procédé de production de protéine d'intérêt, en grande quantité, directement utilisable pour des analyses structurales. L'invention concerne également la protéine recombinante obtenue, la séquence en nucléotides codant pour la protéine recombinante, des vecteurs contenant la séquence en nucléotides, des cellules contenant ladite séquence ou ledit vecteur.



WO 2004/113539 A2

**Procédé de production d'une protéine recombinante  
d'intérêt et protéine produite.**

L'invention a pour objet un nouveau procédé de  
5 production de protéine d'intérêt, en grande quantité,  
directement utilisable pour des analyses structurales.  
L'invention concerne également la protéine recombinante  
obtenue.

Les récepteurs couplés aux protéines G (RCPGs)  
10 constituent une super-famille de protéines membranaires  
caractérisées par 7 domaines transmembranaires (TM I à VII)  
qui jouent un rôle primordial dans la communication  
intercellulaire et la réception des signaux sensoriels [1].

Avec plusieurs centaines de membres identifiés, les  
15 RCPGs constituent la famille structurale et fonctionnelle  
de récepteurs membranaires la plus importante. Ils  
représentent notamment une part significative du génome  
humain connu à ce jour (au moins 700 récepteurs, 0.5 % du  
génom). .

20 Exprimés à la surface de toutes les cellules d'un  
organisme (de la levure à l'homme), ils sont activés par  
une très grande variété de messages extracellulaires  
(peptides, hormones, lipides, molécules odorantes, lumière,  
nucléotides, nucléosides, molécules du goût, etc...). Leur  
25 activation entraîne une cascade intracellulaire de signaux  
par l'intermédiaire de protéines G et résulte en un grand  
nombre de réponses cellulaires (par exemple division ou  
contraction cellulaire, neurotransmission).

De façon générale, les RCPGs sont impliqués dans  
30 chaque fonction physiologique. L'importance de ces  
récepteurs ainsi que la connaissance de leur localisation  
cellulaire les désigne comme des cibles idéales de la  
thérapeutique. Et effectivement, on peut estimer que près  
de 50% des médicaments sur le marché agissent via les

RCPGs. Beaucoup de pathologies sont la conséquence de mutations des RCPGs, et leurs manifestations cliniques sont bien connues, on peut citer par exemple cécité, diabète insipide néphrogénique, hypo- ou hyperthyroïdisme, puberté  
5 précoce, obésité [2].

La découverte que certains récepteurs des chémokines sont des cofacteurs de l'infection par le virus HIV renforce l'idée que les RCPGs sont impliqués dans une grande variété de situations pathologiques [3].

10 Ces considérations générales indiquent clairement la nécessité d'étudier l'architecture fonctionnelle de ces récepteurs, afin de mieux comprendre le processus de transduction du signal et la dynamique de leurs interactions avec diverses molécules (ligands ou  
15 partenaires intracellulaires), ainsi que pour développer de nouveaux outils pharmacologiques et thérapeutiques. Cependant, l'étude de l'architecture fonctionnelle des RCPGs par les méthodes expérimentales « directes » (cristallographie aux rayons X, RMN, spectrométrie de  
20 masse) reste encore très limitée. Une seule structure tridimensionnelle (3D) est actuellement connue, celle de la rhodopsine de bœuf [4], du fait du très fort niveau d'expression naturel de ce récepteur dans la rétine. La connaissance de leur architecture fonctionnelle est donc  
25 abordée actuellement par un faisceau de méthodes théoriques (modélisation), physico-chimiques (photomarquage, fluorescence) et biologiques (mutagénèse dirigée, pharmacologie moléculaire, knock-out, etc...).

Ces études représentent un enjeu industriel et socio-  
30 économique de première importance, compte tenu des applications thérapeutiques potentielles.

Cependant l'étude structurale et fonctionnelle des RCPGs est très difficile pour diverses raisons :

- la nature transmembranaire de ces protéines et leur caractère hydrophobe rendent leur manipulation délicate et conduisent en général à une perte de la fonctionnalité et à une dénaturation après solubilisation ;

5       - les obtenir dans la totalité de leur séquence primaire reste très difficile. La plupart du temps, ils sont exprimés sous forme tronquée [5].

10       - ils sont exprimés en quantité très faible (0.01% des protéines membranaires) ce qui constitue un blocage à leur purification en grande quantité ;

15       - leur poids moléculaire est élevé (supérieur à 40 kD), et ils sont caractérisés par la présence de modifications post-traductionnelles (glycosylation, palmitoylation, phosphorylation) et de traits structuraux particuliers (ponts disulfure) ;

20       - ce sont des protéines multifonctionnelles possédant des domaines avec des rôles différents : liaison des ligands, activation des protéines G, sites allostériques, zones impliquées dans leur régulation / désensibilisation

20       On comprend aisément que l'étape critique qui constitue actuellement un réel blocage est certainement celle de l'obtention des RCPGs en des quantités compatibles avec les approches de biologie structurale « directes ».

25       Jusqu'à présent, aucune stratégie de production en grande quantité, généralisable à tous les RCPGs, permettant en outre leur purification sous forme fonctionnelle d'une manière simple, n'a été mise au point. Ponctuellement, certains récepteurs ont été produits en quantité élevée (mg/l de culture) [6-8], mais les méthodes mises en place  
30       ne s'appliquent pas à une majorité de RCPGs.

Les RCPGs ne représentent dans ce domaine de la production en grande quantité d'une protéine qu'un exemple des difficultés que l'on rencontre lorsqu'il s'agit d'obtenir une quantité importante d'une protéine d'intérêt.

La présente invention a pour but de fournir un procédé de production en grande quantité d'une protéine d'intérêt, particulièrement des RCPGs.

Les inventeurs ont de manière surprenante, montré que  
5 la construction de protéines recombinantes, particulièrement de protéines membranaires, très particulièrement de RCPGs, comprenant au moins un fragment d'une intégrine alpha et la protéine d'intérêt, permet d'obtenir des protéines recombinantes pouvant être  
10 exprimées en grande quantité. Cette stratégie permet en particulier d'obtenir une production desdites protéine en grande quantité dans des microorganismes, particulièrement dans des bactéries. Lorsque les protéines recombinantes de l'invention sont produites dans des bactéries, elle  
15 s'accumulent dans les corps d'inclusion du cytoplasme bactérien. Il est alors nécessaire de renaturer les protéines d'intérêt pour les obtenir sous forme active en une quantité compatible avec une analyse directe de leur structure par exemple par cristallographie aux rayons X ou  
20 par résonance magnétique nucléaire (RMN). Le procédé de l'invention peut permettre en outre la production de protéines non tronquées, particulièrement lorsqu'il est appliqué aux RCPGs.

Les intégrines forment une famille de récepteurs liés  
25 structurellement et fonctionnellement qui participent aux interactions cellule-cellule et cellule-matrice extra-cellulaire. Toutes les intégrines se présentent sous forme d'hétérodimères de sous-unités alpha et bêta, liées de manière non covalente. Sur la base de leur séquence  
30 primaire, toutes les intégrines alpha présentent une région N-terminale constituée de sept séquences en acides aminés répétées (répétition I à VII), chacune comprenant approximativement 60 acides aminés. Certaines sous-unités alpha incluent un domaine d'insertion (domaine-I) d'environ

200 acides aminés, situé entre les répétitions II et III. Les homologues entre les répétitions I et VII comprennent essentiellement des séquences consensus FG et GAP, correspondant aux enchaînements phenylalanine, glycyl-  
5 glycyl, alanyl, prolyl, d'où leur dénomination « répétition FG-GAP ».

A la connaissance des inventeurs, la sous-unité alpha des intégrines (par ailleurs appelée intégrine alpha (intégrine  $\alpha$ ) dans le texte) n'a jamais été utilisée pour  
10 la production de protéines recombinantes d'intérêt dans des cellules autres que des cellules de mammifères et ce en quantité directement compatible avec une analyse structurale de la protéine d'intérêt, ce qui nécessite une quantité de ladite protéine pouvant aller jusqu'à plusieurs  
15 milligrammes.

La présente invention vise à satisfaire cette exigence.

Ainsi, l'invention a pour objet premier l'utilisation d'au moins un fragment d'une intégrine alpha dans la  
20 construction d'au moins une protéine recombinante d'intérêt. L'invention a aussi pour objet l'utilisation d'au moins un fragment d'une intégrine alpha pour la production d'au moins une protéine recombinante d'intérêt.

Dans le présent texte, l'expression "protéine recombinante" ou "protéine recombinante d'intérêt" se  
25 rapporte à la protéine recombinante produite selon l'invention. Cette protéine recombinante peut en particulier comprendre l'enchaînement de plusieurs (au moins deux) protéine d'intérêt, fusionnées, éventuellement  
30 séparées par des séquences espaceurs et/ou des séquences de clivage.

L'expression "protéine d'intérêt" se rapporte à la séquence peptidique correspondant à une protéine d'intérêt que l'on veut produire (ou que l'on a produite).

Ainsi, l'on comprend qu'une "protéine recombinante" est constituée d'une ou plusieurs "protéine d'intérêt" éventuellement séparées par des séquences espaceurs et/ou des séquences de clivage.

5 Par fragment d'une intégrine alpha, on entend aussi bien la séquence en acide aminé complète de l'intégrine alpha utilisée qu'une séquence partielle. La séquence de l'intégrine alpha utilisée peut être native ou mutée. Préférentiellement selon l'invention, la séquence utilisée  
10 est une séquence comprenant l'extrémité N-terminale de l'intégrine alpha utilisée, encore plus préférentiellement une séquence correspondant à l'extrémité N-terminale de l'intégrine alpha utilisée.

Selon un mode particulier de l'invention, le fragment  
15 de l'intégrine alpha utilisée comprend au moins les modules FG-GAP IV à VII et une portion du module FG-GAP III de l'intégrine alpha utilisée.

Le fragment de l'intégrine alpha utilisée peut être un fragment de toute intégrine alpha connue. On citera  
20 particulièrement les intégrines  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 4$ ,  $\alpha 5$ ,  $\alpha 6$ ,  $\alpha 7$ ,  $\alpha 8$ ,  $\alpha 9$ ,  $\alpha 10$ ,  $\alpha 11$ ,  $\alpha D$ ,  $\alpha E$ ,  $\alpha L$ ,  $\alpha M$ ,  $\alpha X$ ,  $\alpha IIb$  ou encore  $\alpha V$ .

Selon un mode de réalisation particulier de l'invention, on utilise un fragment de 287 acides aminés,  
25 correspondant à la partie de l'extrémité N-terminale de l'intégrine alpha 5 qui s'étend entre les positions 231 et 517, selon la numérotation tenant compte de la présence du peptide signal. Si l'on ne tient pas compte du peptide signal, le fragment utilisable dans l'invention s'étend des  
30 positions 190 (résidu G) à 476 (résidu G) de l'intégrine alpha5.

Lorsque l'on utilise d'autres intégrines alpha, les fragments utilisables selon l'invention sont les fragments homologues aux fragments ci-dessus définis. Par exemple

quand il s'agit de l'intégrine  $\alpha V$ , le fragment utilisable selon l'invention correspond à la partie de l'extrémité N-terminale de l'intégrine  $\alpha V$  qui s'étend des positions 211 (résidu G) à 495 (résidu G) selon la numérotation tenant  
5 compte de la présence du peptide signal. Si l'on ne tient pas compte du peptide signal, le fragment utilisable dans l'invention s'étend des positions 181 (résidu G) à 465 (résidu G) de l'intégrine  $\alpha V$ . Quand il s'agit de l'intégrine  $\alpha IIb$ , le fragment utilisable selon l'invention  
10 correspond à la partie de l'extrémité N-terminale de l'intégrine  $\alpha IIb$  qui s'étend des positions 224 (résidu G) à 508 (résidu Q), selon la numérotation tenant compte de la présence du peptide signal. Si l'on ne tient pas compte du peptide signal, le fragment utilisable dans l'invention  
15 s'étend des positions 193 (résidu G) à 477 (résidu Q) de l'intégrine  $\alpha IIb$ .

Selon un mode particulier de l'invention, le fragment de l'intégrine alpha utilisée comprend au moins une séquence en acides aminés choisie parmi les séquences SEQ  
20 ID n°1 (fragment de l'intégrine  $\alpha 5$  humaine), SEQ ID n°2 (fragment de l'intégrine V humaine) et SEQ ID N°3 (fragment de l'intégrine  $\alpha IIb$  humaine), de la liste de séquence fournie en annexe.

Selon un autre mode particulier de l'invention, le  
25 fragment de l'intégrine alpha utilisé comprend au moins une séquence en acides aminés codée par l'une des séquences en nucléotides, choisie parmi les séquences SEQ ID N°4 (fragment de l'intégrine  $\alpha 5$  humaine), SEQ ID N°5 (fragment de l'intégrine V humaine) et SEQ ID N°6 (fragment de  
30 l'intégrine  $\alpha IIb$  humaine), de la liste de séquence fournie en annexe.

Selon encore un autre mode de réalisation particulier de l'invention, il est possible que le fragment de l'intégrine alpha soit utilisé dans la construction de

plusieurs (au moins deux) protéines d'intérêt recombinantes. Dans ce cas, les protéines recombinantes seront fusionnées lors de la traduction. Cela peut s'avérer nécessaire dans le cas d'une protéine d'intérêt pour laquelle la construction selon l'invention ne permet pas sa production directe (protéine réfractaire). Il est alors nécessaire de coupler en tandem la séquence de ladite protéine réfractaire à une protéine recombinante d'intérêt que les constructions selon l'invention permettent de produire. Ainsi, selon cette forme particulière de réalisation de l'invention, la construction selon l'invention comprendra au moins un fragment d'ADN codant au moins un fragment d'une intégrine alpha, puis au moins un ADN codant au moins une première protéine recombinante d'intérêt et au moins un ADN codant au moins une seconde protéine recombinante d'intérêt. Selon ce mode de réalisation particulier de l'invention, l'ADN codant la seconde protéine d'intérêt sera inséré dans la construction en phase en aval de la séquence d'ADN codant la première protéine d'intérêt. Ce mode de réalisation particulier peut être combiné à l'un quelconque des modes de réalisations particuliers décrits précédemment.

Préférentiellement selon l'invention, le fragment d'intégrine alpha est situé dans la protéine recombinante d'intérêt préparée selon l'invention, en amont de la séquence de la protéine d'intérêt (ou des séquences des protéines d'intérêt) que l'on vise à produire, c'est-à-dire du côté N-terminal de la protéine recombinante d'intérêt (ou des protéines recombinantes d'intérêt) que l'on cherche à construire et/ou à produire.

L'invention a également pour objet une protéine recombinante caractérisée en ce qu'elle comprend, fusionnés, au moins un fragment d'une intégrine alpha tel que défini précédemment et au moins une protéine d'intérêt.

La (ou les) protéine(s) d'intérêt, partie(s) de la protéine recombinante de l'invention, peut être toute protéine que l'on cherche à produire, particulièrement une protéine membranaire, très particulièrement un récepteur  
5 couplé aux protéines G (RCPGs). A titre d'exemple pour ces derniers, on peut citer les récepteurs de la vasopressine et de l'ocytocine (V1a, V2, OTR), les récepteurs des leucotriènes (BLT1, BLT2, CysLT1, CysLT2), les récepteurs adrénergiques (bêta 3), les récepteurs canabinoïdes (CB1),  
10 les récepteurs des chimiokines (CCR5, CXCR4) le récepteur AT1 de l'angiotensine II, le récepteur B2 de la bradykinine.

La protéine recombinante de l'invention (quel que soit le mode de réalisation de l'invention) peut en outre  
15 comprendre toute séquence en acides aminés qui permet de la purifier de manière aisée. Ainsi selon un mode particulier de l'invention, la protéine recombinante peut comprendre une séquence de 6 résidus histidine (tag 6xHIS). Ce tag 6xHIS peut être incorporé dans la séquence de la protéine  
20 en vue de sa purification sur colonne de Ni-NTA (Nickel-nitrilotriacetic acid) agarose. Préférentiellement, cette séquence est à l'extrémité C-terminale de la protéine recombinante de l'invention. Lorsque la protéine recombinante selon l'invention est constituée d'au moins  
25 deux protéines fusionnées, le tag 6xHIS est préférentiellement situé en aval de la dernière des protéines d'intérêt que l'on cherche à produire.

La séquence codant la protéine recombinante peut en outre comprendre au moins une séquence codant au moins un  
30 site de clivage par une endoprotéase.

De manière avantageuse, la séquence codant pour les derniers résidus de l'intégrine peut être mutée pour constituer un site de clivage pour une endoprotéase (facteur Xa, thrombine), qui permettra après expression et

purification de la protéine recombinante, de séparer la protéine d'intérêt de son partenaire de fusion. Selon une forme de réalisation particulière de l'invention, le résidu L (position 285) peut être modifié par mutation en un  
5 résidu I, les résidus E et G (positions 286 et 287) étant conservés. Un résidu R supplémentaire peut être introduit par mutagenèse. L'enchaînement ainsi constitué (IEGR) correspond au site de clivage par le facteur Xa qui coupe la protéine après le résidu R.

10        Sous une autre forme de réalisation, le site de clivage au facteur Xa peut être transformé en un site de clivage à la thrombine. Pour cela les résidus I, E, et G peuvent être remplacés par des résidus L, V et P. Le résidu R est conservé afin d'obtenir l'enchaînement LVPR. Comme le  
15 fragment intégrine a été incorporé dans le vecteur côté 3' par un site BamHI (séquence ggatcc), on obtient donc la séquence ggatcc codant pour deux résidus G et S juste après LVPR. L'enchaînement LVPRGS constitue le site de clivage à la thrombine, qui coupe la protéine après le résidu R.

20        On comprend donc que dans une forme plus élaborée, la protéine recombinante de l'invention comprend, de son extrémité N-Terminale à son extrémité C-Terminale, le fragment d'intégrine alpha comportant le site de clivage par une endoprotéase, la (ou les) protéine(s) d'intérêt et  
25 le tag 6xHIS.

       Sous une forme particulière, la protéine recombinante de l'invention comprend, de son extrémité N-Terminale à son extrémité C-Terminale, le fragment d'intégrine alpha comportant le site de clivage par le facteur Xa, la (ou  
30 les) protéine(s) d'intérêt et le tag 6xHIS.

       Sous une autre forme particulière, la protéine recombinante de l'invention comprend, de son extrémité N-Terminale à son extrémité C-Terminale, le fragment d'intégrine alpha comportant le site de clivage par la

thrombine, la (ou les) protéine(s) d'intérêt et le tag 6xHIS.

Il peut encore être nécessaire, lorsque la protéine recombinante selon l'invention comporte plus d'une protéine d'intérêt fusionnées entre-elles, que lesdites protéines d'intérêt puissent être séparées après leur synthèse, par exemple avant purification. Ainsi, il est possible d'insérer entre les différentes séquences d'ADN codant les différentes protéines d'intérêt, au moins une séquence d'ADN codant un site de clivage pour une endoprotéase. Il est possible que des sites de clivage pour des endonucléases différentes soient insérés dans une même protéine recombinante.

Il peut être nécessaire de rendre le clivage de la protéine recombinante encore plus efficace. A cet égard il est possible d'insérer dans la construction selon l'invention une séquence codant une séquence peptidique servant de bras espaceur, préférentiellement situé en amont du site de clivage par une endoprotéase.

Ainsi selon un mode de réalisation particulier de l'invention, la protéine recombinante comprend en outre une séquence peptidique servant de bras espaceur, préférentiellement situé en amont du site de clivage par une endoprotéase.

Dès lors, dans une forme encore plus élaborée, la protéine recombinante de l'invention comprend, de son extrémité N-Terminale à son extrémité C-Terminale, le fragment d'intégrine alpha comportant un bras espaceur et le site de clivage par une endoprotéase, la (ou les) protéine(s) d'intérêt et le tag 6xHIS.

Sous une forme particulière, la protéine recombinante de l'invention comprend, de son extrémité N-Terminale à son extrémité C-Terminale, le fragment d'intégrine alpha comportant un bras espaceur, le site de clivage par le



nucléotides codant pour une protéine recombinante d'intérêt, telle que définie précédemment.

L'invention a aussi pour objet l'utilisation d'au moins un fragment d'une séquence en nucléotides codant pour  
5 au moins un fragment d'une intégrine alpha tel que défini précédemment, pour la production d'une protéine recombinante d'intérêt telle que définie précédemment.

L'invention a en outre pour objet une séquence en nucléotides codant pour une protéine recombinante d'intérêt  
10 comprenant au moins un fragment d'une séquence en nucléotides codant pour au moins un fragment d'une intégrine alpha, tel que défini précédemment, et une séquence en nucléotides codant pour au moins une protéine d'intérêt, telle que définie précédemment.

15 Préférentiellement, la séquence en nucléotides codant pour au moins un fragment d'une intégrine alpha utilisable selon l'invention ou comprise dans la séquence en nucléotides codant pour une protéine recombinante d'intérêt selon l'invention, peut être choisie parmi les séquences en  
20 nucléotides SEQ ID n°4, SEQ ID n°5 et SEQ ID N°6, de la liste de séquence fournie en annexe.

L'invention a encore pour objet un vecteur comprenant une séquence en nucléotides codant pour une protéine recombinante d'intérêt, telle que définie précédemment,  
25 comprenant au moins un fragment d'une séquence en nucléotides codant pour au moins un fragment d'une intégrine alpha et une séquence en nucléotides codant pour au moins une protéine d'intérêt. Le vecteur peut être un vecteur eucaryote tel qu'un plasmide ou un virus. Le  
30 vecteur peut aussi être tout vecteur procaryote tel qu'un plasmide ou un phage.

De préférence, le vecteur est un vecteur d'expression, c'est-à-dire capable de permettre la transcription et la traduction de la séquence en nucléotides qu'il contient.

A titre d'exemple, il est possible de citer les vecteurs de la famille pET vendu par la société Novagen ou ceux de la famille pGEX vendus par la société AmershamBiosciences.

5 L'invention a encore pour objet une cellule dans laquelle une séquence en nucléotides codant pour une protéine recombinante d'intérêt, telle que définie précédemment, a été introduite. Selon un mode particulier de l'invention, la séquence a été introduite sous la forme  
10 d'un vecteur tel que défini précédemment.

Par cellule, on entend ici aussi bien une cellule eucaryote qu'une cellule procaryote, particulièrement une bactérie. Toute bactérie capable de permettre l'expression d'une protéine à partir d'une séquence en nucléotide peut  
15 être utilisée selon l'invention. A titre d'exemple, il est possible de citer toutes les bactéries qui dérivent de BL21, BL21 star, Rosetta, BLR, Origami, Tuner, Novablue, toute disponible commercialement.

L'invention a encore pour objet un procédé de  
20 production d'au moins une protéine d'intérêt, caractérisé en ce que dans une première étape on introduit dans une cellule une séquence en nucléotides codant pour une protéine recombinante d'intérêt telle que définie précédemment, et que dans une deuxième étape on place la  
25 cellule dans des conditions suffisantes pour permettre l'expression de la protéine recombinante d'intérêt.

Le procédé de l'invention peut en outre comprendre une étape supplémentaire au cours de laquelle la protéine recombinante d'intérêt peut être coupée par action d'une  
30 endoprotéase, (facteur Xa, thrombine, par exemple), au site créé dans les derniers résidus de l'intégrine afin de séparer la protéine d'intérêt de son partenaire de fusion.

Le procédé de l'invention peut également comprendre une étape supplémentaire au cours de laquelle la protéine

recombinante d'intérêt, ou la (ou les) protéines d'intérêt séparée(s) de son (de leurs) partenaire(s) de fusion, peut (peuvent) être purifiée(s).

Selon le procédé de l'invention, la séquence en  
5 nucléotides codant pour une protéine recombinante d'intérêt peut-être introduite dans la cellule par toute méthode connue. A titre d'exemple de méthodes utilisables, il est possible de citer pour les cellules procaryotes, le choc thermique ou l'électroporation. Pour les cellules  
10 eucaryotes, on peut citer l'électroporation, la méthode du précipité au phosphate de calcium, l'utilisation de polymères cationiques tels que le DEAE-dextran ou encore toute méthode utilisant des liposomes cationiques ou des dendrimères activés. On peut aussi utiliser des rétrovirus  
15 pour faire du transfert de gènes ainsi que les techniques utilisant des microprojectiles pour délivrer de l'ADN dans des cellules cibles.

De même, toute condition suffisante permettant l'expression de la protéine recombinante d'intérêt, connues  
20 de l'homme du métier, sont utilisables selon le procédé de l'invention.

Enfin, toute méthode de purification de la (ou des) protéine(s), connue de l'homme du métier, peut être utilisée selon le procédé de l'invention. A titre  
25 d'exemple, on peut citer les méthodes de chromatographie d'affinité, de chromatographie d'échange d'ions, de chromatographie d'interaction hydrophobes ou encore de filtration sur tamis moléculaire.

Particulièrement, lorsque la protéine recombinante  
30 d'intérêt comprend le tag 6xHIS, la purification sur colonne Nickel-acide nitrilotriacétique agarose (Ni-NTA) représente une méthode de purification particulièrement satisfaisante dans le cadre du procédé de l'invention.

Les techniques utilisables selon l'invention sont connues de l'homme du métier. Celui-ci peut se référer aux nombreux manuels disponibles et en particulier à "Molecular Cloning, a laboratory manual. 2nd édition, Sambrook, 5 Fritsch, Maniatis eds., CSH laboratory press, (1989)".

Outre les dispositions qui précèdent, l'Invention comprend encore d'autres dispositions qui ressortiront de la description qui va suivre, qui se réfère à des exemples de mise en œuvre de l'Invention ainsi qu'aux dessins 10 annexés, dans lesquels :

La Figure 1 représente une construction correspondant à un vecteur selon l'invention.

La Figure 2 représente la production de la protéine de fusion intégrine  $\alpha 5$ -récepteur V2 de la vasopressine selon 15 le procédé de l'invention (colonne de gauche : poids moléculaires des protéines de l'échantillon marqueur ; flèche : position de la protéine recombinante intégrine  $\alpha 5$ -récepteur V2 de la vasopressine, NI : protéines d'un échantillon non-induit, 2h, 3h et 4h : protéines d'un 20 échantillon induit après 2h, 3h et 4h d'induction).

La Figure 3 représente la protéine recombinante intégrine  $\alpha 5$ -récepteur V2 de la vasopressine de la figure 2 après purification et migration sur gel d'électrophorèse. (colonne de gauche : poids moléculaires des protéines de 25 l'échantillon marqueur ; flèche : position de la protéine recombinante intégrine  $\alpha 5$ -récepteur V2 de la vasopressine)

La Figure 4 représente le résultat de la purification de la protéine recombinante de fusion intégrine  $\alpha 5$ -récepteur V2-récepteur CXCR4 ( $\alpha 5$ -V2-CXCR4) par 30 chromatographie d'affinité

S6M: surnageant de solubilisation en tampon 6M urée, déposé sur la résine Ni-NTA agarose

FT: échantillon non retenu sur la résine

W: fraction wash contenant 15 mM imidazole

E100: fusion purifiée éluée en tampon contenant 100 mM imidazole

La flèche indique la position de la protéine de fusion

5  $\alpha$ 5-V2-CXCR4.

Les exemples suivants sont illustratifs de l'Invention et ne la limitent aucunement.

**Exemple 1** : Construction d'un vecteur permettant  
10 l'expression dans les bactéries d'une protéine recombinante d'intérêt :

Un ADN complémentaire codant pour la protéine d'intérêt que l'on veut exprimer, est positionné, dans le vecteur pET21a (+) (vendu par la société Novagen) en phase  
15 avec un fragment d'ADN complémentaire de l'intégrine  $\alpha$ 5, par l'utilisation de sites de restriction appropriés. Le fragment de l'intégrine  $\alpha$ 5 est délimité par les sites NdeI et BamHI. Le site NdeI a l'avantage d'incorporer un codon ATG qui est le codon initiateur de la traduction. Ce codon  
20 initiateur code pour une méthionine (M). Le site NdeI constitué par la séquence CATATG est donc intéressant pour sous-cloner un fragment d'ADN puisque la séquence cible se trouve directement en phase avec le codon ATG. Celui-ci constitue alors le résidu 1. En ce qui concerne l'intégrine  
25 alpha5, l'ATG du site NdeI est positionné en amont de sa séquence nucléotidique. Dans ce cas, l'ATG codera pour une M1 et le G de l'intégrine sera le résidu 2. Le fragment de 287 résidus sera couplé à la méthionine 1 et on obtient donc un partenaire de fusion de 288 résidus : M1-G288.

30 Le vecteur fournit directement la séquence codant pour le tag 6xHIS qui sera localisé à l'extrémité C-terminale de la protéine recombinante d'intérêt. Un site EcoRI est localisé dans le vecteur du côté de l'extrémité N-terminale du site tag. Ainsi l'ADN complémentaire codant pour la

protéine d'intérêt est inséré entre le site Bam HI marquant l'extrémité C-terminale du fragment d'ADN complémentaire de l'intégrine  $\alpha 5$  et le site EcoRI situé du côté de l'extrémité N-terminale du tag 6xHIS.

5 La figure 1 montre le schéma d'une telle construction.

**Exemple 2 :** Expression du récepteur V2 humain de la vasopressine :

10 Construction du vecteur :

L'ADN complémentaire du récepteur V2 humain de la vasopressine (Cotte et al., J. BIOL. Chem. 273, 29462-29468, 1998), est inséré entre les sites BamHI et EcoRI du vecteur obtenu à l'exemple 1.

15 Etape 1 : Préparation de l'ADN complémentaire du récepteur V2 humain de la vasopressine :

Des sites de reconnaissances pour les enzymes de restriction BamHI et EcoRI sont ajoutés de part et d'autre de la séquence de l'ADN complémentaire du récepteur V2  
20 humain de la vasopressine. Cela est réalisé par la technique de PCR classique. L'ADN complémentaire du récepteur V2 humain de la vasopressine est amplifié à partir du vecteur pRK5-V2, (Cotte et al., J. BIOL. Chem. 273, 29462-29468, 1998), à l'aide de deux oligonucléotides  
25 amorces permettant d'insérer les sites de restriction recherchés :

oligo sens (permet l'incorporation du site BamHI) : 5' ATG GGT CGC GGA TCC ATG CTC ATG GCG TCC ACC ACT TCC 3'

oligo antisens (permet l'incorporation du site EcoRI)  
30 : 5' CGA CGG AAT TCT GCG ATG AAG TGT CCT TGG CCA G 3' .

La réaction de PCR est réalisée dans 50 microlitres d'un mélange réactionnel comprenant :

|              |        |
|--------------|--------|
| - pRK5-V2    | 20 ng  |
| - oligo sens | 100 ng |

- oligo antisens 100 ng
- Pfu Turbo polymerase (Stratagene) 2.5 U
- tampon Pfu 10X (Stratagene) 5 µl
- dNTP 80 µM final pour chacun des 4

5 selon les paramètres de cycle suivants :

- dénaturation initiale à 95°C pendant 2 minutes puis
- 25 cycles : 95°C, 30 secondes puis 55°C, 1 minute, 72°C, 1,5 minutes puis
- 10 - élongation finale à 72°C pendant 10 minutes.

La présence du fragment amplifié (fragment PCR V2 amplifié) est contrôlée sur gel agarose 1%.

Etape 2 : Purification du fragment amplifié (fragment PCR V2 amplifié) :

15 La zone du gel d'agarose où l'on visualise le fragment d'ADN amplifié est découpée et l'ADNc est purifié à l'aide du kit de purification Qiaquick gel extraction kit (référence Qiagen 28706), en suivant strictement le protocole préconisé par le fournisseur.

20 Etape 3 : Coupure du fragment PCR V2 amplifié par les enzymes BamHI et EcoRI :

Cela est réalisé en une seule étape à l'aide des enzymes vendues par New England Biolabs (NEB) par une incubation de 3 heures à 37°C, dans un volume final de 50  
25 microlitres contenant :

- insert V2 (fragment amplifié) 100 à 200 ng 24 µl
- tampon EcoRI 10X NEB 5 µl
- sérum albumine bovine NEB 100X (10 mg/ml) 1 µl
- EcoRI ( 40 U ) 2 µl
- 30 - BamHI ( 40 U ) 2 µl
- eau 16 µl

En fin de réaction, les deux enzymes sont inactivées par chauffage à 80°C. pendant 20 minutes.

Le fragment PCR V2 est alors purifié à partir d'un gel d'agarose 1% selon le protocole décrit ci-dessus.

Etape 3 : sous-clonage du fragment PCR V2 amplifié dans les sites BamHI et EcoRI du vecteur pET21a de l'exemple 1 :

La ligation est réalisée par incubation à température ambiante (20-25°C) pendant 4 heures dans un milieu comprenant :

- fragment PCR V2 BamHI / EcoRI  
10 (100 à 200 ng) 8 µl
- vecteur pET21a (30 ng)  
coupé par BamHI / EcoRI 3 µl
- tampon ligase 10X (NEB) 2.5 µl
- T4 DNA ligase (NEB) 2 µl
- 15 - eau 9.5 µl.

Le produit de ligation, protéine de fusion intégrine-récepteur V2 humain de la vasopressine, est ensuite utilisé pour une transformation des bactéries Rosetta (DE3) afin de réaliser les tests d'expression du récepteur.

20 Introduction du vecteur d'expression dans une bactérie et expression de la protéine d'intérêt :

Transformation :

Le vecteur obtenu précédemment est ensuite introduit dans une bactérie de souche Rosetta (DE3) par la technique du choc thermique en suivant le protocole de transformation préconisé par le fournisseur, en l'occurrence Novagen.

25 20 µl de bactéries Rosetta (DE3) référence Novagen 70954-4 et 1 µl de pET21a-intégrine/V2 (quelques nanogrammes) sont incubés sur glace 30 minutes, puis 30 maintenus à 42°C pendant 30 secondes et de nouveau sur glace pendant 2 minutes pour réaliser un choc thermique.

80 µl de milieu SOC (voir composition dans Molecular Cloning, a laboratory manual. 2nd édition, Sambrook, Fritsch, Maniatis eds. CSH laboratory press, 1989) sont

alors ajoutés, puis l'ensemble est incubé à 37°C pendant 1 heure sous agitation à 300 tours/minutes.

Le milieu d'incubation est alors étalé sur boîtes de pétri contenant du milieu LB agar + ampicilline à 100  
5 microgrammes / ml. Les boîtes sont incubées à 37°C 16 heures. Les bactéries d'une colonie sont alors cultivées à 37°C dans 10 ml de milieu LB contenant 100 µg/ml d'ampicilline (ou de son analogue la carbenicilline), et la suspension cellulaire est agitée à 300 tours/minutes.

10 Expression de la protéine :

Lorsque la densité optique de la culture atteint 0.6 U, l'expression de la protéine recombinante est induite par addition d'IPTG 1 mM.

Des échantillons sont prélevés 2, 3 ou 4 heures après  
15 induction. Pour cela, 1 ml de suspension bactérienne, à densité optique de 0,6, est prélevé dans chaque culture. L'échantillon est centrifugé pendant 2 minutes à 12000 tours par minutes. Le surnageant est éliminé et le culot est resuspendu dans 60 µl de tampon de lyse (25 mM Tris,  
20 pH8,3, 185 mM glycine, 0,1% SDS). 60 µl de tampon de dépôt au SDS (10% glycérol, 5% 2-mercaptoethanol, 25 mM Tris-HCl, pH 6,5, 8% SDS, bleu de bromophénol (quelques grains)) sont alors ajoutés et 10 µl de l'échantillon lysé (extraits protéiques totaux) sont alors déposés sur gel  
25 d'acrylamide/bis-acrylamide 12% - SDS 0.1 %. Après migration des protéines, elles sont colorées au bleu de Coomassie selon les techniques habituelles.

La Figure 2 représente les résultats obtenus. Les échantillons induits sont comparés à des contrôles non  
30 induits (NI) mais ayant été cultivés pendant un temps équivalent. Il est évident que la protéine de fusion α5-récepteur V2, qui présente un poids moléculaire apparent autour de 65 kDa, constitue une des protéines majoritaires de la bactérie, ce qui est une condition requise pour une

purification du récepteur en quantité compatible avec des analyses de sa structure via des approches de cristallographie ou RMN.

1 ml de culture ainsi réalisée a permis d'obtenir  
5 environ 3 µg de récepteur de la vasopressine.

**Exemple 3 : Expression d'autres récepteurs :**

Le résultat obtenu à l'exemple 2 a été reproduit avec la même efficacité, pour d'autres RCPGs, tels que le récepteur β3-adrénergique, les récepteurs BLT2, Cys-LT1 et  
10 Cys-LT2 des leucotriènes LTB4, LTD4 et LTC4, le récepteur cannabinoïde de type 1, le récepteur V1a de la vasopressine et le récepteur de l'ocytocine.

**Exemple 4 : Purification de la protéine de fusion fragment d'intégrine α5- récepteur V2 de la vasopressine**  
15 obtenu à l'exemple 2 :

La méthode utilisée est celle décrite par Porath J. et collaborateurs, (Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258, 598-599, 1975)

20 L'exemple exposé ici est un essai qui a permis de purifier 3 mg de protéine fusion à partir d'une culture bactérienne de 100 ml.

Une colonie isolée sur LB agar + Ampicilline (100 µg / ml) est piquée et cultivée dans 10 ml de milieu de culture  
25 LB + carbenicilline (100 µg / ml). La culture est réalisée à 37°C, sous agitation à 300 tours par minute. Lorsque la densité optique de la culture atteint 0,6, la culture est arrêtée et conservée au réfrigérateur (cet échantillon est appelé préculture). Le lendemain, dans un Erlenmeyer de 500  
30 ml, 100 ml de milieu de culture LB + carbenicilline (100 µg / ml) sontensemencés avec 2 ml de préculture et laissé 37°C, à 300 tours par minute jusqu'à ce que la densité optique de la culture ait atteint 0,6. 0,1 mM d'IPTG sont alors ajoutés à la culture afin d'induire l'expression de

la protéine recombinante. La culture est poursuivie environ 3 heures, jusqu'à obtention d'une densité optique de 2,4 (facteur de stimulation de 4).

La culture est alors centrifugée à 4000 tours par minute pendant 5 minutes. Le surnageant est éliminé et le culot peut être lysé directement ou conservé à -80°C.

Pour la lyse, le culot est repris par homogénéisation à la pipette dans 6 ml de Tris-HCl 20 mM, pH 8,00 + inhibiteurs de protéases (leupeptine 5 µg / ml ; benzamidine 10 µg / ml et PMSF 10 µg / ml). Ces trois inhibiteurs de protéases seront incorporés à tous les tampons utilisés par la suite.

Les bactéries sont lysées par sonication à l'aide d'une microsonde conique Branson (duty cycle 50%, output control 5, fréquence 1 burst par seconde pendant 30 secondes, puis repos 30 secondes ; ce cycle est répété 5 fois). Le tube est conservé dans la glace pendant la sonication. Le milieu est alors centrifugé pendant 30 minutes à 15000 tours par minute à 4°C. Le surnageant est conservé pour contrôle sur gel d'électrophorèse.

Le culot contient la protéine d'intérêt puisque celle-ci est accumulée en corps d'inclusion.

Le culot est repris par homogénéisation à la pipette dans 5 ml de Tris-HCl 20 mM, pH 8,00. Les étapes de lyse et de centrifugation sont répétées une fois.

Les surnageants des centrifugations sont conservés pour contrôle sur gel d'électrophorèse.

Le culot est repris par homogénéisation à la pipette dans 5 ml de Tris-HCl, pH 8,00, urée 1M. Un barreau aimanté est placé dans l'échantillon et celui-ci est agité doucement pendant 1h30. Le tube est conservé dans la glace pendant cette étape qui correspond à un lavage des corps d'inclusion et permet d'éliminer des protéines membranaires ou cytoplasmiques associées aux corps d'inclusion mais qui

sont considérées comme contaminantes par rapport à la protéine recombinante.

L'ensemble est alors centrifugé à 15000 tours par minute pendant 30 minutes à 4°C. Le surnageant est conservé  
5 pour contrôle sur gel d'électrophorèse.

Afin de solubiliser les corps d'inclusion et donc la protéine d'intérêt, le culot est repris par homogénéisation à la pipette dans 5 ml de Tris-HCl 20 mM, pH 8,00, urée 6M, SDS 0.2%.

10 Un barreau aimanté est placé dans l'échantillon et celui-ci est agité doucement pendant 3 heures, dans la glace.

La protéine d'intérêt (la protéine de fusion fragment d'intégrine  $\alpha 5$ -récepteur V2 de la vasopressine) est alors  
15 complètement dénaturée.

L'ensemble est alors centrifugé à 15000 tours par minute pendant 30 minutes à 4°C. Le surnageant contient la protéine d'intérêt et constitue l'échantillon qui sera mis en contact avec la résine Ni-NTA (Nickel-nitriloacetic  
20 acid) afin de purifier la fusion alpha5-V2 par chromatographie d'affinité.

3 ml de résine Ni-NTA agarose Superflow (Qiagen, ref 30430) sont équilibrés dans du Tris-HCl 20 mM, pH 8,00, urée 6M, SDS 0,2%, NaCl 150 mM, imidazole 5 mM. On rajoute  
25 dans l'échantillon contenant la protéine d'intérêt une quantité suffisante de NaCl et d'imidazole pour obtenir une concentration finale de 150 mM de NaCl et de 5 mM d'imidazole. L'échantillon et la résine sont mis en contact et on laisse incubé à 4°C pendant 16 heures et sous  
30 agitation douce. Le mélange échantillon/résine est déposé dans une colonne plastique. On laisse Après la décantation, la fraction "flowthrough" est récupérée à un débit faible, pour contrôle sur gel d'électrophorèse.

La résine est alors lavée avec 3 x 9 ml d'une solution de Tris-HCl 20 mM, pH 8,00, urée 6M, SDS 0,2%, NaCl 150 mM, imidazole 20 mM, pour éliminer toutes les protéines non retenues de façon spécifique sur les groupements Nickel.

5 Les éluats de lavage sont conservés pour contrôle sur gel d'électrophorèse.

La protéine d'intérêt est alors détachée de la résine par passage de 3 ml d'une solution de Tris-HCl 20 mM pH 8,00, urée 6M, SDS 0,2%, NaCl 150 mM, imidazole 100 mM. Un

10 aliquot de la protéine purifiée est conservé pour contrôle sur gel d'électrophorèse.

10 µl du milieu contenant la protéine purifiée sont mélangés à 10 µl de tampon de dépôt au SDS et l'ensemble est déposé sur un gel d'électrophorèse.

15 10 µl d'échantillon purifié contiennent de 5 à 10 µg de protéine soit dans 3 ml d'éluat, 1,5 à 3 mg de protéine recombinante.

La figure 3 montre la protéine purifiée déposée sur gel d'électrophorèse.

20 L'échantillon purifié est dialysé contre une solution de Tris-HCl 20 mM, pH 8,00, urée 6M, NaCl 150 mM pour éliminer le SDS et l'imidazole. Pour cela, l'échantillon est placé dans une cassette de dialyse Pierce (membrane de MWCO de 10000) et la dialyse se fait dans un Becher

25 contenant un litre du tampon. La dialyse est effectuée à 4°C pendant au moins 24 heures.

L'échantillon est récupéré et la quantité de protéine d'intérêt obtenue est dosée par mesure d'absorption (excitation à 280 nM, absorption entre 235 et 500 nM). En

30 général, on obtient une D.O de 1 à 1.5, ce qui est équivalent à une concentration de 0.5 à 1 mg / ml, ce qui est équivalent à une concentration de l'ordre de 10 µM.

La protéine purifiée et dénaturée (car solubilisée dans l'urée 6M) est utilisée pour les essais de renaturation.

**Exemple 5 :** construction d'un vecteur permettant l'expression simultanée de deux protéines d'intérêt (ici deux RCPGs) dans *Escherichia coli* et leur purification.

Un ADN complémentaire codant pour une protéine d'intérêt, en l'occurrence ici le récepteur CXCR4 humain des chimiokines, est inséré dans le vecteur pET21a(+)- $\alpha$ 5V2 décrit auparavant dans l'exemple 2. Cet ADN doit être en phase avec celui codant pour la fusion  $\alpha$ 5V2 et est positionné entre les sites de restriction SacI et HindIII par exemple. Le vecteur fournit directement la séquence codant pour le tag 6xHIS qui sera donc localisé à l'extrémité C-terminale du récepteur CXCR4 et permettra donc sa purification dans une étape ultérieure.

Dans l'exemple, une version optimisée (« bactérialisée ») du CXCR4 est insérée dans le vecteur mais la version eucaryote naturelle de ce récepteur (Herzog H, Hort YJ, Shine J and Selbie LA. Molecular cloning, characterization and localization of the human homolog to the reported bovine NPY Y3 receptor : lack of NPY binding and activation. DNA Cell Biol. 12, 465-471, 1993) peut également être utilisée de la même façon.

Etape 1 : Préparation de l'ADN complémentaire du récepteur CXCR4 humain.

Des sites de reconnaissance pour les enzymes de restriction SacI et HindIII sont ajoutés de part et d'autre de la séquence codant pour le récepteur CXCR4 humain au cours d'une réaction PCR classique. L'ADN complémentaire de ce récepteur est amplifié à partir du vecteur pET101/D-TOPO (Invitrogen) dans lequel il est sous-cloné et à partir de deux oligonucléotides amorces permettant d'insérer les sites de restriction considérés.

Oligo sens (incorporation du site SacI) : 5'  
CGAGCTAAGGC GAGCTC A ATGGAAGGCATTAGCATTTATAC 3'

Oligo antisens (incorporation du site HindIII) : 5'  
CGACGGCCCC AAGCTT GCTGCTATGAAAGCTGCTGCTTTC 3'

5 La réaction de PCR est réalisée dans 50 microlitres et est composée de :

|    |                                    |        |
|----|------------------------------------|--------|
|    | pET101/D-TOPO                      | 20 ng  |
|    | Oligo sens                         | 100 ng |
|    | Oligo antisens                     | 100 ng |
| 10 | Pfu Turbo polymerase (Stratagene)  | 2.5 U  |
|    | Tampon Pfu 10 X                    | 5 µl   |
|    | dNTP 80 µM final pour chacun des 4 |        |

Les paramètres de la réaction sont :

dénaturation initiale à 95°C pendant 2 minutes puis 25  
15 cycles : 95°C, 30 s ; 55°C, 1 min ; 72°C, 1.5 min puis  
élongation finale à 72°C, 10 min.

La présence du fragment amplifié est vérifiée sur gel d'agarose 1%.

20 Etape 2 : purification du fragment amplifié

La zone du gel d'agarose dans laquelle on visualise le fragment d'ADN amplifié est découpée et l'ADNc est purifié à l'aide du kit de purification Qiaquick gel extraction (Qiagen référence 28706) en suivant strictement le  
25 protocole préconisé par le fournisseur.

Etape 3 : coupure du fragment PCR CXCR4 amplifié par SacI et HindIII

Cela est réalisé en deux réactions successives à  
30 l'aide d'enzymes de restriction vendues par NEB Biolabs au cours d'incubation à 37°C.

1<sup>ère</sup> réaction pendant 3h:

|                  |        |
|------------------|--------|
| fragment PCR     | 500 ng |
| tampon NEB 1 10X | 5 µl   |

|                             |       |
|-----------------------------|-------|
| sérum albumine bovine (BSA) | 1 µl  |
| SacI                        | 40 U  |
| Eau qsp                     | 50 µl |

Le fragment PCR amplifié et coupé par SacI est purifié  
5 à partir d'un gel agarose selon le protocole décrit à l'étape 2.

2<sup>ème</sup> réaction pendant 3h :

|    |                                                 |       |
|----|-------------------------------------------------|-------|
|    | fragment PCR récupéré de la réaction précédente |       |
| 10 | tampon NEB 2 10X                                | 5 µl  |
|    | BSA                                             | 1 µl  |
|    | HindIII                                         | 40 U  |
|    | Eau qsp                                         | 50 µl |

Le fragment PCR amplifié et coupé par SacI / HindIII  
15 est purifié à partir d'un gel d'agarose selon le protocole décrit à l'étape 2.

Etape 4 : sous-clonage du fragment PCR CXCR4 amplifié  
et coupé SacI / HindIII dans le vecteur pET21a(+)-α5V2.

20 Cette étape est réalisée par ligation à température ambiante pendant 16H00.

|    |                             |        |
|----|-----------------------------|--------|
|    | Fragment PCR coupé          | 200 ng |
|    | Vecteur pET21a(+)-α5V2      |        |
|    | coupé par les mêmes enzymes | 50 ng  |
| 25 | Tampon ligase 10X (NEB)     | 2.5 µl |
|    | T4 DNA ligase (NEB)         | 2 µl   |
|    | Eau qsp                     | 25 µl  |

Le vecteur obtenu qui code pour une triple protéine de  
30 fusion intégrine-récepteur V2-récepteur CXCR4, est ensuite utilisé pour une transformation des bactéries Rosetta (DE3) dans le but d'exprimer cette fusion et de la purifier.

Etape 5 : transformation des bactéries Rosetta.

Suivre à la lettre le protocole décrit dans l'exemple 2.

Etape 6 : expression de la fusion  $\alpha 5V2$ -CXCR4

5 Suivre à la lettre le protocole décrit dans l'exemple 2 mais le milieu de culture LB est remplacé par du milieu Hyperbroth (Athena Enzyme Systems) et le temps optimum d'induction est de 4 heures.

10 Etape 7 : purification de la fusion  $\alpha 5V2$ -CXCR4. Cette étape est illustrée par la figure 4.

15 Suivre à la lettre le protocole de l'exemple 4 mais lors de l'étape de lavage de la résine Ni-NTA agarose, une concentration de 15 mM d'imidazole au lieu de 20 mM est utilisée dans la solution de lavage. L'élution est réalisée à 100 mM comme dans l'exemple 4.

Le bras espaceur peut également être inséré en amont du site de clivage à la thrombine juste après le site EcoRI.

## Références

1. Bockaert J. and Pin JP. Molecular tinkering of G protein-coupled receptors : an evolutionary success. EMBO J. 18, 1723-1729, 1999.
2. Bockaert J. Les récepteurs à sept domaines transmembranaires : physiologie et pathologie de la transduction. Médecine/Sciences 11, 382-394, 1995.
3. Samson M., Libert F., Doranz B.J., Rucker J., Liesnard C., Farber C.M., Saragosti S., Lapoumeroulie C., Cogniaux J., Forceille C., Muyldermans G., Verhofstede C., Burtonboy G., Georges M., Imai T., Rana S., Yi Y., Smyth R.J., Collman R.G., Doms R.W., Vassart G. and Parmentier M. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382, 722-725, 1996.
4. Palczewski K., Kumasaka T., Hori T., Behnke CA., Motoshima H., Fox BA., LeTrong I., Teller DC., Okada T., Stenkamp RE., Yamamoto M. and Miyano M. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 289, 739-745, 2000.
5. Kiefer H., Vogel R. and Maier K. Bacterial expression of G protein-coupled receptors : prediction of expression levels from sequence. Receptors and Channels 7, 109-119, 2000.
6. Tucker J. and Grisshammer R. Purification of rat neurotensin receptor expressed in Escherichia coli. Biochem J., 317, 891-899, 1996.
7. Kiefer H., Krieger J., Olszewski JD., Von Heijne G., Prestwich GD. And Breer H. Expression of an olfactory receptor in Escherichia coli : purification, reconstitution and ligand binding. Biochemistry 35, 16077-16084, 1996.

8. Wei HM. and Grisshammer R. Purification and characterization of the human adenosine A<sub>2a</sub> receptor functionally expressed in escherichia coli. Eur. J. Biochem. 269, 82-92, 2002.

## REVENDICATIONS

1) Utilisation d'au moins un fragment d'une intégrine alpha pour la production dans une cellule, à l'exception  
5 d'une cellule de mammifères, d'au moins une protéine recombinante d'intérêt.

2) Utilisation selon la revendication 1, caractérisée en ce que la cellule est une cellule procaryote,  
10 particulièrement une bactérie.

3) Utilisation selon l'une quelconque des revendications 1 ou 2, caractérisée en ce que le fragment d'intégrine alpha est une séquence en acides aminés  
15 complète de l'intégrine alpha ou une séquence partielle;

4) Utilisation selon l'une quelconque des revendications 1 à 3, caractérisée en ce que le fragment d'intégrine alpha est une séquence comprenant l'extrémité  
20 N-terminale de l'intégrine alpha utilisée.

5) Utilisation selon l'une quelconque des revendications 1 à 4, caractérisée en ce que l'intégrine alpha est native ou mutée.  
25

6) Utilisation selon l'une quelconque des revendications 1 à 5, caractérisée en ce que le fragment d'intégrine alpha comprend au moins les modules FG-GAP IV à VII et une portion du module FG-GAP III de l'intégrine  
30 alpha utilisée.

7) Utilisation selon l'une quelconque des revendications 1 à 6, caractérisée en ce que le fragment d'intégrine alpha provient d'une intégrine alpha choisie

parmi les intégrines  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 4$ ,  $\alpha 5$ ,  $\alpha 6$ ,  $\alpha 7$ ,  $\alpha 8$ ,  $\alpha 9$ ,  $\alpha 10$ ,  $\alpha 11$ ,  $\alpha D$ ,  $\alpha E$ ,  $\alpha L$ ,  $\alpha M$ ,  $\alpha X$ ,  $\alpha I Ib$ , ou encore  $\alpha V$ .

8) Utilisation selon la revendication 7, caractérisée  
5 en ce que le fragment d'intégrine alpha provient d'une  
intégrine alpha choisie parmi les intégrines  $\alpha 5$ ,  $\alpha V$  ou  
 $\alpha I Ib$ .

9) Utilisation selon la revendication 8, caractérisée  
10 en ce que le fragment de l'intégrine alpha 5 s'étend entre  
les positions 231 et 517 (en tenant compte de la présence  
du peptide signal) ou des positions 190 à 476 (en ne tenant  
pas compte de la présence du peptide signal).

15 10) Utilisation selon la revendication 8,  
caractérisée en ce que le fragment de l'intégrine  $\alpha V$   
s'étend entre les positions 211 et 495 (en tenant compte de  
la présence du peptide signal) ou des positions 181 à 465  
(en ne tenant pas compte de la présence du peptide signal).

20

11) Utilisation selon la revendication 8,  
caractérisée en ce que le fragment de l'intégrine  $\alpha I Ib$   
s'étend entre les positions 224 et 508 (en tenant compte de  
la présence du peptide signal) ou des positions 193 à

25

12) Utilisation selon l'une quelconque des  
revendications 1 à 11, caractérisée en ce que le fragment  
d'intégrine alpha comprend au moins une séquence en acides  
aminés choisie parmi les séquences SEQ ID n°1, SEQ ID n°2  
30 et SEQ ID N°3 de la liste de séquence fournie en annexe.

13) Utilisation selon l'une quelconque des  
revendications 1 à 12, caractérisée en ce que le fragment  
d'intégrine alpha comprend au moins une séquence en acides

aminés codée par l'une des séquences en nucléotides choisie parmi les séquences SEQ ID n°4, SEQ ID n°5 et SEQ ID N°6, de la liste de séquence fournie en annexe.

5           14) Utilisation selon l'une quelconque des revendications 1 à 13, caractérisée en ce que le fragment d'intégrine alpha est situé dans la (ou les) protéine(s) recombinante(s) d'intérêt préparée(s) selon l'invention, en amont de la (ou des) séquence(s) de la (ou des) protéine(s)  
10 d'intérêt que l'on vise à produire.

          15) Utilisation selon l'une quelconque des revendications 1 à 14, caractérisée en ce que la (ou les) protéine(s) recombinante(s) comprend (comprennent) au moins  
15 un site de clivage par une endoprotéase.

          16) Utilisation selon l'une quelconque des revendications 1 à 15, caractérisée en ce que la (ou les) protéine(s) recombinante(s) comprend (comprennent) au moins  
20 un bras espaceur.

          17) Utilisation selon la revendication 16, caractérisée en ce que le bras espaceur est constitué par la séquence peptidique SEQ ID N°8 de la liste de séquence  
25 fournie en annexe.

          18) Utilisation selon la revendication 16, caractérisée en ce que la protéine recombinante comprend au moins un bras espaceur codé par la séquence en acide  
30 nucléique SEQ ID N°7 de la liste de séquence fournie en annexe.

          19) Protéine recombinante caractérisée en ce qu'elle comprend au moins un fragment d'une intégrine alpha tel que

décrit dans l'une quelconque des revendications 1 à 14 et au moins une protéine membranaire d'intérêt.

20) Protéine recombinante selon la revendication 19  
5 caractérisée en ce que la protéine d'intérêt est un récepteur couplé aux protéines G.

21) Protéine recombinante selon la revendication 20,  
caractérisée en ce que le un récepteur couplé aux protéines  
10 G est choisi parmi les récepteurs de la vasopressine et de l'ocytocine (V1a, V2, OTR), les récepteurs des leucotriènes (BLT1, BLT2, CysLT1, CysLT2), les récepteurs adrénergiques (bêta.3), les récepteurs canabinoïdes (CB1), les récepteurs des chimiokines (CCR5, CXCR4) le récepteur AT1 de  
15 l'angiotensine II, le récepteur B2 de la bradykinine.

22) Protéine recombinante selon l'une quelconque des revendications 19 à 21, caractérisée en ce qu'elle comprend au moins une séquence de 6 résidus histidine (tag 6xHIS).  
20

23) Protéine recombinante selon la revendication 22, caractérisée en ce que la séquence de 6 résidus histidine est à l'extrémité C-terminale de la protéine.

24) Utilisation d'au moins un fragment d'une séquence en nucléotides codant pour au moins un fragment d'une intégrine alpha tel que décrit dans l'une quelconque des revendications 1 à 14, dans la construction d'une séquence en nucléotides codant pour au moins une protéine  
25  
30 recombinante d'intérêt telle que décrite dans l'une quelconque des revendications 19 à 23.

25) Séquence en nucléotides codant pour au moins une protéine recombinante d'intérêt telle que décrite dans l'une quelconque des revendications 19 à 23.

5        26) Vecteur comprenant une séquence en nucléotides telle que décrite à la revendication 25.

10       27) Cellule, à l'exception d'une cellule de mammifères, dans laquelle une séquence en nucléotides telle que décrite à la revendication 25 ou un vecteur tel que décrit à la revendication 26 a été introduit.

15       28) Procédé de production d'au moins une protéine d'intérêt, caractérisé en ce que dans une première étape on introduit dans une cellule, à l'exception d'une cellule de mammifères, une séquence en nucléotides codant pour au moins une protéine recombinante d'intérêt, telle que décrite à la revendication 25, et que dans une deuxième étape on place la cellule dans des conditions permettant l'expression de la (ou des) protéine(s) recombinante(s) d'intérêt.

25       29) Procédé selon la revendication 28, caractérisé en ce qu'il comporte en outre une étape supplémentaire au cours de laquelle la (ou les) protéine(s) recombinante(s) d'intérêt est (sont) coupée(s) par action d'une endoprotéase.

30       30) Procédé selon l'une quelconque des revendications 28 ou 29, caractérisé en ce qu'il comporte en outre une étape supplémentaire au cours de laquelle la (ou les) protéine(s) recombinante(s) d'intérêt ou la (ou les) protéine(s) d'intérêt séparée(s) de son (leurs) partenaire(s) de fusion, est (sont) purifiée(s).

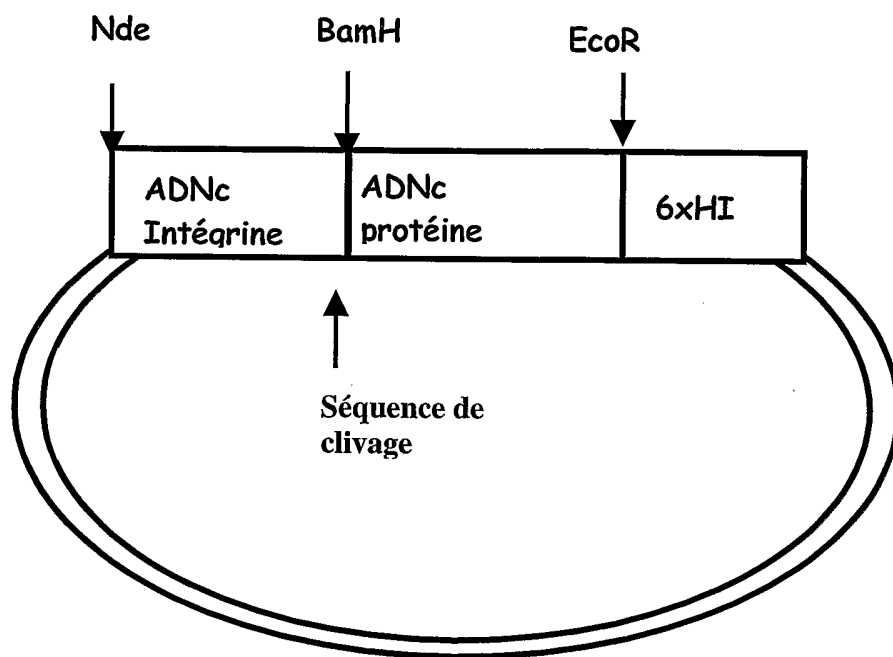


Figure 1

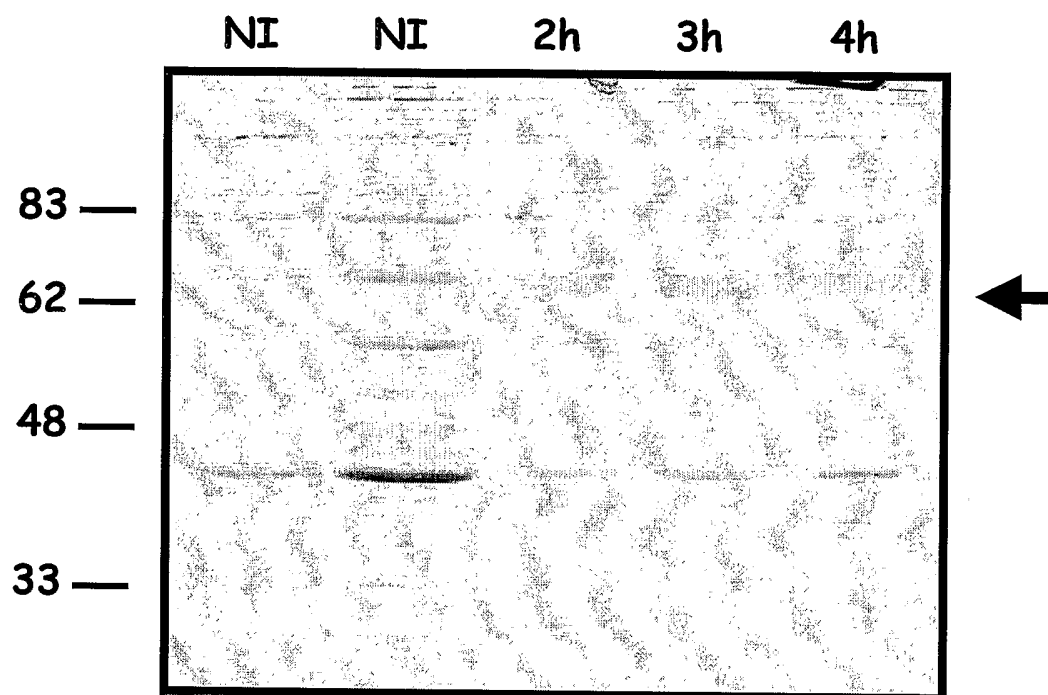


Figure 2

3/4

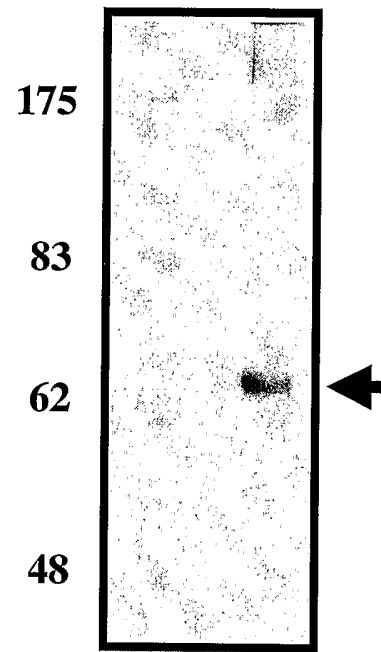


Figure 3

4/4

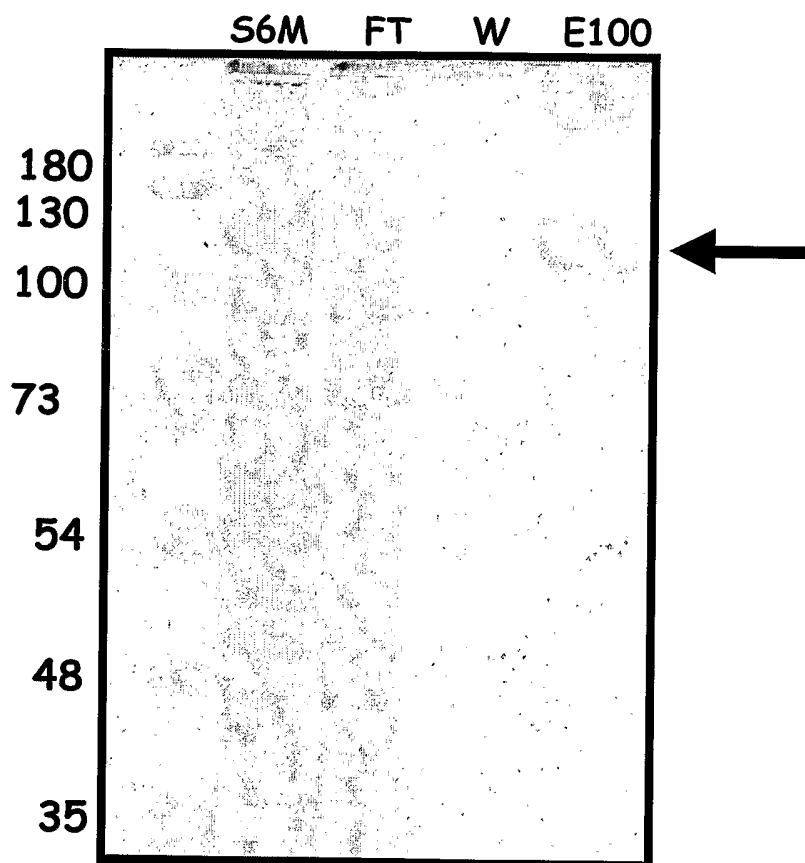


Figure 4

## SEQUENCE LISTING

<110> INSERM  
CNRS

<120> Procédé de production d'une protéine recombinante d'intérêt et protéine produite.

<130> 32412/PCT

<140> PCT/FR04/XXXXXX

<141> 2004-06-18

<150> FR03/07411

<151> 2003-06-19

<160> 6

<170> PatentIn version 3.1

<210> 1

<211> 288

<212> PRT

<213> Homo sapiens

<400> 1

Met Gly Gln Ile Leu Ser Ala Thr Gln Glu Gln Ile Ala Glu Ser Tyr  
1 5 10 15

Tyr Pro Glu Tyr Leu Ile Asn Leu Val Gln Gly Gln Leu Gln Thr Arg  
20 25 30

Gln Ala Ser Ser Ile Tyr Asp Asp Ser Tyr Leu Gly Tyr Ser Val Ala  
35 40 45

Val Gly Glu Phe Ser Gly Asp Asp Thr Glu Asp Phe Val Ala Gly Val  
50 55 60

Pro Lys Gly Asn Leu Thr Tyr Gly Tyr Val Thr Ile Leu Asn Gly Ser  
65 70 75 80

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

Phe Gly Tyr Ala Val Ala Ala Thr Asp Val Asn Gly Asp Gly Leu Asp  
100 105 110

Asp Leu Leu Val Gly Ala Pro Leu Leu Met Asp Arg Thr Pro Asp Gly  
115 120 125

Arg Pro Gln Glu Val Gly Arg Val Tyr Val Tyr Leu Gln His Pro Ala  
130 135 140

Gly Ile Glu Pro Thr Pro Thr Leu Thr Leu Thr Gly His Asp Glu Phe  
145 150 155 160

Gly Arg Phe Gly Ser Ser Leu Thr Pro Leu Gly Asp Leu Asp Gln Asp  
165 170 175

Gly Tyr Asn Asp Val Ala Ile Gly Ala Pro Phe Gly Gly Glu Thr Gln  
180 185 190

Gln Gly Val Val Phe Val Phe Pro Gly Gly Pro Gly Gly Leu Gly Ser  
195 200 205

Lys Pro Ser Gln Val Leu Gln Pro Leu Trp Ala Ala Ser His Thr Pro  
210 215 220

Asp Phe Phe Gly Ser Ala Leu Arg Gly Gly Arg Asp Leu Asp Gly Asn  
225 230 235 240

Gly Tyr Pro Asp Leu Ile Val Gly Ser Phe Gly Val Asp Lys Ala Val  
245 250 255

Val Tyr Arg Gly Arg Pro Ile Val Ser Ala Ser Ala Ser Leu Thr Ile  
260 265 270

Phe Pro Ala Met Phe Asn Pro Glu Glu Arg Ser Cys Ser Leu Glu Gly  
275 280 285

<210> 2  
<211> 286  
<212> PRT  
<213> Homo sapiens

<400> 2

Met Gly Gln Leu Ile Ser Asp Gln Val Ala Glu Ile Val Ser Lys Tyr  
1 5 10 15

Asp Pro Asn Val Tyr Ser Ile Lys Tyr Asn Asn Gln Leu Ala Thr Arg  
20 25 30

Thr Ala Gln Ala Ile Phe Asp Asp Ser Tyr Leu Gly Tyr Ser Val Ala  
35 40 45

Val Gly Asp Phe Asn Gly Asp Gly Ile Asp Asp Phe Val Ser Gly Val  
50 55 60

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

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

Phe Gly Phe Ser Val Ala Ala Thr Asp Ile Asn Gly Asp Asp Tyr Ala  
100 105 110

Asp Val Phe Ile Gly Ala Pro Leu Phe Met Asp Arg Gly Ser Asp Gly  
115 120 125

Lys Leu Gln Glu Val Gly Gln Val Ser Val Ser Leu Gln Arg Ala Ser  
130 135 140

Gly Asp Phe Gln Thr Thr Lys Leu Asn Gly Phe Glu Val Phe Ala Arg  
145 150 155 160

Phe Gly Ser Ala Ile Ala Pro Leu Gly Asp Leu Asp Gln Asp Gly Phe  
165 170 175

Asn Asp Ile Ala Ile Ala Ala Pro Tyr Gly Gly Glu Asp Lys Lys Gly  
180 185 190

Ile Val Tyr Ile Phe Asn Gly Arg Ser Thr Gly Leu Asn Ala Val Pro  
195 200 205

Ser Gln Ile Leu Glu Gly Gln Trp Ala Ala Arg Ser Met Pro Pro Ser  
210 215 220

Phe Gly Tyr Ser Met Lys Gly Ala Thr Asp Ile Asp Lys Asn Gly Tyr  
225 230 235 240

Pro Asp Leu Ile Val Gly Ala Phe Gly Val Asp Arg Ala Ile Leu Tyr  
245 250 255

Arg Ala Arg Pro Val Ile Thr Val Asn Ala Gly Leu Glu Val Tyr Pro.  
260 265 270

Ser Ile Leu Asn Gln Asp Asn Lys Thr Cys Ser Leu Pro Gly  
275 280 285

<210> 3  
<211> 286  
<212> PRT  
<213> Homo sapiens

<400> 3

Met Gly Leu Leu Ala Gln Ala Pro Val Ala Asp Ile Phe Ser Ser Tyr  
1 5 10 15

Arg Pro Gly Ile Leu Leu Trp His Val Ser Ser Gln Ser Leu Ser Phe  
20 25 30

Asp Ser Ser Asn Pro Glu Tyr Phe Asp Gly Tyr Trp Gly Tyr Ser Val  
35 40 45

Ala Val Gly Glu Phe Asp Gly Asp Leu Asn Thr Thr Glu Tyr Val Val  
50 55 60

Gly Ala Pro Thr Trp Ser Trp Thr Leu Gly Ala Val Glu Ile Leu Asp  
65 70 75 80

Ser Tyr Tyr Gln Arg Leu His Arg Leu Arg Ala Glu Gln Met Ala Ser  
85 90 95

Tyr Phe Gly His Ser Val Ala Val Thr Asp Val Asn Gly Asp Gly Arg

100 105 110  
 His Asp Leu Leu Val Gly Ala Pro Leu Tyr Met Glu Ser Arg Ala Asp  
 115 120 125  
 Arg Lys Leu Ala Glu Val Gly Arg Val Tyr Leu Phe Leu Gln Pro Arg  
 130 135 140  
 Gly Pro His Ala Leu Gly Ala Pro Ser Leu Leu Leu Thr Gly Thr Gln  
 145 150 155 160  
 Leu Tyr Gly Arg Phe Gly Ser Ala Ile Ala Pro Leu Gly Asp Leu Asp  
 165 170 175  
 Arg Asp Gly Tyr Asn Asp Ile Ala Val Ala Ala Pro Tyr Gly Gly Pro  
 180 185 190  
 Ser Gly Arg Gly Gln Val Leu Val Phe Leu Gly Gln Ser Glu Gly Leu  
 195 200 205  
 Arg Ser Arg Pro Ser Gln Val Leu Asp Ser Pro Phe Pro Thr Gly Ser  
 210 215 220  
 Ala Phe Gly Phe Ser Leu Arg Gly Ala Val Asp Ile Asp Asp Asn Gly  
 225 230 235 240  
 Tyr Pro Asp Leu Ile Val Gly Ala Tyr Gly Ala Asn Gln Val Ala Val  
 245 250 255  
 Tyr Arg Ala Gln Pro Val Val Lys Ala Ser Val Gln Leu Leu Val Gln  
 260 265 270  
 Asp Ser Leu Asn Pro Ala Val Lys Ser Cys Val Leu Pro Gln  
 275 280 285

<210> 4  
 <211> 864  
 <212> DNA  
 <213> Homo sapiens

<400> 4  
 atgggccaga tcctgtctgc cactcaggag cagattgcag aatcttatta ccccgagtac 60  
 ctgatcaacc tgggttcaggg gcagctgcag actcgccagg ccagttccat ctatgatgac 120  
 agctacctag gatactctgt ggctgttgggt gaattcagtg gtgatgacac agaagacttt 180  
 gttgctggtg tgcccaaagg gaacctcact tacggctatg tcaccatcct taatggctca 240  
 gacattcgat cctctacaa cttctcaggg gaacagatgg cctcctactt tggctatgca 300  
 gtggccgcca cagacgtcaa tggggacggg ctggatgact tgctggtggg ggcaccctg 360  
 ctcattggatc ggaccctga cgggcgccct caggaggtgg gcagggtcta cgtctacctg 420  
 cagcaccag cggcataga gccacgccc acccttacct tcactggcca tgatgagttt 480

```

ggccgatttg gcagctcctt gacccccctg ggggacctgg accaggatgg ctacaatgat      540
gtggccatcg gggctcccctt tgggtggggag acccagcagg gagtagtggt tgtatttcct      600
gggggcccag gagggctggg ctctaagcct tcccagggtt tgcagcccct gtgggcagcc      660
agccacaccc cagacttcctt tggctctgcc cttcgaggag gccgagacct ggatggcaat      720
ggatatcctg atctgattgt ggggtccttt ggtgtggaca aggctgtggt atacaggggc      780
cgccccatcg tgtccgctag tgcctccctc accatcttcc ccgccatggt caaccagag      840
gagcggagct gcagcttaga gggg                                           864

```

<210> 5  
 <211> 858  
 <212> DNA  
 <213> Homo sapiens

```

<400> 5
atgggtcagc ttatttcgga tcaagtggca gaaatcgat ctaaatacga cccaatggt      60
tacagcatca agtataataa ccaattagca actcggactg cacaagctat ttttgatgac      120
agctatttgg gttattctgt ggctgtcggg gatttcaatg gtgatggcat agatgacttt      180
gtttcaggag ttccaagagc agcaaggact ttgggaatgg tttatatatta tgatgggaag      240
aacatgtcct cttatacaa ttttactggc gagcagatgg ctgcatatct cggattttct      300
gtagctgcca ctgacattaa tggagatgat tatgcagatg tgttttattgg agcacctctc      360
ttcatggatc gtggctctga tggcaaacct caagaggtgg ggcaggtctc agtgtctcta      420
cagagagctt caggagactt ccagacgaca aagctgaatg gatttgaggt ctttgcacgg      480
tttggcagtg ccatagctcc tttgggagat ctggaccagg atgggtttcaa tgatattgca      540
attgctgctc catatggggg tgaagataaa aaaggaattg tttatatctt caatggaaga      600
tcaacaggct tgaacgcagt cccatctcaa atccttgaag ggcagtgggc tgctcgaagc      660
atgccaccaa gctttggcta ttcaatgaaa ggagccacag atatagacaa aaatggatat      720
ccagacttaa ttgtaggagc ttttggtgta gatcgagcta tcttatacag ggccagacca      780
gttatcactg taaatgctgg tcttgaagtg taccctagca ttttaaatca agacaataaa      840
acctgctcac tgcctgga                                           858

```

<210> 6  
 <211> 858  
 <212> DNA  
 <213> Homo sapiens

```

<400> 6
atgggtctcc tggcccaggc tccagttgcg gatattttct cgagttaccg cccaggcatc      60
cttttgtggc acgtgtcctc ccagagcctc tcctttgact ccagcaacc agagtacttc      120
gacggctact gggggtactc ggtggccgtg ggcgagttcg acggggatct caacactaca      180
gaatatgtcg tcggtgcccc cacttgagc tggaccctgg gagcggtgga aattttggat      240
tcctactacc agaggctgca tcggctgcgc gcagagcaga tggcgtcgta ttttgggcat      300

```

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| tcagtggctg tcactgacgt caacggggat gggaggcatg atctgctggt gggcgctcca  | 360 |
| ctgtatatgg agagccgggc agaccgaaaa ctggccgaag tggggcgtgt gtatttgttc  | 420 |
| ctgcagccgc gaggccccc cgcgctgggt gccccagcc tcctgctgac tggcacacag    | 480 |
| ctctatgggc gattcggctc tgccatcgca cccctgggcg acctcgaccg ggatggctac  | 540 |
| aatgacattg cagtggctgc cccctacggg ggtcccagtg gccggggcca agtgctggtg  | 600 |
| ttcctgggtc agagtgaggg gctgaggtca cgtccctccc aggtcctgga cagccccctc  | 660 |
| cccacaggct ctgcctttgg cttctccctt cgagggtgcc tagacatcga tgacaacgga  | 720 |
| taccagacc tgatcgtggg agcttacggg gcccaaccagg tggctgtgta cagagctcag  | 780 |
| ccagtgggtga aggcctctgt ccagctactg gtgcaagatt cactgaatcc tgctgtgaag | 840 |
| agctgtgtcc tacctcag                                                | 858 |