



(72) BOUGUELERET, LYDIE, FR

(72) CHUMAKOV, ILYA, FR

(71) GENSET, FR

(51) Int.Cl.⁷ C07K 14/47, C12N 15/63, C07K 16/18, G01N 33/00,
A61K 38/17, C12N 15/12

(30) 1997/08/29 (97/10823) FR

(54) **DEFENSINE HUMAINE Def-X, GENE ET cDNA, COMPOSITION
LES CONTENANT ET APPLICATIONS AU DIAGNOSTIC ET A
LA THERAPIE**

(54) **HUMAN DEFENSIN Def-X, GENE AND DNAc, COMPOSITION
CONTAINING SAME AND DIAGNOSTIC AND
THERAPEUTIC APPLICATIONS**

(57) La présente invention concerne une nouvelle défensine polypeptidique humaine Def-X, homologue de l'HNP-4, son ADN génomique et ADNc, des vecteurs, des cellules transformées par lesdits vecteurs, l'utilisation dudit polypeptide comme agent antibiotique, cytotoxique, de réparation et de régulation endocrine ou comme pesticide ainsi que des compositions cosmétiques ou pharmaceutiques pour le traitement des infections microbiennes, notamment bactériennes, fongiques, et virales, ou parasitaires, de cancers, de l'inflammation et de déficit immunitaire. L'invention concerne également des méthodes et des kits de diagnostic pour la détermination d'une infection microbienne ou parasitaire et d'une inflammation, ou pour le dépistage de prédisposition à des déficiences immunitaires ou des maladies cancéreuses.

(57) The invention concerns a novel human polypeptide defensin, homologous of HNP-4, its genomic DNA and DNAc, vectors, cells transformed by said vectors, the use of said polypeptide as antibiotic, cytotoxic, repairing and endocrine regulating agent or as pesticide as well as cosmetic or pharmaceutical compositions for treating microbial infections, in particular bacterial, fungal, and viral, or parasitic, cancers, inflammation and immunodeficiency. The invention also concerns diagnostic methods and kits for determining a microbial or parasitic infection and an inflammation, or for detecting predisposition to immunodeficiency or cancerous diseases.



PCT

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : C07K 14/47, C12N 15/12, 15/63, C07K 16/18, A61K 38/17, G01N 33/00	A1	(11) Numéro de publication internationale: WO 99/11663 (43) Date de publication internationale: 11 mars 1999 (11.03.99)
(21) Numéro de la demande internationale: PCT/FR98/01864 (22) Date de dépôt international: 28 août 1998 (28.08.98) (30) Données relatives à la priorité: 97/10823 29 août 1997 (29.08.97) FR (71) Déposant (pour tous les Etats désignés sauf US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chèvrefeuilles, F-77000 Vaux-le-Penil (FR). (74) Mandataires: MARTIN, Jean-Jacques etc.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).		(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Publiée Avec rapport de recherche internationale.
(54) Title: HUMAN DEFENSIN DEF-X, GENE AND DNAC, COMPOSITION CONTAINING SAME AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS (54) Titre: DEFENSINE HUMAINE DEF-X, GENE ET cDNA, COMPOSITION LES CONTENANT ET APPLICATIONS AU DIAGNOSTIC ET A LA THERAPIE (57) Abstract The invention concerns a novel human polypeptide defensin, homologous of HNP-4, its genomic DNA and DNAC, vectors, cells transformed by said vectors, the use of said polypeptide as antibiotic, cytotoxic, repairing and endocrine regulating agent or as pesticide as well as cosmetic or pharmaceutical compositions for treating microbial infections, in particular bacterial, fungal, and viral, or parasitic, cancers, inflammation and immunodeficiency. The invention also concerns diagnostic methods and kits for determining a microbial or parasitic infection and an inflammation, or for detecting predisposition to immunodeficiency or cancerous diseases. (57) Abrégé La présente invention concerne une nouvelle défensine polypeptidique humaine Def-X, homologue de l'HNP-4, son ADN génomique et ADNc, des vecteurs, des cellules transformées par lesdits vecteurs, l'utilisation dudit polypeptide comme agent antibiotique, cytotoxique, de réparation et de régulation endocrine ou comme pesticide ainsi que des compositions cosmétiques ou pharmaceutiques pour le traitement des infections microbiennes, notamment bactériennes, fongiques, et virales, ou parasitaires, de cancers, de l'inflammation et de déficit immunitaire. L'invention concerne également des méthodes et des kits de diagnostic pour la détermination d'une infection microbienne ou parasitaire et d'une inflammation, ou pour le dépistage de prédisposition à des déficiences immunitaires ou des maladies cancéreuses.		

DEFENSINE HUMAINE DEF-X, GENE ET cDNA, COMPOSITION LES CONTENANT ET APPLICATIONS AU DIAGNOSTIC ET A LA THERAPIE

La présente invention concerne une nouvelle défensine polypeptidique humaine Def-X, homologue de l'HNP-4, son ADN génomique et ADNc.

L'invention concerne également des vecteurs de clonage et d'expression, des cellules transformées par lesdits vecteurs. L'invention a aussi pour objet l'utilisation desdits polypeptides comme agent antibiotique, cytotoxique, de réparation et de régulation endocrine ou comme pesticide ainsi que des compositions cosmétiques ou pharmaceutiques pour le traitement des infections microbiennes, notamment bactériennes, fongiques, et virales, ou parasitaires, de cancers, de l'inflammation et de déficit immunitaire. Enfin, l'invention comprend des méthodes et des kits de diagnostic pour la détermination d'une infection microbienne ou parasitaire et d'une inflammation, ou pour le dépistage de prédisposition à des déficiences immunitaires ou des maladies cancéreuses.

Les substances antimicrobiennes sont des éléments primordiaux de la défense des organismes multicellulaires. Parmi ces substances, on trouve aussi bien des composés inorganiques simples (péroxyde d'hydrogène, acide hypochloreux, oxyde nitrique) que des peptides et protéines complexes. Ils sont présents sur les premières lignes de défense, à la surface des muqueuses de différents organes, notamment dans les cellules épithéliales de l'intestin et des poumons, selon les espèces, ainsi que dans les organelles microbicides des cellules phagocytaires d'origine hématopoïétique, où ils furent tout d'abord mis en évidence. Leur synthèse *de novo* ou leur libération à partir de sites de stockage - organelles de type lysosomes, granules cytoplasmiques, capables de les stocker sous une forme inactive ou latente - peuvent être induites rapidement, ce qui les rend particulièrement importants dans les phases précoces de résistance aux infections (Martin et al., 1995).

Les protéines antimicrobiennes d'une taille inférieure à cent acides aminés sont arbitrairement appelées peptides antimicrobiens. Plusieurs familles de peptides antimicrobiens ont été identifiées, qui diffèrent quant à la présence en leur sein de ponts disulfures, quant à leur composition en acides aminés, à leur conformation structurale et à leur spectre d'activité. Les peptides antimicrobiens comportant six cystéines conservées forment la famille des défensines. Cette famille est composée de peptides antimicrobiens présents dans de nombreuses espèces, abondants, d'environ 3-4 kDa (Ganz et Lehrer, 1994). Ces peptides sont formés de 30 à 40 acides aminés, dont six

cystéines invariantes qui forment trois liens disulfides intramoléculaires. Ils ont une conformation complexe, sont amphipathiques, riches en feuillet bêta anti-parallèles, mais dépourvus d'hélices alpha (Lehrer et Ganz, 1992). L'action antimicrobienne des défensines résulterait de leur insertion dans les membranes des cellules cibles, permettant la formation de canaux voltage-dépendants. White et al. (1995) décrivent les mécanismes possibles d'insertion membranaire et de formation de pores multimériques par les défensines, qui permettent la perméabilisation des membranes des cellules cibles, par exemple des cellules microbiennes ou tumorales. La structure cristallographique de la défensine humaine de neutrophile HNP-3 (voir ci-dessous) a été déterminée, et un mécanisme particulier de dimérisation des défensines humaines de neutrophile est en outre suggéré. La connaissance élargie de cette famille de peptides et la comparaison de leurs séquences et spectres d'activité permettront de mieux comprendre ces mécanismes et leurs spécificités, ainsi que les résidus acides aminés plus particulièrement impliqués dans ces phénomènes.

Les défensines se répartissent en trois familles de peptides, structurellement différents : les défensines "classiques", les bêta-défensines et les défensines des insectes. Ces familles présentent des différences concernant la position et l'espacement des résidus cystéines conservés, ainsi que ceux d'autres acides aminés conservés (proline, glycine) (Ganz et Lehrer, 1995).

Les défensines humaines, de type classique, proviennent essentiellement de deux sources. Elles ont d'abord été identifiées par purification peptidique à partir d'extraits de neutrophiles. Quatre défensines ont ainsi été isolées: "human neutrophil peptides" HNP-1, HNP-2, HNP-3, et HNP-4. Les trois premières sont des produits différents du même gène (Ganz et Lehrer, 1995). Ces trois peptides représentent 99 % du contenu des neutrophiles en défensines, alors que HNP-4 y est aussi présent, mais à des concentrations 100 fois plus faibles. Plus récemment, deux défensines entériques humaines, HD-5 et HD-6, ont été caractérisées dans l'intestin grêle et plus précisément dans les cellules de Paneth (Bevins et al., 1996). Alors que 16 gènes de défensines entériques ont été mis en évidence chez la souris, seuls ces deux homologues ont été identifiés chez l'homme (Mallow et al., 1996).

Les défensines ont une action antimicrobienne sur un large spectre de microorganismes *in vitro* (Martin et al., 1995). Ce spectre d'action, particulièrement large, comprend des bactéries, Gram-positives et Gram-négatives, plusieurs champignons, des mycobactéries, des parasites dont les spirochètes et plusieurs virus à enveloppe dont les virus HSV et HIV. Elles sont également cytotoxiques pour plusieurs

catégories de cellules normales et malignes, dont les cellules résistantes au TNF-alpha et au facteur cytolytique NK (Kagan et al., 1994). La grande quantité de cibles des défensines et leur abondance dans les cellules sanguines spécialisées dans la défense immunitaire, ainsi que l'augmentation dramatique de leur concentration au cours d'infections sévères, suggèrent que ces molécules joueraient un rôle important dans l'immunité naturelle aux infections et aux cancers. Notamment, l'augmentation de la transcription des gènes des défensines et la libération de granules cytoplasmiques contenant des défensines pré-synthétisées en réponse à des stimuli, contribuent à la réponse antimicrobienne locale, les défensines pouvant participer à la réaction d'inflammation, aux processus de réparation et à la régulation endocrine pendant l'infection. Les défensines hématopoïétiques pourraient contribuer au phénomène de lyse des cellules cancéreuses, phénomène médié par les neutrophiles au cours de la réponse immunitaire anticorps-dépendante. Le rôle physiologique précis des défensines entériques n'est pas clairement établi. Elles pourraient endiguer la prolifération de la flore intraluminaire ou empêcher la translocation de bactéries à travers la muqueuse intestinale (Mallow et al., 1996). L'abondance de l'ARNm de défensine dans les cellules de Paneth renforce l'hypothèse que ces cellules épithéliales joueraient un rôle clé dans la défense immunitaire de l'intestin. Il a par ailleurs été montré que leur schéma d'expression coïncide avec l'apparition des cellules de Paneth au cours de l'embryogenèse. Mallow et al. (1996) ont suggéré que de faibles taux d'expression de défensines entériques chez le fœtus serait le témoin d'une immaturité de la défense locale, ce qui prédisposerait les enfants nés prématurément à des infections dues aux microorganismes intestinaux.

Une concentration des défensines correspondant à 10 % du taux normal est constatée chez des patients atteints de "specific granule deficiency", une maladie rare du développement des granulocytes. Les sujets atteints souffrent d'infections fréquentes, provoquées par des bactéries communes (Ganz et Lehrer, 1995).

Les défensines modifiées biochimiquement sont de potentiels agents prophylactiques et thérapeutiques contre les infections (Ganz et Lehrer, 1995). La recherche concernant ces peptides antimicrobiens ou d'autres molécules participant de l'immunité naturelle, acquiert une importance particulière depuis que se développent des phénomènes de résistance des microorganismes aux antibiotiques traditionnels (Bevins et al., 1996).

La structure primaire de défensines, notamment des défensines humaines, a fait l'objet d'études récentes (White et al., 1995 ; Mallow et al., 1996). Les défensines

classiques comprennent 29 à 35 acides aminés, mais dérivent de précurseurs - préproprotéines - comprenant 90 à 100 acides aminés. La maturation protéolytique des défensines humaines de neutrophiles en peptides matures est couplée avec leur adressage vers les granulocytes ; la fonction du propeptide inclurait l'inactivation de la forme précurseur de la défensine et un support à l'acquisition de la conformation active du peptide mature (Martin et al., 1995). Les homologues peptidiques sont maximales au niveau des signaux peptides, et minimales au niveau des peptides matures, qui comportent néanmoins six résidus cystéines totalement conservés. Si la conservation de ces résidus semble nécessaire à l'acquisition de structures secondaires impliquées dans l'activité des défensines, les différences de séquences existant au sein de la très large famille de ces peptides antimicrobiens, notamment à leur extrémité N-terminale, mais aussi dans d'autres régions non conservées, semblent être des déterminants importants de leur spectre d'activité, et de leur efficacité antimicrobienne ou cytotoxique. L'identification de nouveaux membres de cette famille de peptides, et notamment de défensines humaines, est donc nécessaire à la compréhension de leur mécanisme d'action et de leur spécificité, ainsi qu'à leur utilisation comme agents anti-infectieux et/ou cytotoxiques, ou au dessin de peptides variants présentant des spectres spécifiques et/ou d'efficacité diminuée ou augmentée.

Sparkes et al. (1989), ont localisé le gène codant pour HNP-1 sur le chromosome 8, dans la région 8p23. Bevins et al. (1995), et Mallow et al. (1996), ont localisé les deux gènes codant pour HD-5 et HD-6 sur le chromosome 8, plus précisément dans la région 8p21-pter, région incluant la région précédemment identifiée comme portant les défensines hématopoïétiques. Les gènes codant pour les défensines entériques humaines HD-5 et HD-6 contiennent deux exons, alors que ceux codant pour les défensines hématopoïétiques en contiennent trois, les deux derniers exons codant pour le prépropeptide, aussi bien chez l'homme, que chez le cobaye et le lapin (Mallow et al., 1996). La comparaison des séquences génomiques des gènes *HD-5* et *HD-6* a révélé une très forte similarité des séquences flanquantes non codantes en 5', suggérant que celles-ci contiennent l'information nécessaire à la tissu-spécificité de l'expression de ces gènes; ces mêmes régions portent en outre de nombreux sites de fixation pour des facteurs de transcription, dont deux sites AP2 et six sites IL6, suggérant des voies de régulation de l'expression de ces gènes au cours des processus inflammatoires. De façon plus générale, le très important degré de similarité des séquences et de l'organisation génomique des défensines HNP-1, 2, 3, 4 et HD-5 et 6, a conduit Bevins et al. (1995) à

un modèle d'évolution tentant de relier l'organisation chromosomique de la famille, et les fractions homologues de chaque paire de gènes.

Il est enfin intéressant de noter que la région chromosomique 8p23 est impliquée dans de nombreuses pathologies, notamment cancéreuses : on citera par exemple le carcinome hépatocellulaire (Becker et al., 1996), le cancer du poumon non à petites cellules (Sundareshan et Augustus, 1996), le cancer de la prostate (Ichikawa et al., 1996), et le carcinome colorectal (Yaremko et al., 1994). Bien que ceci n'ait jamais été documenté, il est possible qu'une déficience en l'une ou l'autre des défensines humaines ait un rôle dans la prédisposition à de telles pathologies, ou dans leur développement.

La présente invention concerne une nouvelle défensine humaine, Def-X, homologue de la défensine HNP-4.

La présente invention a donc pour objet un polypeptide isolé choisi parmi les polypeptides suivants :

- a) polypeptide dont la séquence d'acides aminés est la séquence SEQ ID N° 3 ;
- b) polypeptide homologue, variant, ou modifié du polypeptide dont la séquence d'acides aminés est la séquence SEQ ID N° 3 ;
- c) polypeptide dont la séquence d'acides aminés est la séquence d'acides aminés d'un fragment biologiquement actif d'un polypeptide tel que défini en a) ou b) ;
- d) polypeptide comprenant au moins un fragment tel que défini en c).

Dans la présente description, on entendra désigner également par « polypeptide » une protéine ou un peptide.

Selon un mode préféré, le polypeptide selon l'invention est caractérisé en ce qu'il est constitué de l'un au moins des fragments suivants :

- a) peptide signal dont la séquence d'acides aminés est la séquence SEQ ID N° 4, correspondant à la séquence comprise entre la position 1 et la position 19, extrémités comprises, de la séquence d'acides aminés SEQ ID N° 3 ;
- b) région pro dont la séquence d'acides aminés est la séquence SEQ ID N° 5, correspondant à la séquence comprise entre la position 20 et la position 63, extrémités incluses, de la séquence d'acides aminés SEQ ID N° 3 ;
- c) peptide mature dont la séquence d'acides aminés est la séquence SEQ ID N° 6, correspondant à la séquence comprise entre la position 64 et la position 94, extrémités incluses, de la séquence d'acides aminés SEQ ID N° 3 ; ou
- d) fragment homologue, variant ou modifié d'un peptide selon a), b) ou c).

De façon encore préférée, les polypeptides selon la présente invention correspondent à la structure primaire de la défensine mature définie précédemment, c'est-à-dire la structure correspondant à la séquences d'acides aminés SEQ ID N° 6 suivante :

5 Ile Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly Thr Cys
Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr.

ses homologues, variants ou formes modifiées ainsi que leurs fragments biologiquement actifs et les polypeptides les contenant.

10 Il est bien entendu que les polypeptides de l'invention sont sous forme non naturelle, c'est-à-dire qu'ils ne sont pas pris dans leur environnement naturel mais qu'ils ont pu être obtenus par purification à partir de sources naturelles ou bien obtenus par recombinaison génétique ou par synthèse chimique comme cela sera décrit ci-après.

Par « polypeptide homologue », on entend un polypeptide dont la séquence d'acides aminés présente au minimum 80 %, et préférentiellement 90 %, d'acides
15 aminés en commun.

Par « polypeptide variant », on entend désigner un polypeptide muté ou correspondant à un polymorphisme pouvant exister, notamment chez l'être humain et pouvant présenter une troncature, une substitution, une délétion et/ou une addition d'au moins un acide aminé comparé au polypeptide selon l'invention.

20 Par « polypeptide modifié », on entend désigner un polypeptide obtenu par recombinaison génétique ou par synthèse chimique comme cela sera décrit ci-après, présentant une modification par rapport à la séquence normale. Ces modifications pourront notamment porter sur les domaines pré-, pro- ou mature du polypeptide selon l'invention, sur les acides aminés à l'origine d'une spécificité de spectre ou d'efficacité
25 de l'activité, ou à l'origine de la conformation structurale, de la charge, ou de l'hydrophobicité, et de la capacité de multimérisation et d'insertion membranaire du polypeptide selon l'invention. On pourra ainsi créer des polypeptides d'activité équivalente, augmentée ou diminuée, et de spécificité équivalente, plus étroite, ou plus large. Les modifications pourront aussi porter sur les séquences impliquées dans la
30 maturation, le transport et l'adressage du polypeptide.

Par « fragment biologiquement actif » d'un polypeptide selon l'invention, on entend désigner un fragment polypeptidique ayant conservé au moins une activité du polypeptide dont il est issu, en particulier :

- capable d'être reconnu par un anticorps spécifique d'un polypeptide selon
35 l'invention ; et/ou

- capable d'agir comme antibiotique ; et/ou
- capable d'agir comme agent cytotoxique ; et/ou
- capable d'agir comme agent antitumoral ; et/ou
- capable de moduler la réparation de tissu, la régulation endocrine ou le processus d'inflammation, notamment durant une infection.

Selon l'invention, les fragments biologiquement actifs de polypeptides selon l'invention auront au minimum 10 acides aminés, de préférence 15 acides aminés.

Comme cela a été indiqué précédemment, parmi les fragments biologiquement actifs, un fragment préféré est le peptide mature de séquence d'acides aminés SEQ ID N° 6.

Parmi les homologues du peptide mature, il faut citer les polypeptides dans lesquels jusqu'à 5 acides aminés ont été modifiés, tronqués à l'extrémité N- ou C-terminale, ou bien délétés, ou bien ajoutés, ce qui représente environ 80 % de la séquence.

Les fragments biologiquement actifs de ce peptide mature comportent de préférence de 10 à 15 acides aminés, dont l'intérêt pourra être de pouvoir être obtenus facilement par synthèse chimique.

Comme cela est indiqué, les modifications du polypeptide mature auront pour objectif notamment de :

- moduler l'activité de la défensine,
- modifier sa spécificité, tant au niveau des microorganismes sur lesquels elle est active que sur sa localisation tissulaire,
- modifier sa biodisponibilité.

Les composés précédents peuvent être obtenus en utilisant la chimie combinatoire, dans laquelle il est possible de faire varier systématiquement des parties de polypeptide avant de les tester sur des modèles, cultures cellulaires ou des microorganismes par exemple, pour sélectionner les composés les plus actifs ou présentant les propriétés recherchées.

La synthèse chimique présente également l'avantage de pouvoir utiliser :

- des acides aminés non naturels, ou
- des liaisons non peptidiques.

Ainsi, afin d'améliorer la durée de vie des peptides, il pourra être intéressant d'utiliser des acides aminés non naturels, par exemple sous forme D, ou bien des analogues d'acides aminés, notamment des formes soufrées par exemple.

Enfin, la structure de la défensine mature ou de ses homologues, variants ou modifiés, de même que les fragments correspondant, pourront être intégrés dans des structures chimiques de type polypeptidique ou autres. Ainsi, il pourra être intéressant de prévoir aux extrémités N- et C-terminales des composés non reconnus par les protéases.

5 L'invention comprend également les acides nucléiques codant pour un polypeptide selon l'invention.

Selon un mode préféré, les acides nucléiques selon l'invention seront choisis parmi les acides nucléiques suivants :

- a) acide nucléique de séquence SEQ ID N° 1 (génomique) ;
- 10 b) acide nucléique de séquence SEQ ID N° 2 (cDNA) ;
- c) acide nucléique équivalent, homologue, muté ou modifié, par rapport aux acides nucléiques selon a) ou b) ;
- d) fragment des séquences a), b) ou c) ayant au moins dix bases ;
- e) acide nucléique capable de s'hybrider avec l'une des séquences telles que définies
- 15 en a), b), c) ou d).

Il est entendu que la présente invention ne concerne pas les séquences génomiques dans leur environnement chromosomique naturel ; il s'agit de séquences qui ont été isolées, c'est-à-dire qu'elles ont été prélevées directement ou indirectement, leur environnement ayant été au moins partiellement modifié.

20 Il peut ainsi s'agir d'ADN génomique, d'ADNc, ou d'ARN, comportant ou non des nucléotides non naturels ; il peut s'agir d'acides nucléiques naturels isolés, ou d'acides nucléiques de synthèse.

Par acide nucléique équivalent, on entendra un acide nucléique codant pour les polypeptides selon l'invention, compte tenu de la dégénérescence du code génétique,

25 et les ADNc et ARN correspondants.

Par acide nucléique homologue, on entendra un acide nucléique dont la séquence présente une homologie d'au moins 80 %, de préférence 90 %, avec les séquences nucléiques selon l'invention.

Par acide nucléique muté, on entendra tout acide nucléique codant pour un polypeptide variant selon l'invention, et tout acide nucléique comportant, par rapport aux séquences SEQ ID N° 1 et SEQ ID N° 2, au moins une mutation dans les séquences promotrices et/ou régulatrices, lesquelles pourront avoir un effet sur l'expression du polypeptide notamment sur son taux d'expression et la tissu-spécificité de celle-ci. Les séquences présentant un polymorphisme présent chez l'être humain sont donc incluses

35 dans l'invention. Parmi ces polymorphismes, certains pourront conduire à des

déficiences immunitaires, de réponse aux infections, à des prédispositions et/ou au développement de cancers.

5 Par acide nucléique modifié, on entendra tout acide nucléique codant pour un polypeptide modifié selon l'invention, ou tout acide nucléique obtenu par mutagenèse selon des techniques bien connues de l'homme de l'art, et comportant des modifications par rapport aux séquences normales, notamment des mutations dans les séquences régulatrices et/ou promotrices, notamment conduisant à une modification du taux et/ou de la tissu-spécificité de l'expression du polypeptide.

10 La présente invention concerne l'ensemble des amorces et sondes, qui pourront être marquées selon des méthodes bien connues de l'homme du métier, permettant de mettre en évidence, notamment par des techniques basées sur l'hybridation ou sur l'amplification, par exemple par PCR, les séquences nucléiques selon l'invention, y compris de discriminer les séquences normales des séquences mutées.

15 Parmi les fragments d'acides nucléiques intéressants, il faut citer en particulier les oligonucléotides anti-sens, c'est-à-dire dont la structure assure, par hybridation avec la séquence cible, une inhibition de l'expression du produit correspondant. Il faut encore citer les oligonucléotides sens qui, par interaction avec des protéines impliquées dans la régulation de l'expression du produit correspondant, induiront soit une inhibition, soit une activation de cette expression.

20 Il pourra s'agir de séquences qui agissent aussi bien au niveau des séquences exoniques ou introniques décrites que sur les séquences flanquantes, notamment les promoteurs et/ou régions 5' UTR.

25 La présente invention concerne également des vecteurs de clonage ou d'expression comportant une séquence nucléotidique telle que décrite précédemment.

Ces vecteurs de clonage ou d'expression pourront comporter des éléments assurant l'expression de la séquence dans une cellule hôte, notamment des séquences promotrices et des séquences de régulation efficaces dans ladite cellule.

30 Le vecteur en cause pouvant être à répllication autonome ou bien destiné à assurer l'intégration de la séquence au sein des chromosomes de la cellule hôte.

Dans le cas de systèmes à répllication autonome, en fonction de la cellule hôte, procaryote ou eucaryote, on utilisera de préférence des systèmes de type plasmidique ou des systèmes viraux, les virus vecteurs pouvant être notamment des adénovirus (Perricaudet et al., 1992), des rétrovirus, des poxvirus ou des virus

herpétiques (Epstein et al., 1992). L'homme de métier connaît les technologies utilisables pour chacun de ces virus.

Ainsi, il est connu d'utiliser comme vecteur viral des virus défectifs dont la culture est effectuée dans des cellules de complémentation, ceci évitant les risques
5 éventuels de prolifération d'un vecteur viral infectieux.

Lorsque l'on souhaitera l'intégration de la séquence dans les chromosomes de la cellule hôte, il sera nécessaire de prévoir de part et d'autre de la séquence nucléotidique à intégrer une ou plusieurs séquences provenant de la cellule hôte afin d'assurer la recombinaison. Il s'agit là également de procédés qui sont largement décrits
10 dans la technique antérieure. On pourra, par exemple, utiliser des systèmes de type plasmidique ou viral ; de tels virus seront, par exemple, les rétrovirus (Temin, 1986) ou les AAV, Adenovirus Associated Virus (Carter, 1993).

L'invention concerne également les cellules procaryotes ou eucaryotes transformées par un vecteur tel que décrit précédemment et ceci afin d'assurer
15 l'expression d'une défensine Def-X naturelle, normale ou variante, ou modifiée, ou bien, par exemple, d'un de ses fragments.

Comme cela a été indiqué précédemment, la présente invention concerne également les polypeptides obtenus par culture des cellules ainsi transformées et récupération de la protéine exprimée, ladite récupération pouvant être effectuée de façon
20 intracellulaire ou bien de façon extracellulaire dans le milieu de culture lorsque le vecteur a été conçu pour assurer la sécrétion de la protéine par le biais, par exemple, d'une séquence "signal", le polypeptide étant sous forme d'un pré-polypeptide ou prépro-polypeptide. Les constructions permettant la sécrétion des polypeptides sont connues, aussi bien pour des systèmes procaryotes que des systèmes eucaryotes. Dans le
25 cadre de la présente invention, certains des polypeptides Def-X pourront comporter leur propre système de sécrétion ou d'insertion membranaire.

Il est bien entendu que les polypeptides recombinants selon l'invention peuvent être obtenus sous forme glycosylée ou non glycosylée et présenter ou non la structure tertiaire naturelle.

30 Parmi les cellules utilisables pour la production de ces polypeptides, il faut citer bien entendu les cellules bactériennes (Olins et Lee, 1993), mais également les cellules de levure (Buckholz, 1993), de même que les cellules animales, en particulier les cultures de cellules de mammifère (Edwards et Aruffo, 1993) mais également les cellules d'insectes dans lesquelles on peut utiliser des procédés mettant en oeuvre des
35 baculovirus par exemple (Luckow, 1993).

Les cellules ainsi obtenues peuvent permettre de préparer des polypeptides naturels, variants ou modifiés, Def-X, mais également des fragments de ces polypeptides, notamment des polypeptides pouvant correspondre aux fragments biologiquement actifs.

5 La présente invention concerne, en outre, les mêmes polypeptides selon l'invention mais obtenus par synthèse chimique et pouvant comporter des acides aminés non naturels ou modifiés.

10 Les polypeptides selon la présente invention, en particulier la défensine mature, de même que les homologues, dérivés ou polypeptides matures modifiés, peuvent être obtenus par synthèse chimique et ce en utilisant l'une quelconque des nombreuses synthèses peptidiques connues, par exemple les techniques mettant en œuvre des phases solides ou des techniques utilisant des phases solides partielles, par condensation de fragments ou par une synthèse en solution classique.

15 Lorsque les composés selon la présente invention sont synthétisés par la méthode en phase solide, l'acide aminé C-terminal est fixé sur un support solide inerte et comporte des groupes protecteurs de son groupement amino en alpha (et si cela est nécessaire, des protections sur ses groupes fonctionnels latéraux).

20 A la fin de cette étape, le groupe protecteur du groupement amino terminal est éliminé et on fixe le second acide aminé comportant lui aussi les protections nécessaires.

Les groupes protecteurs N-terminaux sont éliminés après que chaque acide aminé a été fixé, par contre on maintient, bien entendu, la protection sur les chaînes latérales.

25 Lorsque la chaîne polypeptidique est complète, on clive le peptide de son support et on élimine les groupes de protection latéraux.

La technique de synthèse en phase solide est décrite notamment dans Stewart et al. (1984) et Bodanszky (1984).

30 Il ne sera pas ici évoqué les détails de la synthèse, il convient simplement de rappeler que les groupes protecteurs préférés pour les groupements alpha-amino sont des groupes protecteurs de type uréthane (BOC ou FMOC). Quant aux réactifs de couplage, ils sont très nombreux, parmi eux il faut bien entendu citer plus particulièrement la N,N'-diisopropyl-carbodiimine (DIC) mise en œuvre en général dans le DMF ou le DCM.

35 Lorsque l'on souhaitera utiliser des amino-acides non naturels, il pourra être nécessaire de prévoir d'autres types de réactif et en particulier d'autres types de système de protection.

La présente invention concerne également les anticorps polyclonaux ou monoclonaux obtenus par réaction immunologique d'un organisme humain ou animal avec un agent immunogène constitué par un polypeptide selon l'invention, notamment un polypeptide obtenu par culture d'une des cellules précédemment décrites, ou par
5 synthèse chimique comme indiqué précédemment.

L'invention s'étend donc aux anticorps monoclonaux et polyclonaux ou un de leurs fragments, anticorps chimériques, capables de reconnaître spécifiquement un polypeptide selon l'invention.

L'invention comprend aussi les anticorps selon l'invention, caractérisés en ce
10 qu'ils sont marqués.

Les anticorps marqués pourront être, par exemple, immunoconjugués à des enzymes telles que la peroxydase ou la phosphatase alcaline, ou marqués à l'aide de composés fluorescents, de la biotine ou encore radiomarqués. Les techniques de marquage sont bien connues de l'homme du métier et ne seront pas développées dans la
15 présente description.

L'invention s'étend également à l'utilisation d'un polypeptide selon l'invention comme agent antimicrobien, notamment antibactérien, antifongique, antiviral et/ou antiparasitaire, comme agent cytotoxique, à visée notamment anticancéreuse, et/ou comme agent de modulation des processus d'inflammation, de réparation tissulaire et de
20 régulation endocrine, notamment corticostatique.

Selon un autre aspect, l'invention concerne une composition pharmaceutique comprenant un polypeptide selon l'invention, pouvant être associée à un véhicule pharmaceutiquement acceptable.

Une telle composition pourra être administrée par voie systémique, locale ou
25 topique.

Son mode d'administration, sa posologie, ses formes galéniques optimales pourront être déterminés selon les critères généralement pris en compte dans l'établissement d'un traitement adapté à un patient, notamment son âge, son poids corporel, la tolérance de traitement, ses effets secondaires constatés, etc..

L'invention comprend également une composition pharmaceutique comprenant un vecteur selon l'invention capable d'exprimer *in vivo* un polypeptide selon l'invention, pouvant être associé à un véhicule pharmaceutiquement acceptable.
30

Il est également possible de prévoir l'expression de polypeptides ou leurs fragments *in vivo*, notamment par le biais de la thérapie génique et en utilisant les
35 vecteurs qui ont été décrits précédemment.

Dans le cadre de la thérapie génique, il est possible également de prévoir l'utilisation des séquences des gènes ou des ADNc précédemment décrits, "nus", cette technique a notamment été développée par la société Vical, qui a montré qu'il était, dans ces conditions, possible d'exprimer le polypeptide dans certains tissus sans avoir recours
5 au support d'un vecteur viral notamment.

Toujours dans le cadre de la thérapie génique, il est également possible de prévoir l'utilisation de cellules transformées ex-vivo, lesquelles pourront être ensuite réimplantées, soit telles quelles, soit au sein de systèmes de type organoïde, tel que cela est également connu dans l'état de la technique (Danos et al. 1993). On peut également
10 envisager l'utilisation d'agents facilitant le ciblage d'un type cellulaire déterminé, la pénétration dans les cellules ou le transport vers le noyau.

Lesdites compositions pharmaceutiques sont, selon l'invention, destinées à la prévention et/ou au traitement des infections microbiennes, notamment les infections microbiennes d'origines bactériennes, de bactéries Gram-positives ou Gram-négatives, mycobactériennes, fongiques et virales, ou parasitaires, notamment de spirochètes.
15

Selon un mode préféré, l'invention concerne avantageusement les compositions pharmaceutiques selon l'invention caractérisées en ce que les infections virales sont des infections liées à des virus à enveloppe, notamment les virus HSV et HIV.

20 L'invention a également pour objet des compositions pharmaceutiques selon l'invention, destinées à la prévention et/ou au traitement des cancers, notamment les mélanomes, le cancer du foie, de la prostate, du poumon non à petites cellules ou le carcinome colorectal.

L'invention comprend, en outre, des compositions pharmaceutiques selon
25 l'invention, destinées à augmenter les défenses immunitaires, à augmenter les défenses immunitaires en cas d'immunodéficience acquise ou à prévenir l'immunodéficience, notamment pour le traitement du psoriasis, ou à moduler les processus inflammatoires dans les cas notamment de maladies à inflammation chronique.

Les polypeptides selon la présente invention sont plus particulièrement
30 utilisables sous forme topique externe, par exemple sur la peau et les muqueuses. Ces formes topiques externes peuvent être aussi bien à usage pharmaceutique, dermatologique qu'à usage cosmétique.

En particulier, ces compositions peuvent être utilisées comme agent antiseptique pharmaceutique ou bien comme antiseptique dans certains cosmétiques, soit

pour assurer un nettoyage de la peau ou des phanères et/ou à titre de conservateur des compositions.

Les compositions topiques selon la présente invention peuvent être utilisées notamment dans certaines affections cutanées, oculaires, vaginales ou buccales. Elles
5 peuvent également être utilisées comme agent cosmétique additionnel, notamment dans certains shampooings traitants.

L'invention concerne également la mise en évidence de l'absence ou d'une quantité anormale de protéine ou d'acide nucléique correspondant à la défensine X comme marqueur d'une infection ou de pathologies qui seront décrites ci-après.

10 L'invention concerne également la mise en évidence d'une forme anormale de la protéine ou la présence d'un acide nucléique anormal correspondant à une défensine mutée qui peut éventuellement être totalement inactive. Dans ce cas, la présence de cette forme anormale peut être un marqueur de prédisposition à certaines affections, notamment l'immunodéficience et/ou des cancers.

15 C'est pourquoi, la présente invention concerne une méthode de diagnostic d'une immunodéficience et/ou d'une prédisposition à certains types de cancers, caractérisée en ce qu'on met en évidence dans un prélèvement de patient la présence d'une défensine anormale et/ou d'une séquence codant pour une défensine anormale.

Les méthodes de diagnostic selon la présente invention permettent,
20 notamment, la mise en évidence d'une immunodéficience, et/ou d'une prédisposition à un ou des cancers, notamment ceux cités précédemment, en particulier dans des familles à risque. Ce type de diagnostic sera en général effectué par mise en évidence des formes mutées de la protéine ou des séquences d'acide nucléique.

Mais l'invention concerne également des méthodes de diagnostic de
25 l'inflammation, d'immunodéficience, de prédisposition à des affections de type cancer et/ou d'infections dues à des microorganismes ou liées à un déficit immunitaire ou phénomène inflammatoire, caractérisées en ce qu'elles comprennent le dosage d'un polypeptide ou d'un acide nucléique selon l'invention dans un échantillon biologique et la comparaison du résultat dudit dosage obtenu avec la quantité de polypeptide ou
30 d'acide nucléique présente normalement dans un échantillon biologique équivalent.

Dans ce cas, le dosage peptidique permettra, en général, une détection d'une infection microbienne ou parasitaire et/ou d'une inflammation. Les dosages peptidiques peuvent être réalisés par tout procédé connu, ELISA ou RIA par exemple. La mise en évidence d'une forme anormale de la défensine-X peut être réalisée, par exemple, à

l'aide d'un anticorps monoclonal spécifique de cette forme, en particulier les anticorps objet de l'invention.

Selon un mode de réalisation préféré, l'invention comprend avantageusement les méthodes caractérisées en ce qu'elles mettent en œuvre une sonde et/ou une amorce
5 oligonucléotidique selon l'invention.

On préférera en général les méthodes dans lesquelles tout ou partie de la séquence correspondant au polypeptide Def-X est amplifiée préalablement par dosage d'acide nucléique selon l'invention, ces méthodes d'amplification pouvant être réalisées par des méthodes dites PCR ou PCR-like. Par PCR-like on entendra désigner toutes les
10 méthodes mettant en œuvre des reproductions directes ou indirectes des séquences d'acides nucléiques, ou bien dans lesquelles les systèmes de marquage ont été amplifiés, ces techniques sont bien entendu connues, en général il s'agit de l'amplification de l'ADN par une polymérase ; lorsque l'échantillon d'origine est un ARN il convient préalablement d'effectuer une transcription reverse. Il existe actuellement de très
15 nombreux procédés permettant cette amplification, par exemple les méthodes dites NASBA "Nucleic Acid Sequence Based Amplification" (Compton 1991), TAS "Transcription based Amplification System" (Guatelli et al. 1990), LCR "Ligase Chain Reaction" (Landegren et al. 1988), "Endo Run Amplification" (ERA), "Cycling Probe Reaction" (CPR), et SDA "Strand Displacement Amplification" (Walker et al. 1992),
20 bien connues de l'homme du métier.

L'invention concerne en outre des kits ou nécessaires de diagnostic pour la détermination d'une infection microbienne ou parasitaire, d'une inflammation, d'une immunodéficience et/ou d'une prédisposition à des affections de type cancer, caractérisés en ce qu'ils comprennent un anticorps selon l'invention.

25 Les kits ou nécessaires de diagnostic pour la détermination d'une infection microbienne ou parasitaire, d'une inflammation, d'une immunodéficience et/ou de prédisposition à des affections de type cancer, caractérisés en ce qu'ils comprennent une sonde et/ou une amorce selon l'invention font également partie de l'invention.

L'invention a, enfin, pour objet l'utilisation de polypeptide selon l'invention
30 comme pesticide, notamment pour la culture de végétaux d'intérêt industriel comme, par exemple, les plantes vivrières telles que le maïs, le blé, le soja, le riz ou le colza, les plantes fourragères, les arbres fruitiers, la vigne ou les plantes ornementales.

D'autres caractéristiques et avantages de la présente invention apparaîtront à la lecture des exemples ci-après, illustrés par les figures dont les légendes sont décrites
35 ci-dessous.

Légendes des figuresFigure 1

Séquence génomique de hDef-X.

Est présentée la totalité de la séquence d'ADN génomique de hDef-X qui présente une
5 homologie significative avec le gène codant pour hDef-4 (HNP-4).

La séquence présente les sites suivants, dont la présence est déduite par homologie avec la séquence hDef-4 :

	• CAAT box	1711-1714
	• TATA box	1758-1767
10	• mRNA start	1836
	• exon 1	1836-1874
	• site d'épissage 1	GTCAGT
	• insertion Alu	2155-2335
	• insertion fragment de L1	2710-2780
15	• site d'épissage 2	CAG
	• exon 2	3394-3577
	• début de phase codante	3406
	• site d'épissage 3	GTGAGA
	• site d'épissage 4	CAG
20	• exon 3	4164-4379
	• fin de phase codante	4276
	• site de polyadénylation	4374-4379.

Figure 2

Alignement des séquences génomiques des défensines humaines Def-X et Def-4 (HNP-
25 4).

Alignement de la totalité de la séquence d'ADN génomique de la nouvelle défensine Def-X présentant une homologie avec l'ADN génomique de hDef-4 (GenBank accession number U18745).

Les annotations présentent les positions sur la séquence de hDef-4 des signaux CAAT
30 box, TATA box, sites d'épissage, débuts et fins d'introns/d'exons, début de transcription, site de polyadénylation.

Figure 3

Alignement des séquences d'ADNc de hDef-4 (HNP-4) et hDef-X.

Les séquences présentent une homologie globale de 61,4 %. L'alignement révèle une
35 insertion d'environ 75 bases en aval du codon STOP, présentes sur la séquence de hDef-

4, mais non sur celle de hDef-X ; l'homologie forte reprend sur toute la région comprise entre l'extrémité de cette insertion et celle de l'ADNc. En dehors de cette région d'insertion, le degré d'homologie entre séquences nucléiques est donc remarquable.

Figure 4

5 Séquence peptidique de la protéine hDef-X.

La position des sites de clivage du signal peptide et de la région pro ont été déduites de l'alignement des séquences peptidiques de hDef-4 et hDef-X.

Figure 5

10 Alignement des séquences peptidiques des défensines humaines connues hDef-1, hDef-4, hDef-5, et hDef-6 avec hDef-X.

- * L'étoile indique un acide aminé conservé sur les cinq séquences.
 - Le point indique un acide aminé dont la classe est conservée sur les cinq séquences (acide aminé soit identique, soit faisant l'objet d'une substitution conservative).
 - ^ six flèches indiquent les positions des six cystéines conservées au travers de la
- 15 classe des défensines classiques et responsables de la structure tridimensionnelle nécessaire à l'activité de ces peptides.

EXEMPLES

20 Exemple 1 : Identification du gène codant pour hDef-X

Isolement du BAC B0725B12

Afin d'analyser la région 8p23 du génome humain, notamment dans la région connue comme portant des gènes codant pour des défensines humaines, on a isolé un BAC ("Bacterial Artificial Chromosome") correspondant à ladite région. Une banque

25 de BACs couvrant le génome humain complet a été préparée à partir de l'ADN d'une lignée lymphoblastique humaine dérivée de l'individu n° 8445 des familles du CEPH. Cette lignée a été utilisée comme source d'ADN de haut poids moléculaire. L'ADN a été partiellement digéré par l'enzyme de restriction BamHI, puis cloné au site BamHI du plasmide pBeloBacII. Les clones ainsi obtenus ont été "poolés" et criblés selon une

30 procédure d'analyse tridimensionnelle précédemment décrite pour le criblage des banques de YACs ("Yeast Artificial Chromosome") (Chumakov et al., 1992 et 1995). Les pools tridimensionnels obtenus ont été criblés par PCR à l'aide des amorces encadrant le marqueur SHGC-10793, pour Neutrophil defensin 4 precursor (GeneBank : numéro d'accèsion U18745) ; un clone du BAC B0725 B12 a été ainsi isolé.

Après digestion par l'enzyme de restriction NotI, la taille de l'insert porté par ce BAC a été déterminée sur un gel d'agarose 0,8 % après migration par électrophorèse en champ alterné (CHEF) (4 heures à 9Volts/cm, avec un angle de 100°, à 11°C en tampon 0,5 x TAE). On a ainsi mis en évidence que le BAC B0725B12 porte un insert
5 de 220 kb, avec un site interne pour l'enzyme NotI.

Localisation chromosomique du BAC B0725B12 par hybridation in situ fluorescente (FISH)

La localisation chromosomique du BAC dans la région candidate 8p23.1-23.2 a été confirmée par hybridation in situ fluorescente (FISH) sur chromosomes
10 métaphasiques, selon la méthode décrite par Cherif et al., (1990).

Séquençage de l'insert du BAC B0725B12

Afin de séquencer l'insert du BAC B0725B12, on a préparé une banque de sous-clones à partir de l'ADN soniqué de ce BAC.

Les cellules issues d'un litre de culture "overnight" ont été traitées par lyse
15 alcaline selon les techniques classiques. Après centrifugation du produit obtenu dans un gradient de chlorure de césium, 12 µg d'ADN du BAC B0725B12 ont été purifiés. 3 µg d'ADN ont été soniqués afin d'obtenir des fragments dont les tailles se distribuent uniformément de 1,2 kb à 1,5 kb. Les fragments obtenus ont été traités dans un volume de 50 µl avec 2 unités de Vent polymérase pendant 20 minutes à 70°C, en présence des 4
20 déoxytriphosphates (100 µM). Les fragments aux extrémités franches résultant de cette étape ont été séparés par électrophorèse en gel 1 % d'agarose à bas point de fusion (60 Volts pendant 3 heures). Les fragments groupés selon leurs tailles ont été excisés et les bandes obtenues traitées par l'agarose. Après extraction au chloroforme et dialyse sur colonnes Microcon 100, l'ADN en solution a été ajusté à une concentration de 100 ng/µl.
25 Une ligation a été effectuée "overnight" en mettant en présence 100 ng de l'ADN fragmenté du BAC B0725B12 et 20 ng d'ADN du vecteur BluescriptSK linéarisé par digestion enzymatique, et traité par la phosphatase alcaline. Cette réaction a été réalisée dans un volume final de 10 µl en présence de 40 unités/µl de T4 ADN ligase (New England Biolabs). Les produits de ligation ont ensuite servi à transformer par
30 électroporation, soit une souche XL-Blue (pour les plasmides multicopies), soit une souche D10HB (pour les sous-clones issus du BAC). Les clones lacZ⁻ résistant à l'antibiotique ont été repiqués individuellement en microplaques pour stockage et séquençage.

On a ainsi obtenu 960 sous-clones correspondant à l'insertion de fragments
35 de 1,2 kb à 1,5 kb au site BamHI (rendu franc) du plasmide BluescriptSK.

Les inserts de ces sous-clones ont été amplifiés par PCR sur cultures bactériennes conduites "overnight", en utilisant les amorces des vecteurs flanquant les insertions. La séquence des extrémités de ces inserts (en moyenne 500 bases de chaque côté) a été déterminée par séquençage automatique fluorescent sur séquenceur ABI 377, 5 équipé du logiciel ABI Prism DNA Sequencing Analysis (version 2.1.2).

Les fragments de séquence provenant des sous-BACs ont été assemblés par le logiciel Gap4 de R. Staden (Bonfield et al., 1995). Ce logiciel permet la reconstruction d'une séquence complète à partir de fragments de séquences. La séquence déduite de l'alignement des différents fragments est la séquence consensus.

10 On a enfin utilisé des techniques de séquençage dirigé (marche systématique de l'amorce) pour parfaire les séquences et relier les contigs.

Analyse des séquences pour l'identification de gènes

Les exons potentiels de l'insert du BAC B0725B12 ont été repérés par recherche d'homologie sur les banques publiques de protéines, d'acides nucléiques et 15 d'EST (Expressed Sequence Tags).

Banques de données

On a utilisé des refontes locales des principales banques publiques. La banque de protéines utilisées est constituée par la fusion non redondante des banques 20 Genpept (traduction automatique de GenBank, NCBI ; Benson et al., 1996) ; Swissprot (George et al., 1996) et PIR/NBRF (Bairoch et al., 1996). Les doublons ont été éliminés par le logiciel "nrdb" (domaine public, NCBI ; Benson et al., 1996). Les répétitions internes ont ensuite été masquées par le logiciel "xnu" (domaine public, NCBI ; Benson et al., 1996). La banque résultante, dénommée NRPU (Non-Redundant Protein-Unique) 25 a servi de référence pour les recherches d'homologies protéiques. Les homologies trouvées avec cette banque ont permis de localiser des régions codant potentiellement pour un fragment de protéine au moins apparenté à une protéine connue (exons codants). La banque d'EST utilisée est composée des sous-sections "gbest" (1-9) de Genbank (NCBI ; Benson et al., 1996). Elle contient tous les fragments de transcrits publics.

30 Les homologies trouvées avec cette banque ont permis de localiser des régions potentiellement transcrites (présentes sur l'ARN messenger).

La banque d'acides nucléiques (autres que les EST) utilisée contient toutes autres sous-sections de Genbank et de l'EMBL (Rodriguez-Tome et al., 1996) dont les doublons ont été éliminés comme précédemment.

Logiciels

On a utilisé l'ensemble de logiciels BLAST (Altschul et al., 1990) de recherche d'homologies entre une séquence et des banques de données protéiques ou nucléiques. Les seuils de signification utilisés dépendent de la longueur et de la complexité de la région testée ainsi que de la taille de la banque de référence. Ils ont été ajustés et adaptés à chaque analyse.

Exemple 2 : analyse des séquences nucléiques et peptidiques de hDef-X**Structure du gène codant pour hDef-X**

L'alignement du gène codant pour hDef-X avec ceux codant pour les défensines connues a permis de noter une homologie maximale entre hDef-X et hDef-4 (Figure 2). Le taux global d'homologie des deux séquences nucléiques est de 72 %. Les deux seules régions de l'ADN génomique de hDef-X ne présentant pas d'homologie avec celui de hDef-4 correspondent à deux zones d'insertion de séquence répétée dans la séquence de hDef-X, qui sont absentes sur la séquence de hDef-4 : un élément de type Alu (positions 2155 à 2335) et un fragment d'élément de Line 1 (positions 2710 à 2780).

On note une conservation importante de la région flanquant en 5' la région promotrice, d'où découle probablement une conservation importante des éléments de régulation de la stabilité du messenger et de l'expression du gène.

La forte conservation de la séquence de l'exon 1, non traduit, permet de rattacher définitivement la défensine hDef-X à la classe des défensines classiques hématopoïétiques, soit hDef-1, 2, 3 et 4, par opposition aux défensines entériques hDef-5 et 6, dont la séquence génomique ne comporte que deux exons, tous deux codants.

L'alignement des ADNc de hDef-4 et hDef-X, indiquant une homologie supérieure à 60 %, est présenté Figure 3.

Analyse protéique

La séquence peptidique de la défensine selon l'invention est représentée Figure 4. Les trois domaines de la protéine sont positionnés comme suit :

- peptide signal : aa 1-19
- région pro : aa 20-63
- peptide mature : aa 64-94.

Les degrés d'homologies spécifiques entre hDef-4 et hDef-X ont été calculés, selon la région de la protéine concernée :

- peptide signal : 63,2 %

- région pro : 52,3 %
- peptide mature : 37,9 %.

L'homologie globale est de 49,5 %. Ces chiffres confirment la très forte homologie qui existe entre défensines, homologie maximale au niveau des peptides signaux et minimale au niveau des peptides matures.

On retrouve dans la séquence protéique primaire de Def-X les acides aminés conservés dans la classe des défensines classiques, notamment les six cystéines impliquées dans la structure tridimensionnelle de celles-ci (Figure 5).

Afin de prédire les structures secondaires présentes sur la défensine selon l'invention, on a utilisé les logiciels de prédiction de structure secondaire inclus dans le Protein Interpretation Package, Copyright MRC 1994, Medical Research Council, Hillsroad, Cambridge, United Kingdom.

Ces logiciels ont notamment permis de comparer les structures prédites de Def-X et HNP-4. Profils d'hydrophobicité, structures en alpha-hélices, feuillets β , amphiphilicité sont superposables dans les deux peptides, ce qui suggère des processus analogues d'insertion membranaire et de formation de canaux ioniques multimériques pour ces deux défensines.

Exemple 3 : Recherche de mutations associées à des cas familiaux de cancers

Extraction de l'ADN génomique

L'ADN génomique de patients immunodéficients ou atteints de cancer, est extrait du sang veineux périphérique après lyse cellulaire, digestion protéique, partition organique et finalement précipitation alcoolique, selon des techniques classiques bien connues de l'homme de l'art.

Il est notamment intéressant d'étudier la présence de mutations dans l'ADN génomique d'individus issus de familles à fort taux cancer, tous types de cancers confondus. Une déficience dans un gène de défensine de granulocyte, tel hDef-X peut en effet avoir un rôle dans la prédisposition aux cancers, comme mentionné précédemment.

Amplification de l'ADN génomique

Des amorces oligonucléotidiques sont utilisées pour l'amplification génomique des séquences exoniques dérivées du BAC B0725B12 ; elles sont prédites par analyse informatique, et définies à l'aide du logiciel OSP (Hillier et al., 1991).

Toutes ces amorces contiennent, en amont des bases spécifiquement ciblées par l'amplification, une queue oligonucléotidique universelle commune, destinée à permettre le séquençage des fragments amplifiés (PU : 5'-

TGTAACACGACGGCCAGT-3' pour les amorces en amont, et RP : 5'-CAGGAAACAGCTATGACC-3' pour les amorces en aval).

Les amorces oligonucléotidiques sont synthétisées selon la méthode des phosphoramidites, sur un synthétiseur GENSET UFPS 24.1.

5 L'amplification de chaque séquence exonique prédite est réalisée par réaction d'amplification en chaîne par polymérase (PCR), dans les conditions suivantes :

	Volume final	50 µl
	ADN génomique	100 ng
	MgCl ₂	2 mM
10	dNTP (pour chacun)	200 µM
	Amorce (pour chacune)	7.5 pmoles
	AmpliTaq Gold DNA polymerase (Perkin)	1 unité
	Tampon de PCR (10X = 0.1 M Tris HCl pH 8.3, 0.5 M KCl)	1 X.

15 L'amplification est réalisée dans un thermocycleur Perkin Elmer 9600 ou MJ Research PTC200 avec couvercle chauffant. Après un chauffage à 94°C pendant 10 minutes, 35 cycles sont effectués. Chaque cycle comprend : 30 secondes à 94°C, 1 minute à 55°C et 30 secondes à 72°C. Un segment final d'élongation de 7 minutes à 72°C termine l'amplification.

20 La quantité de produits d'amplification obtenue est déterminée sur microplaque de 96 puits, par fluorométrie, utilisant l'agent intercalant Picogreen (Molecular Probes).

Détection des polymorphismes/mutations

25 Les produits de l'amplification génomique par PCR sont séquencés sur séquenceur automatique ABI 377, en utilisant des amorces fluorescentes marquées par les fluorochromes ABI (Joe, Fam, Rox et Tamra) et l'ADN polymérase Thermosequanase (Amersham).

Les réactions sont réalisées en microplaques de 96 puits, sur thermocycleur Perkin Elmer 9600, dans des conditions classiques de cycles de température :

- 8 cycles : dénaturation : 5 sec. à 94°C ; hybridation : 10 sec. ; élongation : 30 sec. à 30 72°C, puis
- 13 cycles : dénaturation : 5 sec. à 94°C ; élongation : 30 sec. à 72°C.

6 unités de Thermosequanase, et 5-25 ng de produit d'amplification sont utilisés par réaction de séquence.

35 A l'issue des cycles d'amplification, les produits des réactions de séquence sont précipités dans l'éthanol, resuspendus dans du tampon de charge contenant de la

formamide, dénaturés, et déposés sur gels d'acrylamide 4 % ; les électrophorèses (2 heures 30 à 3 000 Volts) sont conduites sur séquenceurs ABI 377 équipés des logiciels ABI de collection et d'analyse (ABI Prism DNA Sequencing Analysis Software, version 2.1.2.).

- 5 Les séquences obtenues chez des patients atteints des déficiences étudiées, notamment chez des patients issus de familles à forte prédisposition aux cancers, sont comparées aux séquences obtenues chez des sujets contrôles, apparentés et non apparentés. Une analyse statistique (calcul de lod score) permet de conclure quant à la signification de la présence d'un site d'hétérozygotie et à son association avec une
- 10 prédisposition aux cancers.

Exemple 4 : Recherche de mutations ponctuelles

- Les mutations ponctuelles identifiées comme indiqué ci-dessus, peuvent ensuite être mises en évidence chez des sujets présentant une potentielle déficience dans
- 15 le gène codant pour hDef-X, selon de nombreuses méthodes connues de l'homme de l'art. Parmi celles-ci, on peut citer la liste non exhaustive suivante :
- séquençage
 - « single nucleotide primer extension » (Syvanen et al., 1990)
 - RFLP
 - 20 • recherche de « single strand conformation polymorphism »
 - méthodes basées sur un clivage des régions misappariées (clivage enzymatique par la S1 nucléase, clivage chimique par différents composés tels que la pipéridine ou le tétrioxide d'osmium)
 - mise en évidence d'hétéroduplex en électrophorèse
 - 25 • méthodes basées sur l'utilisation d'« allele specific oligonucleotide » (ASO, Stoneking et al., 1991)
 - méthode OLA (« dual color oligonucleotide ligation assay, Samiotaki et al., 1994)
 - méthode ARMS (« amplification refractory mutation system »), ou ASA (« allele specific amplification »), ou PASA (« PCR amplification of specific allele ») (Wu et
 - 30 al., 1989).

REFERENCES

- Altschul, Stephen F., Gish W., Miller W., Myers E. W., & Lipman D.J. Basic local alignment search tool. J. Mol. Biol. 215:403-10 (1990).
- 5 Bairoch A. & Apweiler R. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res. 24: 21-25 (1996).
- 10 Becker S.A., Zou, Y.Z. & Slagle, B.L. Frequent loss of chromosome 8p in hepatitis B virus-positive hepatocellular carcinomas from China. Cancer Res. 56 (21) : 5092-7 (1996).
- 15 Benson D. A., Boguski M., Lipman D. J. & Ostell J. GenBank. Nucleic Acids Res. 24: 1-5 (1996).
- 20 Bodansky M., Principles of peptide synthesis, (1984).
- 25 Bevens, C.L., Jones, D.E., Dutra, A., Schaffzin, J. & Muenke, M. Human enteric defensin genes : chromosomal map position and a model for possible evolutionary relationships. Genomics 31 : 95-106 (1996).
- 30 Bonfield J. K., Smith K. F. & Staden R. A new DNA sequence assembly program. Nucleic Acids Res. 23 : 4992-9 (1995).
- Buckholz R.G. Yeast Systems for the Expression of Heterologous Gene Products. Curr. Op. Biotechnology 4: 538-542 (1993).
- Carter B.J. Adeno-Associated virus vectors. Curr. Op. Biotechnology 3: 533-539 (1993).
- Cherif D., Julier C., Delattre O., Derré J., Lathrop G.M., & Berger R. : Simultaneous localization of cosmid and chromosome R-banding by fluorescence microscopy - Applications to regional mapping of chromosome 11. Proc.Natl.Acad.Sci. USA. 87 : 6639-6643 (1990).

Chumakov I., Rigault P., Guillou S., Ougen P., Billault A., Guasconi G., Gervy P., Le Gall I., Soularue P., Grinas P. et al. Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* 359: 380-386 (1992).

- 5 Chumakov I.M., Rignault P., Le Gall I. et al. A YAC contig map of the human genome. *Nature* 377 suppl : 175-183 (1995).

Compton J. Nucleic Acid Sequence-Based Amplification. *Nature* 350: 91-92 (1991).

- 10 Danos O., Moullier P. & Heard J.M. Réimplantation de cellules génétiquement modifiées dans des néo-organs vascularisés. *Médecine/Sciences* 9:62-64 (1993).

Edwards C.P. et Aruffo A. Current applications of COS cell based transient expression systems. *Curr. Op. Biotechnology* 4: 558-563 (1993).

15

Epstein A. : Les vecteurs herpétiques pour le transfert de gènes - *Médecine/Sciences* 8: 902-911 (1992).

Ganz T. & Lehrer R.I. Defensins. *Curr. Op. Immunology*. 6 : 584-9 (1994).

20

Ganz T. & Lehrer R.I. Defensins. *Pharmac. Ther.* Vol. 66 : 191-205 (1995).

George D. G., Barker W. C., Mewes H. W, Pfeiffer F. & Tsugita A. The PIR-International Protein Sequence Database. *Nucleic Acids Res.* 24: 17-20 (1996).-

25

Guatelli J.C. et al. Isothermal in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878 (1990).

- 30 Hillier L. & Green P. OSP : a computer program for choosing PCR and DNA sequencing primers. *PCR Methods Appl.* 1: 124-8 (1991).

- 35 Ichikawa, T., Nihei, N., Kuramochi, H., Kawana, Y., Killary, A.M., Rinker-Schaeffer, C.W., Barrett, J.C., Isaacs, J.T., Kugoh, H., Oshimura, M. & Shimazaki, J. Metastasis suppressor genes for prostate cancer. *Prostate Suppl.* 6 : 31-35 (1996).

- Kagan, B.L., Ganz, T. & Lehrer, R.I. Defensins : a family of antimicrobial and cytotoxic peptides. *Toxicology* 87 : 131-149 (1994).
- Landegren U., Kaiser R., Sanders J. & Hood L. A ligase-mediated gene detection
5 technique. *Science* 241: 1077-1080 (1988).
- Lehrer & Ganz. Defensins : endogenous antibiotic peptides from human leukocytes. *Ciba Found. Symp.* 171 : 276-290 (1992).
- 10 Luckow V.A. Baculovirus systems for the expression of human gene products. *Curr. Op. Biotechnology* 4: 564-572 (1993).
- Mallow, E.B., Harris, A., Salzman, N., Russel, J.P., DeBerardinis, R.J., Ruchelli, E., & Bevens, C.L. Human enteric defensins. Gene structure and developmental expression. *J.*
15 *Biol. Chem.* 271 (8) : 4038-4045 (1996).
- Martin, E., Ganz, T. & Lehrer, R.I. Defensins and other endogenous peptide antibiotics of vertebrates. *J. Leukocyte Biol.* 58 : 128-136 (1995).
- 20 Olins P.O. et Lee S.C. Recent advances in heterologous gene expression in *E. coli*. *Curr. Op. Biotechnology* 4: 520-525 (1993).
- Perricaudet M., Stratford-Perricaudet L., & Briand P. : La thérapie génique par adénovirus - *La Recherche* 23: 471-473 (1992).
- 25 Rodriguez-Tome P., Stoehr P. J., Cameron G. N., & Flores T. P. The European Bioinformatics Institute (EBI) databases. *Nucleic Acids Res.* 24: 6-12 (1996).
- Samiotaki M., Kwiatkowski M., Parik J., & Landegren U. Dual-color detection of DNA
30 sequence variants through ligase-mediated analysis. *Genomics* 20: 238-242 (1994).
- Sparkes, R.S., Kronenberg, M., Heinzmann, C., Daher, K.A., Klisak, I., Ganz, T. & Mohandas, T. Assignment of defensin gene(s) to human chromosome 8p23. *Genomics* 5
(2) : 240-4 (1989).

Stewart, J.M. et Yound, J.D. Solid Phase Peptides Synthesis. Pierce Chem. Company, Rockford, Ill, 2ème éd.. (1984).

5 Stoneking M., Hedgecock D., Higuchi R.G., Vigilant L., & Erlich H.A. Population variation of human DNA control region sequences by enzymatic amplification and sequence-specific oligonucleotide probes. Am. J. Hum. Genet. 48: 370-382 (1991).

10 Sundareshan, T.S. & Augustus, M. Cytogenetics of non-small cell lung cancer : simple technique for obtaining high quality chromosomes by fine needle aspirate cultures. Cancer Genet. Cytogenet. 91 (1) : 53-60 (1996).

15 Syvänen A.C., Aalto-Setälä K., Harju L., Kontula K. & Soderlund H. A primer-guided nucleotide incorporation assay in the genotyping of Apo E. Genomics 8 : 684-692 (1990).

20 Temin H.M. : Retrovirus vectors for gene transfer. In Kucherlapati R., ed. Gene Transfer, New York, Plenum Press, 149-187 (1986).

Walker G.T., Fraiser M.S., Schram J.L., Little M.C., Nadeau J.G., & Malinowski D.P.
25 Strand displacement amplification : an isothermal in vitro DNA amplification technique. Nucleic Acids Res. 20: 1691-1696 (1992).

White, S.H., Wimley, W.C. & Selsted, M.E. Structure, function, and membrane integration of defensins. Curr. Op. Structural Biology. 5 : 521-527 (1995).

25 Wu D.Y., Ugozzoli L., Pal B.K. & Wallace R.B. Allele-specific amplification of b-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA 86 : 2757-2760 (1989).

30 Yaremko, M.L., Wasylyshyn, M.L., Paulus, K.L., Michelassi, F. & Westbrook, C.A. Deletion mapping reveals two regions of chromosome 8 allele loss in colorectal carcinomas. Genes Chromosomes Cancer. 10 (1) : 1-6 (1994).

LISTE DE SEQUENCES(1) INFORMATIONS GENERALES:

- (i) DEPOSANT:
- (A) NOM: GENSET SA
 - (B) RUE: 24 RUE ROYALE
 - (C) VILLE: PARIS
 - (E) PAYS: FRANCE
 - (F) CODE POSTAL: 75008
- (ii) TITRE DE L' INVENTION: DEFENSINE HUMAINE DEF-X, GENE ET CDNA, COMPOSITION LES CONTENANT ET APPLICATIONS AU DIAGNOSTIC ET AL LA THERAPIE
- (iii) NOMBRE DE SEQUENCES: 6
- (iv) ADDRESS DE LA CORRESPONDENCE :
- Lola A. Bartoszewicz
 - Sim & McBurney
 - 330 University Avenue, 6th Floor
 - Toronto, Ontario M5G 1R7
- (v) FORME DECHIFFRABLE PAR ORDINATEUR:
- (A) ORDINATEUR: IBM PC compatible
 - (B) SYSTEME D' EXPLOITATION: PC-DOS/MS-DOS
 - (C) LOGICIEL: PatentIn Release #1.0, Version #1.25 (EPO)
- (vi) DATE DE LA DEMANDE AU COURANT :
- (A) NUMÉRO DE LA DEMANDE :
 - (B) DATE DE DÉPOT :
 - (C) CLASSIFICATION :
- (vii) DONNÉES RELATIVES À LA PRIORITÉ :
- (A) NUMÉRO DE LA DEMANDE : FR 97/10823
 - (B) DATE DE DÉPOT : 28 AOÛT 1998
 - (C) CLASSIFICATION :
- (viii) INFORMATION DE LA MANDATAIRES :
- (A) NOM : Lola A. Bartoszewicz
 - (B) NUMÉRO DE REFERENCE : 10488-13 LAB

1a

(2) INFORMATIONS POUR LA SEQ ID NO: 1:

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 4415 PAIRES DE BASE
 - (B) TYPE: NUCLEOTIDE
 - (C) NOMBRE DE BRINS: DOUBLE
 - (D) CONFIGURATION: LINEAIRE
- (ii) TYPE DE MOLECULE: ADN
- (vi) ORIGINE:
 - (A) ORGANISME: Homo sapiens
- (ix) CARACTERISTIQUE:
 - (A) NOM/CLE: Exon 1
 - (B) EMBLACEMENT: 1836..1874
- (ix) CARACTERISTIQUE:
 - (A) NOM/CLE: Exon 2
 - (B) EMBLACEMENT: 3394..3577
- (ix) CARACTERISTIQUE:
 - (A) NOM/CLE: Exon 3
 - (B) EMBLACEMENT: 4161..4380

(ix) CARACTERISTIQUE:

(A) NOM/CLE: start CDS

(B) EMPLACEMENT: 3406..3408

(ix) CARACTERISTIQUE:

(A) NOM/CLE: stop CDS

(B) EMPLACEMENT: 4276..4278

(ix) CARACTERISTIQUE:

(A) NOM/CLE: site de polyAdenylation

(B) EMPLACEMENT: 4374..4379

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACACCATTG TCTTCATGTA ACCCCATTAG CTATACCCTC TAGTGCAAGG AAACCATAGG	60
GCCTAGGTCA CACCATGAGG CTGCNCTTAC AAGTTATGCA AAAACTATGG ACTTGGGAGA	120
CCTGTGCGTA ACAACATCAC ACNCCAAATT TAACCAGCTC TCCCCATAAC AGCACGCTCA	180
TGTGTTACTG AGGAAATGCC TGTGGATTGG AGTGTGTTCT GTGTGCAGGA GGCTGGTCCA	240
GGTTTCACTT CTGCAGGACA CTGGACGTTT CCCAAAACCA GCAGACTTTC CCCACGTGCA	300
CACACACCCC TTCTCATTTT GCCTCTACAT CCATATCCAC TGGGCCCTTC AGGCACCTAC	360
TAA TGCCCTA GAACCTAAA CCATCATCTG GGGCCCAGTT CCCTGAATGG CCCTAATCTC	420
TTCCTCTGCT GGAATGAGTC CAGTGCCAC TTCCTCCAC GGTGAATTG CTGGGCTGCT	480
ACAGATCAGG AACTCACTGC TTCCTCATAG GGGCAGCCGA CTTCACTGCT CTGCAACAGC	540
GACCACCCCT AGCGAGGCTT GAGATGCCTC TTGCCTCCTT AAGACTGAGG GAGACGCTTC	600
AGCTCTCACT CCACTGCCCC AAGTCCTCCA CAGCGCGGTG CCTGCTGCCT TCACACAGAG	660
CTGCAGGGGN AGGTCCTGTG TATCCGGCCT GCTGGACCAG CGCTGTGCAC AACCTCCCA	720
TGGCAACAGT GGCTGCCCCG CCTGCACACT GGGCTTGGCA ACCTCGCTGT AGGTATTTAT	780
TCCCTCAGGA GTGACTGCAT TCTTTTCCCA TTTCCAGAAA ACTGATGCCA TTTACCTCAC	840
TATGAGGAGG AGGAGGAGGA GGAGGGTGGA GAGTGGTACA TTTTAAAATG TGCATATTC	900
TCCCTAGGAC TCCCCCTCAA ATAACCCAGG AGGGACCATA CCAGCTCATT CCTGTGTATC	960
CCAAGCATAN GAGTAATCAT CCCACTCATG CTGAGTGTAT GGTGGCCATT AAGCCTGCCC	1020
TGA ACTGGCT TTACAACAAG GTGTTTGAGC ACACAGCACC GTCTTGCTGC CACCTTGGCC	1080
CCCTCCCTTG TGAGACCTCT GAGACACATT NAGGTCTCAC CTAAAAATCT CAGGATTTCT	1140
AGGCCCAAAN CGGTCCTAAA AAATTGTTCA GTCTGAACTC TCTAAGGTCA AGAGAAGAGG	1200

TGGTTGCTCC CTCTAAGAAA CCACATGTTG CATGTACATC CTTAATTCCG GAAAGTCCAA	1260
CAAACTGCCC CTGCTTAGCA ACACAAGCCG AGGTGGTACT CCTCTCACCC GGGCATTCTC	1320
CAACACACCT GTTTGTCCAA ACAGCTTTGA TTTGTTTTTA TAGTTGGACC CCAGGTTCCT	1380
AGGAGGCTGG TTCAGGCCAT ATTCCAAATC CTCATCTGTG TGTGAGTGGC ATTCTTAGCC	1440
TAGCCTCCTT ACAGGGTGGA TACTATGATA CACAGCCAGG CTGTCCCAGT GGCTTTCAAT	1500
ATTCTTTTGG TCCAGATAGT TCAGCCTCAG CACCAGTGTA GGCATCACAG GGTCATTGT	1560
CTTAGGAGTC ATGGAGAATT CATAGTTGGT AGCTACCTGG GCCTGGCCAG GGCTGACCAT	1620
AGACAAGGCA TCCCTCTGTG AACTCCTATT TTAATGCCAG CTTCCCAACA AATTTCTCAA	1680
CTGCTCTTAC CAGCAGGTAT TTA/ACTACT CAATAGAAAG TAACCCTGAA AATTAGGACA	1740
CCTGTTCCCA AAAGACCCTT AAATAGGGGA AGTCCTTTCN CTGCTTGTGC ACAGCTGCTG	1800
ATGTGGCAAC ATGAGGCCTG GGACAGGGGA CTGTCTCTG CCCACTCTGG TAGCCTCAGG	1860
TAGCTTAACA ATCTGTCAGT AATACAATAC AAAACTTAAA CTTTCATACT GCGGTTCCAC	1920
CCAGGAAGCT GTGTTCCCAA TCTGACCCGT GATTATGGGG CCACCTCAGA GGNACCCAG	1980
TGAGGGAATA TTTTGCCATC TGGGACTGTT GGTGCTGGG GGCAGTGGCT ATGAGCTCAG	2040
TTAATAA/CT CAAGCAGTTT CCTTCCAAAC ACACATGTCC TACTTAACGT GTCC/ACAGA	2100
GATGATCATA CTCATANGCT GCTAAAACAT TANTTTTATT TTGAGAAAAG TCTATTCATG	2160
TTCTTGGCCC ATGGAGTTTT CATTNATTA NTTTATTTAT TTTGCAGAGA TGGAGTCTCA	2220
CTATGTTGCT CAAGCTGGTC TCCA/CTCCT GGGCTCAAGC GATCTTCCTA CTTTGGCCTT	2280
TGAAAGCGCT GAGATTGCCT GTGTGAGCCA TCATGGGGGC TCACTGGCCC ACTGATTAAT	2340
CAGATTAATT GTTTTTTGCT ATTGAANTTG TTTGACTTCC TTGTATATTC GGATATTTAC	2400
CCATTCTAAC ACGTAGGGTT TGCAAATATT TTCTCTCATG TTCTGTGTTG CCTTTTCACT	2460
CAGTTGATGG TTTCTTTTGC TGTGCAGGTG CTTTAGTGTT CAACGCAGCC CCGCTTGTCT	2520
ATTTTCCATT TTATTGCCTG TCCCTTTGAT GTCATAGCCA AGAAATAATT GCCCAGATTA	2580
ATGTCAAAAA GCTTTATCCC TATATATTCT TCTAGTAGTT TATGGTTTCA GATCTTATGT	2640
TTAGGTCTTC AATCCATTGA GTTGATTTTT GTATGTGGTA TAAGAAAAAA GACCACATGT	2700
ATACATATCT CAAATTCTAA GGTAGTATAT ATTAGACACA TACAATGTGT CTATTTACAC	2760
ACATTGAGCT GAAAATAATA AACATATTTT TATCTTTCAA TCAACTCTAT CTCTATCTCA	2820

CTGAACTTGT TTCACCTATA GCCTGATGAG GTTGCTGTCC TCTCTACCCC AGCTCCTATA	2880
GGAGACTGCT CATCCCCTAA CCTCAAAAAC CCCTTCATGA GGGTGATAAT GCCCTTGAAAT	2940
CCTGCAATGA ATTAGTTCTC TACTACAGTG GAATTCAGGT CTGTTATGAG GGTCTGGATC	3000
TCTGAAGAGA AGAGCTCTCA TTTTCAGAAA ATAAGCAGGA TTTATTCCCT GAAATTACTG	3060
AATTAAATCA CTGTTTCGAT TACTTTTTGC AATATTAAA GTAAATATTT AACAGGTAA	3120
AACAGAAAT AATGGTAGGG TCCTTATCAT CACCGTGAAT TCCAAGCTAG CATAGACACT	3180
AAACCTAGAG ATTCACACTA GAATGAAAGC TGGGAGAGCA GAGGAGTCTC AGAAGGATGT	3240
GGAGGCCAAT GGACACCTGC AACCTCTCCA ACGAAATGCC TACCTCCTCT CACTGCAGCA	3300
TCCATCTCTG AGCCTTCTCG CAGCAGAGCT ATAAATTCAG CCTGGCTCCT CCGTTCCCAC	3360
ACATCCACTC CTGCTCTCCC TCCTCTCCTC CAGGTGACTA CAGTTATGAG GACCCTCACC	3420
CTCCTCTCTG CCTTTCTCCT GGTGGCCCTT CAGGCCTGGG CAGAGCCGCT CCAGGCAAGA	3480
GCTCATGAGA TCCCAGCCCA GAAGCAGCCT CCAGCAGATG ACCAGGATGT GGTCATTTAC	3540
TTTTCAGGAG ATGACAGCTG CTCTCTTCAG GTTCCAGGTG AGAGATGCCA GCATGCAGAG	3600
CTACAGACTA GACAGAAAGG CAGGAGACAG GCTCTGGAAT TGGATCTCAG TGGCAGATGT	3660
CACTTAGGTG GCTATACTTA ACATCTCTGG TCCTGGATTT TCTCATATCT AATGGAATA	3720
GAGAACCAA GAAATCTAAG AGATTTTTCT TTCTCCAAA ACTTGATTCC AAGATATGAC	3780
TGTGAAATTC ACTAGATTTA AGATATAAGG AGATGCTACC TAGTTCCTTC TGGAGCCAGA	3840
CAAACAAGCT TAAGTATATA GGAAAATATT TCACCCTGTC TATATAGGAG GTTTTAGAAC	3900
CTGGAGAGGA GCCTAAGAAT GTGTTTCAGGT GTGTGTGTGA TGGGCAGGAA TGCAGAAAAG	3960
TGAAGCAAAG GAGAATGAGT CTCGAATCCT GTGTGACCAG CACTGCTCTG TGTATTTATT	4020
CCTATTGACT GAGATTGTTT GTGCTACCGG CTGTAATACA GCCAACATCA CTCATCAGCC	4080
AACATGTGAC TTCTCCAAGA TTCCCTTTAC CACCCACTGC TGNACCCCGT ACTCAGTTTC	4140
TGATGCTCTC TCTGGGTCCC CAGGCTCAAC AAAGGGCTTG ATCTGCCATT GCAGAGTACT	4200
ATACTGCATT TTTGGAGAAC ATCTTGGTGG GACCTGCTTC ATCCTTGGTG AACGCTACCC	4260
AATCTGCTGC TACTAAGCTT GCAGACTAGA GAAAGAGAGT TCATAATTTT CTTTGAGCAT	4320
TAAAGGGAAT TGTTATTCTT ATACCTTGTC CTCGATTTC TGTCCTCATC CCAAATAAAT	4380
ACTTGGTAAC ATGATTTCGG GGTTTTTTTT TTTT	4415

WO 99/11663

PCT/FR98/01864

5

(2) INFORMATIONS POUR LA SEQ ID NO: 2:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 453 PAIRES DE BASE
- (B) TYPE: NUCLEOTIDE
- (C) NOMBRE DE BRINS: DOUBLE
- (D) CONFIGURATION: LINEAIRE

(ii) TYPE DE MOLECULE: ADNc

(vi) ORIGINE:

- (A) ORGANISME: Homo sapiens

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

CTCTGCCCCAC TCTGGTAGCC TCACGTAGCT TAACAATCTG TGACTACAGT T ATG AGG	57
Met Arg	
1	
ACC CTC ACC CTC CTC TCT GCC TTT CTC CTG GTG GCC CTT CAG GCC TGG	105
Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln Ala Trp	
5 10 15	
GCA GAG CCG CTC CAG GCA AGA GCT CAT GAG ATG CCA GCC CAG AAG CAG	153
Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln Lys Gln	
20 25 30	
CCT CCA GCA GAT GAC CAG GAT GTG GTC ATT TAC TTT TCA GGA GAT GAC	201
Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly Asp Asp	
35 40 45 50	
AGC TGC TCT CTT CAG GTT CCA GGC TCA ACA AAG GGC TTG ATC TGC CAT	249
Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile Cys His	
55 60 65	
TGC AGA GTA CTA TAC TGC ATT TTT GGA GAA CAT CTT GGT GGG ACC TGC	297
Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly Thr Cys	
70 75 80	
TTC ATC CTT GGT GAA CGC TAC CCA ATC TGC TGC TAC TAA GCTTGCAGAC	346
Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr *	
85 90 95	
TAGAGAAAAA GAGTTCATAA TTTTCTTTGA GCATTAAAGG GAATTGTTAT TCTTATACCT	406
TGTCCTCGAT TTCCTGTCCT CATCCCAAAT AAATACTTGG TAACATG	453

(2) INFORMATIONS POUR LA SEQ ID NO: 3:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 94 ACIDES AMINES
- (B) TYPE: ACIDE AMINE
- (C) NOMBRE DE BRINS: SIMPLE
- (D) CONFIGURATION: LINEAIRE

(ii) TYPE DE MOLECULE: PROTEINE

(vi) ORIGINE:

- (A) ORGANISME: Homo sapiens

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: PEPTIDE SIGNAL
- (B) EMBLACEMENT: 1..19

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: REGION PRO
- (B) EMBLACEMENT: 20..63

(ix) CARACTERISTIQUE:

- (A) NOM/CLE: PEPTIDE MATURE
- (B) EMBLACEMENT: 64..94

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

Met Arg Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln
1 5 10 15

Ala Trp Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln
20 25 30

Lys Gln Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly
35 40 45

Asp Asp Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile
50 55 60

Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly
65 70 75 80

Thr Cys Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr
85 90

Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu
35 40

WO 99/11663

PCT/FR98/01864

8

(2) INFORMATIONS POUR LA SEQ ID NO: 6:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 31 ACIDES AMINES
- (B) TYPE: ACIDE AMINE
- (C) NOMBRE DE BRINS: SIMPLE
- (D) CONFIGURATION: LINEAIRE

(ii) TYPE DE MOLECULE: PEPTIDE MATURE

(vi) ORIGINE:

- (A) ORGANISME: Homo sapiens

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

Ile	Cys	His	Cys	Arg	Val	Leu	Tyr	Cys	Ile	Phe	Gly	Glu	His	Leu	Gly
1				5				10						15	
Gly	Thr	Cys	Phe	Ile	Leu	Gly	Glu	Arg	Tyr	Pro	Ile	Cys	Cys	Tyr	
			20					25					30		

REVENDICATIONS

1. Polypeptide isolé choisi parmi les polypeptides suivants :

- a) polypeptide dont la séquence d'acides aminés est la séquence SEQ ID No. 3 ;
- 5 b) polypeptide présentant au minimum 80 % d'identité avec le polypeptide dont la séquence d'acides aminés est la séquence SEQ ID No. 3 ;
- c) polypeptide dont la séquence d'acides aminés est la séquence d'acides aminés d'un fragment d'au moins 10 acides aminés consécutifs d'un polypeptide dont la séquence d'acides aminés présente au minimum 90 % d'identité avec la séquence SEQ ID No.
- 10 3 ;
- d) polypeptide comprenant au moins un fragment tel que défini en c).

2. Polypeptide selon la revendication 1, caractérisé en ce qu'il est constitué de l'un au moins des fragments suivants :

- a) peptide signal dont la séquence d'acides aminés est la séquence SEQ ID No. 4 ;
- 15 b) région pro dont la séquence d'acides aminés est la séquence SEQ ID No. 5 ;
- c) peptide mature dont la séquence d'acides aminés est la séquence SEQ ID No. 6 ; ou
- d) fragment présentant au minimum 90 % d'identité avec un peptide selon a), b), ou c).

3. Polypeptide isolé choisi parmi les polypeptides suivants :

- a) polypeptide dont la séquence d'acides aminés est la séquence SEQ ID No. 6 ;
- 20 b) polypeptide présentant au minimum 80 % d'identité avec le polypeptide dont la séquence d'acides aminés est la séquence SEQ ID No. 6 ;
- c) polypeptide dont la séquence d'acides aminés est la séquence d'acides aminés d'un fragment d'au moins 10 acides aminés consécutifs d'un polypeptide tel que défini en a) ou b) ;
- 25 d) polypeptide comprenant au moins un fragment tel que défini en c).

4. Polypeptide selon l'une des revendications 1 ou 3, caractérisé en ce que ledit fragment comprend au moins 15 acides aminés consécutifs.

5. Acide nucléique codant pour un polypeptide selon l'une des revendications 1 à 4.

30 6. Acide nucléique choisi parmi les acides nucléiques suivants :

- a) acide nucléique de séquence SEQ ID No. 1 ;
- b) acide nucléique de séquence SEQ ID No. 2 ;
- c) acide nucléique présentant 80 % d'identité par rapport aux acides nucléiques selon a) ou b) ;

d) fragment nucléique de séquence SEQ ID No. 1 comprenant au moins un exon du gène Def X.

7. Vecteur de clonage ou d'expression dans une cellule hôte appropriée d'une séquence nucléotidique, caractérisé en ce qu'il comporte une séquence selon l'une des revendications 5 et 6.

8. Vecteur selon la revendication 7, caractérisé en ce qu'il comporte les éléments assurant l'expression de ladite séquence dans ladite cellule hôte.

9. Cellule transformée par un vecteur selon l'une des revendications 7 et 8.

10. Cellule selon la revendication 9, caractérisée en ce qu'il s'agit d'une cellule procaryote.

11. Cellule selon la revendication 9, caractérisée en ce qu'il s'agit d'une cellule eucaryote.

12. Procédé de production d'un polypeptide selon l'une des revendications 1 à 3, caractérisé en ce qu'on cultive une cellule selon l'une des revendications 9 à 11 et en ce que l'on récupère le polypeptide produit.

13. Polypeptide susceptible d'être obtenu par la mise en œuvre du procédé selon la revendication 12.

14. Polypeptide selon l'une des revendications 1 à 4, caractérisé en ce qu'il est obtenu par synthèse chimique.

15. Anticorps monoclonal ou polyclonal ou un de leurs fragments, anticorps chimériques, caractérisé en ce qu'il est capable de reconnaître spécifiquement un polypeptide selon l'une des revendications 1 à 4, et 14.

16. Anticorps selon la revendication 15, caractérisé en ce qu'il est marqué.

17. Sonde ou amorce oligonucléotidique, caractérisée en ce qu'elle est constituée d'un acide nucléique selon l'une des revendications 5 et 6.

18. Sonde selon la revendication 17, caractérisée en ce qu'elle est marquée.

19. Utilisation d'un polypeptide selon l'une des revendications 1 à 4, 13 et 14 comme agent antimicrobien et/ou antiparasitaire.

20. Utilisation d'un polypeptide selon l'une des revendications 1 à 4, 13 et 14 comme agent cytotoxique, notamment à visée anticancéreuse.

21. Utilisation d'un polypeptide selon l'une des revendications 1 à 4, 13 et 14 comme agent de modulation des processus de l'inflammation, de réparation tissulaire et de régulation endocrine, notamment corticostatique.

22. Composition pour usage topique externe, caractérisée en ce qu'elle comporte au moins un polypeptide selon l'une des revendications 1 à 4, 13 et 14.

23. Composition selon la revendication 22, caractérisée en ce qu'il s'agit d'une composition cosmétique.

24. Composition pharmaceutique comprenant un polypeptide selon l'une des revendications 1 à 4, et 13 et 14.

5 25. Composition pharmaceutique comprenant un vecteur selon l'une des revendications 7 et 8, capable d'exprimer *in vivo* un polypeptide selon l'une des revendications 1 à 4.

26. Composition pharmaceutique selon l'une des revendications 22, 24 et 25, caractérisée en ce qu'elle comprend un véhicule pharmaceutiquement acceptable.

10 27. Composition pharmaceutique selon l'une des revendications 22, 24 à 26, destinée à la prévention et/ou au traitement des infections microbiennes ou parasitaires.

28. Composition pharmaceutique selon la revendication 27, caractérisée en ce que les infections microbiennes ou parasitaires sont des infections d'origines bactériennes, de bactéries Gram-positives ou Gram-négatives, mycobactériennes, 15 fongiques, ou liées à des spirochètes.

29. Composition pharmaceutique selon la revendication 27, caractérisée en ce que les infections virales sont des infections liées à des virus à enveloppe, notamment les virus HSV et HIV.

20 30. Composition pharmaceutique selon l'une des revendications 22, 24 à 26, destinée à la prévention et/ou au traitement de cancers, notamment les mélanomes.

31. Composition pharmaceutique selon la revendication 30, caractérisée en ce que le cancer est le cancer du foie, de la prostate, du poumon non à petites cellules ou le carcinome colorectal.

25 32. Composition pharmaceutique selon l'une des revendications 22, 24 à 26 destinée à augmenter les défenses immunitaires, à augmenter les défenses immunitaires en cas d'immunodéficience acquise ou à prévenir l'immunodéficience, notamment pour le traitement du psoriasis.

30 33. Composition pharmaceutique selon l'une des revendications 22, 24 à 26, destinée à moduler les processus inflammatoires, notamment dans les cas de maladies à inflammation chronique.

34. Méthode de diagnostic d'une immunodéficience et/ou d'une prédisposition à des affections de type cancer, caractérisée en ce qu'on met en évidence dans un prélèvement de patient la présence d'une défensine anormale et/ou d'une séquence codant pour une défensine anormale.

35. Méthode de diagnostic d'infections dues à des micro-organismes ou liées à un déficit immunitaire ou à un phénomène inflammatoire, caractérisée en ce qu'elle comprend le dosage d'un polypeptide selon l'une des revendications 1 à 4 ou d'un acide nucléique selon l'une des revendications 5 et 6 dans un échantillon biologique et la
5 comparaison du résultat dudit dosage obtenu avec la quantité dudit polypeptide, respectivement dudit acide nucléique, présente normalement dans un échantillon biologique équivalent.

36. Méthode de diagnostic d'inflammation, d'une immunodéficience et/ou d'une prédisposition à des affections de type cancer, caractérisée en ce qu'elle comprend le
10 dosage d'un polypeptide selon l'une des revendications 1 à 4 ou d'un acide nucléique selon l'une des revendications 4 et 5 dans un échantillon biologique et la comparaison du résultat dudit dosage obtenu avec la quantité dudit polypeptide, respectivement dudit acide nucléique, présente normalement dans un échantillon biologique équivalent.

37. Méthode de diagnostic selon l'une des revendications 34 à 36, caractérisée en
15 ce qu'elle met en œuvre un anticorps selon l'une des revendications 15 et 16.

38. Méthode de diagnostic selon l'une des revendications 34 à 36, caractérisée en ce qu'elle met en œuvre une sonde et/ou une amorce oligonucléotidique selon l'une des revendications 17 et 18.

39. Kit ou nécessaire de diagnostic pour la détermination d'une infection
20 microbienne ou parasitaire, d'une inflammation, d'une immunodéficience et/ou de prédisposition à des affections de type cancer, caractérisé en ce qu'il comprend un anticorps selon l'une des revendications 15 et 16.

40. Kit ou nécessaire de diagnostic pour la détermination d'une infection
25 microbienne ou parasitaire, d'une inflammation, d'une immunodéficience et/ou de prédisposition à des affections de type cancer, caractérisé en ce qu'il comprend une sonde et/ou une amorce selon l'une des revendications 17 et 18.

41. Utilisation d'un polypeptide selon l'une des revendications 1 à 4, 13 et 14, comme pesticide, notamment pour la culture de végétaux d'intérêt industriel.

1/16

ACACCATTG	TCTTCATGTA	ACCCCATAG	CTATACCCTC	TAGTGCAAGG	AAACCATAGG
10	20	30	40	50	60
GCCTAGGTCA	CACCATGAGG	CTGCNCTTAC	AAGTTATGCA	AAACTATGG	ACTTGGGAGA
70	80	90	100	110	120
CCTGTGCGTA	ACAACATCAC	ACNCCAAATT	TAACCAGCTC	TCCCCATAAC	AGCACGCTCA
130	140	150	160	170	180
TGTGTTACTG	AGGAAATGCC	TGTGGATTGG	AGTGTGTTCT	GTGTGCAGGA	GGCTGGTCCA
190	200	210	220	230	240
GGTTTCACTT	CTGCAGGACA	CTGGACGTTT	CCCAAACCA	GCAGACTTTC	CCCACGTGCA
250	260	270	280	290	300
CACACACCCC	TTCTCATTTT	GCCTCTACAT	CCATATCCAC	TGGGCCCTTC	AGGCACCTAC
310	320	330	340	350	360
TAATGCCCTA	GAACCTAAAA	CCATCATCTG	GGGCCAGTT	CCCTGAATGG	CCCTAATCTC
370	380	390	400	410	420
TTCCTCTGCT	GGAATGAGTC	CAGTGCCAC	TTCCTCCAAC	GGTGAAATTG	CTGGGCTGCT
430	440	450	460	470	480
ACAGATCAGG	AACTCACTGC	TTCCTCATAG	GGGCAGCCGA	CTTCACTGCT	CTGCAACAGC
490	500	510	520	530	540
GACCACCCCT	AGCGAGGCTT	GAGATGCCTC	TTGCCTCCTT	AAGACTGAGG	GAGACGCTTC
550	560	570	580	590	600
AGCTCTCACT	CCACTGCCCC	AAGTCCTCCA	CAGCGCGGTG	CCTGCTGCCT	TCACACAGAG
610	620	630	640	650	660
CTGCAGGGGN	AGGTCCTGTG	TATCCGGCCT	GCTGGACCAG	CGCTGTGCAC	AACCCTCCCA
670	680	690	700	710	720
TGGCAACAGT	GGCTGCCCCG	CCTGCACACT	GGGCTTGGCA	ACCTCGCTGT	AGGTATTTAT
730	740	750	760	770	780
TCCCTCAGGA	GTGACTGCAT	TCTTTTCCCA	TTTCCAGAAA	ACTGATGCCA	TTTACCTCAC
790	800	810	820	830	840
TATGAGGAGG	AGGAGGAGGA	GGAGGGTGGA	GAGTGGTACA	TTTTAAAATG	TGCACTATTC
850	860	870	880	890	900
TCCCTAGGAC	TCCCCCTCAA	ATAACCCAGG	AGGGACCATA	CCAGCTCATT	CCTGTGTATC
910	920	930	940	950	960
CCAAGCATAN	GAGTAATCAT	CCCACTCATG	CTGAGTGTAT	GGTGGCCATT	AAGCCTGCCC
970	980	990	1000	1010	1020

Figure 1

2/16

```

TGAAGTGGCT TTAGAACAAG GTGTTTGAGC ACACAGCACC GTCTTGCTGC CACCTTGGCC
  1030      1040      1050      1060      1070      1080

CCCTCCCTTG TGAGACCTCT GAGACACATT NAGGTCTCAC CTAAAAATCT CAGGATTTCT
  1090      1100      1110      1120      1130      1140
AGGCCCAAN CGGTCCTAA AATTGTTC A GTCTGAACTC TCTAAGGTCA AGAGAAGAGG
  1150      1160      1170      1180      1190      1200

TGGTTGCTCC CTCTAAGAAA CCACATGTTG CATGTACATC CTTAATTCCG GAAAGTCCAA
  1210      1220      1230      1240      1250      1260

CAACCTGCC CTGCTTAGCA ACACAAGCCG AGGTGGTACT CCTCTCACCC GGGCATTCTC
  1270      1280      1290      1300      1310      1320

CAACACACCT GTTTGTCCAA ACAGCTTTGA TTTGTTTTTA TAGTTGGACC CCAGGTTCCC
  1330      1340      1350      1360      1370      1380

AGGAGGCTGG TTCAGGCCAT ATTCCAAATC CTCATCTGTG TGTGAGTGGC ATTCTTAGCC
  1390      1400      1410      1420      1430      1440

TAGCCTCCTT ACAGGGTGGA TACTATGATA CACAGCCAGG CTGTCCCAGT GGCTTTCAAT
  1450      1460      1470      1480      1490      1500

ATTCTTTTGG TCCAGATAGT TCAGCCTCAG CACCAGTGTA GGCATCACAG GGTCAATTGT
  1510      1520      1530      1540      1550      1560

CTTAGGAGTC ATGGAGAATT CATAGTTGGT AGCTACCTGG GCCTGGCCAG GGCTGACCAT
  1570      1580      1590      1600      1610      1620

AGACAAGGCA TCCCTCTGTG AACTCCTATT TTAATGCCAG CTTCCCAACA AATTTCTCAA
  1630      1640      1650      1660      1670      1680
                                CAAT box
                                ----
CTGCTCTTAC CAGCAGGTAT TTAACTACT CAATAGAAAG TAACCCTGAA AATTAGGACA
  1690      1700      1710      1720      1730      1740
                                TATA box
                                -----
CCTGTTCCCA AAAGACCCTT AAATAGGGGA AGTCCTTTCTN CTGCTTGTGC ACAGCTGCTG
  1750      1760      1770      1780      1790      1800

                                |->mRNA -----
ATGTGGCAAC ATGAGGCCTG GGACAGGGGA CTGTCCTCTG CCCACTCTGG TAGCCTCACG
  1810      1820      1830      1840      1850      1860

                                Spsite
-- exon 1 ---->#####
TAGCTTAACA ATCTGTCACT AATACAATAC AAAACTTAAA CTTTCATACT GCGGTTCCAC
  1870      1880      1890      1900      1910      1920

CCAGGAAGCT GTGTTCCCAA TCTGACCCGT GATTATGGGG CCACCTCAGA GGGNACCCAG
  1930      1940      1950      1960      1970      1980

```

Figure 1(suite)

3/16

```

TGAGGGAATA TTTTGCCATC TGGGACTGTT GGTTCCTGGG GGCAGTGGCT ATGAGCTCAG
  1990      2000      2010      2020      2030      2040

TTAATAAAGT CAAGCAGTTT CCTTCCAAAC ACACATGTCC TACTTAAAGT GTCCAACAGA
  2050      2060      2070      2080      2090      2100

-----
GATGATCATA CTCATANGCT GCTAAACAT TANTTTTATT TTGAGAAAAG TCTATTCATG
  2110      2120      2130      2140      2150      2160
-----
TTCTTGGCCC ATGGAGTTTT CATTNNATTA NTTTATTTAT TTTGCAGAGA TGGAGTCTCA
  2170      2180      2190      2200      2210      2220

-----
CTATGTTGCT CAAGCTGGTC TCCAACCTCT GGGCTCAAGC GATCTTCCTA CTTTGGCCTT
  2230      2240      2250      2260      2270      2280

-----
TGAAAGCGCT GAGATTGCCT GTGTGAGCCA TCATGGGGGC TCACTGGCCC ACTGATTAAT
  2290      2300      2310      2320      2330      2340

CAGATTAATT GTTTTTTGCT ATTGAANTTG TTTGACTTCC TTGTATATTC GGATATTTAC
  2350      2360      2370      2380      2390      2400

CCATTCTAAC ACGTAGGGTT TGCAAATATT TTCTCTCATG TTCTGTGTTG CCTTTTCACT
  2410      2420      2430      2440      2450      2460

CAGTTGATGG TTTCTTTTGC TGTGCAGGTG CTTTAGTGTT CAACGCAGCC CCGCTTGTCT
  2470      2480      2490      2500      2510      2520

ATTTTCCATT TTATTGCCTG TCCCTTTGAT GTCATAGCCA AGAAATAATT GCCCAGATTA
  2530      2540      2550      2560      2570      2580

ATGTCAAAAA GCTTTATCCC TATATATTCT TCTAGTAGTT TATGGTTTCA GATCTTATGT
  2590      2600      2610      2620      2630      2640

TTAGGTCTTC AATCCATTGA GTTGATTTTT GTATGTGGTA TAAGAAAAAA GACCACATGT
  2650      2660      2670      2680      2690      2700

ATACATATCT CAAATTCTAA GGTAGTATAT ATTAGACACA TACAATGTGT CTATTTACAC
  2710      2720      2730      2740      2750      2760

ACATTGAGCT GAAAATAATA AACATATTTT TATCTTTCAA TCAACTCTAT CTCTATCTCA
  2770      2780      2790      2800      2810      2820

CTGAACTTGT TTCACCTATA GCCTGATGAG GTTGCTGTCC TCTCTACCCC AGCTCCTATA
  2830      2840      2850      2860      2870      2880

GGAGACTGCT CATCCCCTAA CCTCAAAAAC CCCTTCATGA GGGTGATAAT GCCCTTGAAT
  2890      2900      2910      2920      2930      2940

```

Figure 1 (suite)

4/16

```

CCTGCAATGA ATTAGTTCTC TACTACAGTG GAATTCAGGT CTGTTATGAG GGTCTGGATC
  2950      2960      2970      2980      2990      3000

TCTGAAGAGA AGAGCTCTCA TTTTCAGAAA ATAAGCAGGA TTTATTCCCT GAAATTACTC
  3010      3020      3030      3040      3050      3060

AATTAAATCA CTGTTTCGAT TACTTTTTGC AATATTAAAA GTAAATATTT AACAGGTAA
  3070      3080      3090      3100      3110      3120

AAACAGAAAT AATGGTAGGG TCCTTATCAT CACCGTGAAT TCCAGCTAG CATAGACACT
  3130      3140      3150      3160      3170      3180

AAACCTAGAG ATTCACACTA GAATGAAAGC TGGGAGAGCA GAGGAGTCTC AGAAGGATGT
  3190      3200      3210      3220      3230      3240

GGAGGCCAAT GGACACCTGC AACCTCTCCA ACGAAATGCC TACCTCCTCT CACTGCAGCA
  3250      3260      3270      3280      3290      3300

TCCATCTCTG AGCCTTCTCG CAGCAGAGCT ATAAATTCAG CCTGGCTCCT CCGTTCCCAC
  3310      3320      3330      3340      3350      3360

                                     Spsite      CDS start
                                     ###<-----*-----
ACATCCACTC CTGCTCTCCC TCCTCTCCTC CAGGTGACTA CAGTTATGAG GACCCTCACC
  3370      3380      3390      3400      3410      3420

----- Exon 2 -----
CTCCTCTCTG CCTTTCTCCT GGTGGCCCTT CAGGCCTGGG CAGAGCCGCT CCAGGCAAGA
  3430      3440      3450      3460      3470      3480

-----
GCTCATGAGA TGCCAGCCCA GAAGCAGCCT CCAGCAGATG ACCAGGATGT GGTCAATTTAC
  3490      3500      3510      3520      3530      3540

                                     Spsite
----->### ###
TTTTCAGGAG ATGACAGCTG CTCTCTTCAG GTTCCAGGTG AGAGATGCCA GCATGCAGAG
  3550      3560      3570      3580      3590      3600

CTACAGACTA GACAGAAGGA CAGGAGACAG GCTCTGGAAT TGGATCTCAG TGGCAGATGT
  3610      3620      3630      3640      3650      3660

CACTTAGGTG GCTATACTTA ACATCTCTGG TCCTGGATTT TCTCATATCT AAATGGAATA
  3670      3680      3690      3700      3710      3720

GAGAACCAAA GAAATCTAAG AGATTTTCTT TTCTCCAAA ACTTGATTCC AAGATATGAC
  3730      3740      3750      3760      3770      3780

TGTGAAATTC ACTAGATTTA AGATATAAGG AGATGCTACC TAGTTCCTTC TGGAGCCAGA
  3790      3800      3810      3820      3830      3840

```

Figure 1(suite)

5/16

```
CAAACAAGCT TAAGTATATA GGAAAATATT TCACCCTGTC TATATAGGAG GTTTTAGAAC
    3850          3860          3870          3880          3890          3900

CTGGAGAGGA GCCTAAGAAT GTGTTTCAGGT GTGTGTGTGA TGGGCAGGAA TGCAGAAAAG
    3910          3920          3930          3940          3950          3960

TGAAGCAAAG GAGAATGAGT CTCGAATCCT GTGTGACCAG CACTGCTCTG TGTATTTATT
    3970          3980          3990          4000          4010          4020

CCTATTGACT GAGATTGTTT GTGCTACCGG CTGTAATACA GCCAACATCA CTCATCAGCC
    4030          4040          4050          4060          4070          4080

AACATGTGAC TTCTCCAAGA TTCCCTTTAC CACCCACTGC TGNACCCCGT ACTCAGTTTC
    4090          4100          4110          4120          4130          4140

                                Spsite
                                ###<-----
TGATGCTCTC TCTGGGTCCC CAGGCTCAAC AAAGGGCTTG ATCTGCCATT GCAGAGTACT
    4150          4160          4170          4180          4190          4200

----- Exon 3 -----
ATACTGCATT TTTGGAGAAC ATCTTGGTGG GACCTGCTTC ATCCTTGGTG AACGCTACCC
    4210          4220          4230          4240          4250          4260

                                CDS stop
                                ***-----
AATCTGCTGC TACTAAGCTT GCAGACTAGA GAAAAGAGT TCATAATTTT CTTTGAGCAT
    4270          4280          4290          4300          4310          4320

                                                                Poly Ad
                                                                *----->
TAAAGGGAAT TGTTATTCTT ATACCTTGTC CTCGATTTC TGTCTCATC CCAAATAAAT
    4330          4340          4350          4360          4370          4380

ACTTGATAAC ATGATTTCAG GGTTTTTTTT TTTT
    4390          4400          4410
```

Figure 1(suite)

6/16

	10	20	30	40	50
DEF4	GGATCCCCATTTGTCTTCAGTGTAAACC-ATTAGTTAAACCGCCTACTGCAAGGAAACCA				
DEFX	ACACCATTGTCTTCA-TGTAAACCCATTAGCTATACCTCTAGTGCAAGGAAACCA				
	10	20	30	40	50
DEF4	60	70	80	90	100
DEF4	CAAGGCTTGGATCAGATCATGAGGCTGCCCT-ACAAGTTATGCCAAAAAATATGGACTTG				
DEFX	TAGGGCCTAGGTCACACCATGAGGCTGCNCTTACAAGTTATGC-AAAAACCTATGGACTTG				
	60	70	80	90	100
DEF4	120	130	140	150	160
DEF4	GAAGACCTGTCTGTTATAATATCACAC-CCAAATCTAACCAGCTCTGCCAATAACAGCTC				
DEFX	GGAGACCTGTGCGTAACAACATCACACNCCAAATTTAACCAGCTCTCCCCATAACAGCAC				
	120	130	140	150	160
DEF4	180	190	200	210	220
DEF4	TCTCCTATGTTACTAGGAAAATGCCTATGGATTGGAGTGTGTTCTGTGTGCAGGAGGCTG				
DEFX	GCTCATGTGTTACTGAGGAAATGCCTGTGGATTGGAGTGTGTTCTGTGTGCAGGAGGCTG				
	180	190	200	210	220
DEF4	240	250	260	270	280
DEF4	GTCCAGGTTTCACTTCTGCAGGACACTGGACATC-CCCACAACCAGACCTTCCCCAC				
DEFX	GTCCAGGTTTCACTTCTGCAGGACACTGGACGTTTCCCAAAACCAGCAGACTTTCCCCAC				
	240	250	260	270	280
DEF4	300	310	320	330	340
DEF4	GTGCACACACACCCCTTCTCATTTTGCCTCTACATCCATATCCACTGGGCCCTTCAGGCA				
DEFX	GTGCACACACACCCCTTCTCATTTTGCCTCTACATCCATATCCACTGGGCCCTTCAGGCA				
	300	310	320	330	340
DEF4	360	370	380	390	400
DEF4	CCTACTAATGCCCTAGAACCTAAAACCATCATCTGGGGCCCAGTTCCCCAAATAGCCCTA				
DEFX	CCTACTAATGCCCTAGAACCTAAAACCATCATCTGGGGCCCAGTTCCCTGAATGGCCCTA				
	360	370	380	390	400
DEF4	420	430	440	450	460
DEF4	ATTTCTTCCTCTGCTGGAATGAGTCCAGTGCCCACTTCCTCCAAAGGTGAAATTGCTGGG				
DEFX	ATCTCTTCCTCTGCTGGAATGAGTCCAGTGCCCACTTCCTCCAACGGTGAAATTGCTGGG				
	420	430	440	450	460
DEF4	480	490	500	510	520
DEF4	CCTGCAACAGATCAGGAACTCACTGCTTC-TCATAGGGGCAGCCGACTTCACTGCTCTGG				
DEFX	C-TGCTACAGATCAGGAACTCACTGCTTCCTCATAGGGGCAGCCGACTTCACTGCTCTGC				
	480	490	500	510	520

Figure 2

7/16

	540	550	560	570	580	590
DEF4	AACAGCGACCACCCCTAGCGAGGCTTGAGATGCCTCTTCCCTCCTTAAGACTGAGAGCGC					
DEFX	AACAGCGACCACCCCTAGCGAGGCTTGAGATGCCTCTTGCCTCCTTAAGACTGAGGGAGA					
	540	550	560	570	580	590
			600	610	620	630
DEF4	CGCT-----GCCCCAGTCCCTCCATAGCCCAGTGCCTGGCTGCCTTCA					
DEFX	CGCTTCAGCTCTCACTCCACTGCCCCAAGTCCCTCCACAGCGCGGTGCCTG-CTGCCTTCA					
	600	610	620	630	640	650
	640	650	660	670	680	690
DEF4	GCCAGAGCTGCAGGGG-AGGCCCTGAGCACCCAAGTCCTGCTGGACCAGCGCTGTGCACG					
DEFX	CACAGAGCTGCAGGGGNAGGTCCTGTGTATCC--GGCCTGCTGGACCAGCGCTGTGCACA					
	660	670	680	690	700	710
	700	710	720	730	740	750
DEF4	GCCCTCCCATGGCGGCAGGGGCTGCCTGGACTGCATACTGGGTTCAGCAACCTCACTATA					
DEFX	ACCCTCCCATGGCAACAGTGGCTGCCCGGCCTGCACACTGGGCTTGGCAACCTCGCTGTA					
	720	730	740	750	760	770
	760	770	780	790	800	810
DEF4	GGTATTCATTCCCTCAGGAACAACTGCATTCTTTTCTCATTTCCAGAACTCATCCCGT					
DEFX	GGTATTTATTCCCTCAGGAGTGACTGCATTCTTTTCCCATTTCCAGAACTGATGCCAT					
	780	790	800	810	820	830
	820	830	840	850	860	
DEF4	TTACCTCACTACAAGGAGGAGGATG-----GTGGAGAGTGGTACATTTTAAATGT					
DEFX	TTACCTCACTATGAGGAGGAGGAGGAGGAGGGTGGAGAGTGGTACATTTTAAATGT					
	840	850	860	870	880	890
	870	880	890	900	910	920
DEF4	GCACTAGTCTCCCTGGGACTCCCCTTCAAATAACCCAGGAGGGACCACACAAGGGAAAGC					
DEFX	GCACTATTCTCCCTAGGACTCCCCTCAAATAACCCAGGAGGGACCATAACCAGCTCATTC					
	900	910	920	930	940	950
	930	940	950	960	970	980
DEF4	TTATGCATCCCCCCCACCC-AGTGACCATCTTCCCTAACTCTGGGTGTAGGGAGACTCGTA					
DEFX	CTGTGTATCCCAAGCATANGAGTAATCATCCCACTCATGCTGAGTGTATGGTGGCCATTA					
	960	970	980	990	1000	1010
	990	1000	1010	1020	1030	1040
DEF4	AGCCTACG--GGATTGGTTTGGGAACAGGGTATTTGAGCTCACAACACAAGGTGATGCAA					
DEFX	AGCCTGCCCTGAACTGGCTTTAGAACAAGGTGTTTGAGCACACAGCACCG-----					
	1020	1030	1040	1050	1060	

Figure 2 (suite)

8/16

	1050	1060	1070	1080	1090	1100
DEF4	GCTAACACCAATCTCGCTGCAGCTTTGGCCACCATCCTAAGG-GACTTCTGACAGACATT					
	:	:	:	:	:	:
DEFX	-----TCTTGCTGCCACCTTGGCCCCCTCCCTTGTGAGACCTCTGAGACACATT					
	1070	1080	1090	1100	1110	
	1110	1120	1130	1140	1150	1160
DEF4	-AGGTGTCACGCAATCATTTGATGAGTCCTTGGCCTGGAT--GACCTAGACAGTCATTTA					
	:	:	:	:	:	:
DEFX	NAGGTCTCACCTAAAAATCTCAGGATTTCTAGGCCCAAANC GGTCCTAAAAAATTGTTCA					
	1120	1130	1140	1150	1160	1170
	1170	1180	1190	1200	1210	1220
DEF4	GGCTTGA ACTATCTAAGGCCAAGCAAAAAGGTGACTGTCCCCTCTAGGAA-CCACATGCT					
	:	:	:	:	:	:
DEFX	GTCT-GA ACTCTCTAAGGTCAAGAGAAGAGGTGGTTGCTCCCTCTAAGAAACCACATGTT					
	1180	1190	1200	1210	1220	
	1230	1240	1250	1260	1270	
DEF4	ATATGCACATCCTTTACTCGGGAGCCTGCAAC----CTGCCCTATCCAGCAACACAAGCC					
	:	:	:	:	:	:
DEFX	GCATGTACATCCTTAATTCCGGAAAGTCCAACAAACCTGCCCTGCTTAGCAACACAAGCC					
	1230	1240	1250	1260	1270	1280
	1280	1290	1300	1310	1320	1330
DEF4	CAGGCG-TATTCAGTCTCATCCAGGTATTCTCCAAC---CTTACTTGTCTGAATGGCTTG					
	:	:	:	:	:	:
DEFX	GAGGTGGTACTCC-TCTCACCCGGGCATTCTCCAACACACCTGTTTGTCCAACAGCTTT					
	1290	1300	1310	1320	1330	1340
	1340	1350	1360	1370	1380	1390
DEF4	GATTTGTTTTTATGGTTAGACCCAGGG-CCTGGGAGGTGAGTTCAGACCAATTCCAAA					
	:	:	:	:	:	:
DEFX	GATTTGTTTTTATAGTTGGACCCAGGTTCCAGGAGGCTGGTTCAGGCCATATTCCAAA					
	1350	1360	1370	1380	1390	1400
	1400	1410	1420	1430	1440	1450
DEF4	TCCTCATCTGTGTGTGGGTGGCATTTTGATCCTAGTCTCCTCGCAAGGTGTATACAACAA					
	:	:	:	:	:	:
DEFX	TCCTCATCTGTGTGTGAGTGGCATTCTTAGCCTAGCCTCCTTACAGGGTGGATACTATGA					
	1410	1420	1430	1440	1450	1460
	1460	1470	1480	1490	1500	1510
DEF4	TATGCAGGCCAGGCTCTCCTGGTGGCTTTAAATATTCCCTCGGTCCAGGTAGTTCAGCCT					
	:	:	:	:	:	:
DEFX	TACACAG-CCAGGCTGTCCAGTGGCTTTCAATATTCTTTTGGTCCAGATAGTTCAGCCT					
	1470	1480	1490	1500	1510	1520
	1520	1530	1540	1550	1560	1570
DEF4	CAGCCACCAGCATAGGTATCATGGGGTCAATTGTCTTAGGAGTCATGAGGAATCCACAGT					
	:	:	:	:	:	:
DEFX	CAGC-ACCAGTGTAGGCATCACAGGGTCAATTGTCTTAGGAGTCATGGAGAATTCATAGT					
	1530	1540	1550	1560	1570	1580

Figure 2 (suite)

9/16

```

      1580      1590      1600      1610      1620      1630
DEF4  TGATTGCTGCCTGGGCCTGGCCAGGGCTGACCAAGTAGACGAGGGGTCCGTACCTCCGT
      :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  TGGTAGCTACCTGGGCCTGGCCAGGGCTGACCA---TAGACAAGGCATC---CCTCTGT
      1590      1600      1610      1620      1630

      1640      1650      1660      1670      1680      1690
DEF4  GGACTCCTGCTTGAAGTCCAGCTTTCTGCCAAATTTCTCAACTGCCCTTGTTAACAGTTA
      : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  GAAGTCCCTATTTTAATGCCAGCTTCCCAACAAATTTCTCAACTGCTCTTACCAGCAGGTA
      1640      1650      1660      1670      1680      1690

      CAAT box
      1700 ----1710      1720      1730      1740      1750
DEF4  TTAAAGTACCCAATAGAAAGTAACGCTGAAAAATTAGGACACCTGATACCAAAGACCC
      :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  TTAAACTACTCAATAGAAAGTAACCCTGAAAA-TTAGGACACCTGTTCCCAAAGACCC
      1700      1710      1720      1730      1740      1750

      TATA box
      -----      1770      1780      1790      1800
DEF4  TTAAATAAGG-AAGTCCTCTC-CTCTGTGTGCATGGCTGCTCTTG---CTACATAAGACC
      :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  TTAAATAGGGGAAGTCCTTTCNCTGCTTGTGCACAGCTGCTGATGTGGCAACATGAGGCC
      1760      1770      1780      1790      1800      1810

      mRNA start -->
      1810      1820 | 1830      1840      1850      1860 SpSite
DEF4  TGGAACACAGGACTGCTGTCTGCCCTCTCTGCTCGCCCTGCCTAGCTTGAGGATCTGTAA
      :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  TGGGACAGGGGACTGTCCTCTGCCCACTCTGGTAGCCTCACGTAGCTTAACAATCTGTCA
      1820      1830      1840      1850      1860      1870

      --      1880      1890      1900      1910      1920
DEF4  GTAACACAA-----AACTTAAACTTTCACATTGAGGTTTCAATATTGAAGCTGTGTCCCC
      :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  GTAATACAATACAAAACCTTAAACTTTCATACTGCGGTTCCACCCAGGAAGCTGTGTTCCT
      1880      1890      1900      1910      1920      1930

      1930      1940      1950      1960      1970      1980
DEF4  AGTCTGACCTCTCACTGTGGGGCCACCCAGAGGACCCAGCGTGAAGCCCCTGCTGTGAA
      : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  AATCTGACCCGTGATTATGGGGCCACCTCAGAGGGNACCCAGTGAGGGAA-TATTTTG--
      1940      1950      1960      1970      1980      1990

      1990      2000      2010      2020      2030      2040
DEF4  CTTCTATCTGGGTGTCTGGCGGCTGCTGGGGGTAATGGCTACTAGCTAAGTCAATAGAGA
      : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :
DEFX  . ---CCATCTGGGA--CTGTTGGTTGCTGGGGGCAGTGGCTATGAGCTCAGTTAATA----
      2000      2010      2020      2030      2040

```

Figure 2(suite)

	2050	2060	2070	2080	2090	2100
DEF4	AACTCAAAAAGTTTCCTTCCAAACACACGTGTCCTACTTGACATGTCCAATAAAGACGAT					
	::: ::: :					
DEFX	AACTCAAGCAGTTTTCTTCCAAACACACATGTCCTACTTAACGTGTCCAACAGAGATGAT					
	2050	2060	2070	2080	2090	2100
	2110	2120	2130	2140		
DEF4	CA----CAGCTTCT--TAAACATTATTTTATTGTGAGAGAAGCCTCT-----					
	:: : . :					
DEFX	CATACTCATANGCTGCTAAACATTANTTTTATTTTGAGAAAAGTCTATTCATGTTCTTG					
	2110	2120	2130	2140	2150	2160
					2150	
DEF4	-----GCAG-----GTC---CTA---					
	: :					
DEFX	GCCCATGGAGTTTTCATTTNATTANTTTATTTATTTTGCAGAGATGGAGTCTCACTATGT					
	2170	2180	2190	2200	2210	2220
				2160		
DEF4	-----GGTCT-----GTTTTTC-----					
	: :					
DEFX	TGCTCAAGCTGGTCTCCAACCTCGGGCTCAAGCGATCTTCTACTTTGGCCTTTGAAAG					
	2230	2240	2250	2260	2270	2280
					2170	
DEF4	-----AATCAGGTT					
	: :					
DEFX	CGCTGAGATTGCCTGTGTGAGCCATCATGGGGGCTCACTGGCCCCACTGATTAATCAGATT					
	2290	2300	2310	2320	2330	2340
	2180	2190	2200	2210	2220	2230
DEF4	GTTTGTTTTTTGCTATTGA-GTTGTTTGACTTCCTTATGTATTCAGATATTTACCCCTTC					
	: :					
DEFX	AATTGTTTTTTGCTATTGAANTTGTTTGACTTCCTTGTATATTCGGATATTTACCCATTC					
	2350	2360	2370	2380	2390	2400
	2240	2250	2260	2270	2280	2290
DEF4	TACCACGTAGGCTTTGCAAACATTTTCTCTCATTTTCTGGGTTGCCGTTTCCCTCAGTTG					
	: :					
DEFX	TAACACGTAGGGTTTGCAAATATTTTCTCTCATGTTCTGTGTTGCCTTTTCACTCAGTTG					
	2410	2420	2430	2440	2450	2460
	2300	2310	2320	2330	2340	2350
DEF4	ATTGTTTCCTTTGCTATGAAGATGCTTTAGCGTTCAATGCAGCCCCGCTTGTCTATTTTC					
	: :					
DEFX	ATGGTTTCCTTTGCTGTGCAGGTGCTTTAGTGTTCAACGCAGCCCCGCTTGTCTATTTTC					
	2470	2480	2490	2500	2510	2520
	2360	2370	2380	2390	2400	2410
DEF4	CCATTTGTTTATTGCCTGTGCCTTTGGTGT CATAGCCAAGAAATCATTACTCACGTCAAT					
	: :					
DEFX	C-ATTT---TATTGCCTGTCCCTTTGATGTCATAGCCAAGAAATAATTGCCAGATTAAT					
	2530	2540	2550	2560	2570	2580

Figure 2 (suite)

11/16

```

      2420      2430      2440      2450      2460      2470
DEF4  GTCCAAA-GCTTTATCTTTGTATGTGCTTCTCGTAGTTGTATGGTTTCAGGTCTTTTCAA
      ::: ::: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  GTCAAAAGCTTTATCCCTATATATTCTTCTAGTAGTT-TATGGTTTCAGATCTT-----
      2590      2600      2610      2620      2630

      2480      2490      2500      2510      2520      2530
DEF4  GTCTATGTTGAG-TCTTCAATCCATGTTGAGCTGATTTT-TACATGTTGTGAGAGAAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  ----ATGTTTAGGTCTTCAATCCA--TTGAGTTGATTTTGTATGTGGTATAAGAAAAAA
      2640      2650      2660      2670      2680      2690

      2540
DEF4  GACCACGTGTATGCACCT-----
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  GACCACATGTATACATATCTCAAATTCTAAGGTAGTATATATTAGACACATACAATGTGT
      2700      2710      2720      2730      2740      2750

      2550      2560      2570
DEF4  -----AGC---AACTCATGAAC-----CTTACA--CAACTCTTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  CTATTTACACACATTGAGCTGAAAATAATAAACATATTTTATCTTTCAATCAACTCTAT
      2760      2770      2780      2790      2800      2810

      2580      2590      2600      2610      2620      2630
DEF4  ATCTCTCTCACTGAGCTCATTTACCTGTACCCTGATAAGGTCATTGTCCTCTTCACTCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  CTCTATCTCACTGAACCTTGTTCACCTATAGCCTGATGAGGTTGCTGTCCTCTCTACCCC
      2820      2830      2840      2850      2860      2870

      2640      2650      2660      2670      2680      2690
DEF4  GGCCCTACAGGAGACTACTCACCCCATACCTCAGTCGCCCTTCATGAGGGT-ATAAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  AGCTCCTATAGGAGACTGCTCATCCCTAACCTCAAAAACCCCTTCATGAGGGTGATAAT
      2880      2890      2900      2910      2920      2930

      2700      2710      2720      2730      2740      2750
DEF4  GACCTAGAAGCCTGCAATGAGTTACT-CTCTACTCCACCGGAATTCAGGTCTGGCACCAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  GCCCTTGAATCCTGCAATGAATTAGTTCTCTACTACAGTGGAATTCAGGTCTGTTATGAG
      2940      2950      2960      2970      2980      2990

      2760      2770      2780      2790      2800      2810
DEF4  TGTTTAGACCT--GAAGAGAATAGTAGGGCCCATATCAGGAAATAAGAGGCATTTGCTC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  GGTCTGGATCTCTGAAGAGAAGAG---CTCTCATTTTCAGAAAATAAGCAGGATTTATTC
      3000      3010      3020      3030      3040

      2820      2830      2840      2850      2860      2870
DEF4  TCTTAAATTATTGAATGAAAGCACTGTTTCCATT-CTTTTGTAGAATATTAAAGATTTAAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  CCTGAAATTACTGAATTAAATCACTGTTTCGATTACTTTTGTCAATATTAAA-----
      3050      3060      3070      3080      3090

```

Figure 2(suite)

12/16

```

      2880      2890      2900      2910      2920      2930
DEF4  CAGGAAATATTAGGTATTTCTGAAAACAGGAAAAATGCCAGGGTCCTCATCATCACCA
      :: ::::: : : ::::: ::::: ::::: :::::
DEFX  -AGTAAATATTTA--AACAGGTAAACAG-AAATAATGGTAGGGTCCTTATCATCACCG
      3100      3110      3120      3130      3140      3150

      2940      2950      2960      2970      2980
DEF4  TCAACTTCACCTAGGCACAGACACTAACATAGAGCTTC---CTGTGAAGAAAGCTGGG
      : : : : : ::::: : ::::: ::::: : : : : :::::
DEFX  TGAATTCCAAGCTAG-CATAGACACTAACCTAGAGATTACACTAGAATGAAAGCTGGG
      3160      3170      3180      3190      3200      3210

      2990      3000      3010      3020      3030      3040
DEF4  AGAGCAGAGGAGGCATTCCAGGGATGTCAAGGCCAATAGGAGTCGGCATCCTCTCTAACA
      ::::: : : : : ::::: : ::::: : : : : :::::
DEFX  AGAGCAGAGGAGTC-TCAGAAGGATGTGGAGGCCAATGGACACCTGCAACCTCTCCAACG
      3220      3230      3240      3250      3260      3270

      3050      3060      3070      3080      3090      3100
DEF4  AAATGCACACCTCCTCTCACTCAGAAGGCCAAAGGTTTCTTATCTCTGTGCCTTCTCCCA
      ::::: ::::: ::::: : : : : : ::::: ::::: : :
DEFX  AAATGCCTACCTCCTCTCACT-----GC---AGCATCC--ATCTCTGAGCCTTCTCGCA
      3280      3290      3300      3310      3320

      3110      3120      3130      3140      3150      3160
DEF4  GAA-AGCTATAAATCCAAGCTGGCTTCTCCCTCCCCACACAGCTGCTCCTGCTCTCCCTC
      : : ::::: : : ::::: : : : : : : : :::::
DEFX  GCAGAGCTATAAATTCAGCCTGGCTCCTCCGTTCCCACACATCCACTCCTGCTCTCCCTC
      3330      3340      3350      3360      3370      3380

      <----- exon2 -----
      3170      3180      3190      3200      3210      3220
DEF4  CTC-----CAGGTCACCCAGCCATGAGGATTATCGCCCTCCTCGCTGCTATTCTCTTGG
      : : : : : : : : : : : : : : : : : : : :
DEFX  CTCTCCTCCAGGTGACTACAGTTATGAGGACCCTCACCTCCTCTCTGCCTTTCTCCTGG
      3390      3400      3410      3420      3430      3440

      -----
      3230      3240      3250      3260      3270      3280
DEF4  TAGCCCTCCAGGTCCGGGCAGGCCACTCCAGGCAAGAGGTGATGAGGCTCCAGGCCAGG
      : : : : : : : : : : : : : : : : : : : :
DEFX  TGGCCCTTCAGGCCTGGGCAGAGCCGCTCCAGGCAAGAGCTCATGAGATGCCAGCCAGA
      3450      3460      3470      3480      3490      3500

      -----
      3290      3300      3310      3320      3330      3340
DEF4  AGCAGCGTGGGCCAGAAGACCAGGACATATCTATTTCTTTGCATGGGATAAAAGCTCTG
      : : : : : : : : : : : : : : : : : : : :
DEFX  AGCAGCCTCCAGCAGATGACCAGGATGTGGTCATTTACTTTTCAGGAGATGACAGCTGCT
      3510      3520      3530      3540      3550      3560

```

Figure 2 (suite)

13/16

```

----->
      3350      3360      3370      3380      3390      3400
DEF4  CTCTTCAGGTTTCAGGTGAGAGAGGCCAGCATAAAAAGCTACCGAGTCTAGAGAGACGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  CTCTTCAGGTTCCAGGTGAGAGATGCCAGCATGCAGA-GCTAC--AGACTAGACAGAAGG
      3570      3580      3590      3600      3610

      3410      3420      3430      3440      3450      3460
DEF4  ATGGGAGATGGGCTCTGGAATCACATCTCAATGGTGGATGTCACTTAGGTGGCTTTACTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  ACAGGAGACAGGCTCTGGAATTGGATCTCAGTGGCAGATGTCACTTAGGTGGCTATACTT
      3620      3630      3640      3650      3660      3670

      3470      3480      3490      3500      3510      3520
DEF4  ACCATCTCTGGGCCTCGATTTTCTTATCTCGAACTGAATAGAGAGACAAACAAATGTAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  AACATCTCTGGTCCTGGATTTTCTCATATCTAAATGGAATAGAGAACCAAAGAAATCTAA
      3680      3690      3700      3710      3720      3730

      3530      3540      3550      3560      3570      3580
DEF4  GT-AGTCTTCTTTCTCCAAAGACTTGATTCCAAGGTATGTCTATAAAATTCGCTAGGGTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  GAGATTTTCTTTCTCCAAAAGTTGATTCCAAGATATGACTGTGAAATTCAGTAGATTT
      3740      3750      3760      3770      3780      3790

      3590      3600      3610      3620      3630
DEF4  AAGATATGGAGAGACAGATTGACCAGTTCTTTCTGGATCTAAACAAGTA-GAT--ATTAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  AAGATATAAGGAGATG--CTACCTAGTTCCTTCTGGAGCCAGACAAACAAGCTTAAGTAT
      3800      3810      3820      3830      3840      3850

      3640      3650      3660      3670      3680      3690
DEF4  AG-GGAAAATATTTTCATTCTGCCAACAAAGGAAATTTTAAAAACTGGAGATGGGCTTAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  ATAGGAAAATATTTTACCCTGTCTATATAGGAGGTTTTAGAACCTGGAGAGGAGCCTAAG
      3860      3870      3880      3890      3900      3910

      3700      3710      3720      3730      3740      3750
DEF4  AGTATGTTTCAGGTGTGTGTCTGATGGGGCA--AAAGCACACAAATCAGAGCAAAGAGAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  AATGTGTTTCAGGTGTGTGTGTGATGGG-CAGGAATGCAGAAAAGTGA-AGCAAAGGAGAA
      3920      3930      3940      3950      3960      3970

      3760      3770      3780      3790      3800      3810
DEF4  TGAGTCTCAAATCCTGTATGAGCAGCATTGCTCTGTGTATTTATTCTATTGACTAAGGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  TGAGTCTCGAATCCTGTGTGACCAGCACTGCTCTGTGTATTTATTCTATTGACTGAGAT
      3980      3990      4000      4010      4020      4030

      3820      3830      3840      3850      3860      3870
DEF4  TGTTTGTGCTACCGGCACTAATGCAGCCAGCATCACCGGTCAGCCAGCATGTGCATTCTC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DEFX  TGTTTGTGCTACCGGCTGTAATACAGCCAACATCACTCATCAGCCAACATGTGACTTCTC
      4040      4050      4060      4070      4080      4090

```

Figure 2(suite)

14/16

```

3880      3890      3900      3910      3920      3930
DEF4      CAAGATTCCCTTTACCACCCACCGCTGACCTTGGTGCTTAAATTTCTCAGTCTTCCTCTGT
          :::::::::::::::::::::: :::: :  :: :: : ::::: :  :: :::::
DEFX      CAAGATTCCCTTTACCACCCACTGCTGNACCCCGTACTCAGTTTCTGATGCTCTCTCTGG
          4100      4110      4120      4130      4140      4150

          <----- exon3 ----->
3940      3950      3960      3970      3980      3990
DEF4      GTTCCCAGGCTCAACAAGGGGCATGGTCTGCTCTTGCAGATTAGTATTCTGCCGGCGAAC
          :: :::::::::::::: :::: :  ::::: ::::: :  ::::: :::::
DEFX      GTCCCCAGGCTCAACAAAGGGCTTGATCTGCCATTGCAGAGTACTATACTGCATTTTGG
          4160      4170      4180      4190      4200      4210

----- exon3 -----
4000      4010      4020      4030      4040      4050
DEF4      AGAACTTCGTGTTGGGAAGTGCCTCATTGGTGGTGTGAGTTTCACATACTGCTGCACGCG
          ::::: :  :: ::::: ::::: ::::: :  ::::: :  ::::: :  ::::: :
DEFX      AGAACATCTTGGTGGGACCTGCTTCATCCTTGGTGAACGCTACCCAATCTGCTGCT----
          4220      4230      4240      4250      4260      4270

----- exon3 -----
4060      4070      4080      4090      4100      4110
DEF4      TGTCGATTAACATTCTGCTGTCCAAGAGAATGTCATGCTGGGAACGCCATCATCGGTGGT
          : :::
DEFX      -----ACTAA-----

----- exon3 -----
4120      4130      4140      4150      4160      4170
DEF4      GTTAGCTTCACATGCTTCTGCAGCTGAGCTTGCAGAAATAGAGAAAAATGAGCTCATAATT
          ::::::::::: ::::::::::: :  :: :::::::::::
DEFX      -----GCTTGCAGACTAGAGAAAAA-GAGTTCATAATT
          4280      4290      4300

----- exon3 -----
4180      4190      4200      4210      4220      4230
DEF4      TGCTTTGAGAGCTACAGGAAATGGTTGTTTCTCCTATACTTTGTCCTTAACATCTT-TCT
          : ::::::::::: :  :::: :  ::::: :  ::::: :  ::::: :  ::::: :
DEFX      TTCTTTGAGCATTAAAGGGAATTGTTATT---CTTATACCTTGTCTCGATTTCTGTCC
          4310      4320      4330      4340      4350      4360

          Poly Ad
          ----->
4240      4250      4260      4270      4280      4290
DEF4      TGATCCTAAATATATATCTCGTAACAAGATGTCTTTGTTTACACCTCTTTGAAATTTGAT
          : :::: :  ::::: :  ::::: :  ::::: :  ::::: :
DEFX      TCATCCCAAATAAATACTTGGTAACATGATTTCCGGGTTTTTTTTTTTTT
          4370      4380      4390      4400      4410

```

Figure 2(suite)

15/16

	10	20	30	40	50	60
DEF4	GTCTGCCCTCTCTGCTCGCCCTGCCTAGCTTGAGGATCTGTCACCCCAGCCATGAGGATT					
	:	:	:	:	:	:
DEFX	CTCTGCCCACTCTGGTAGCCTCACGTAGCTTAACAATCTGTGACTACAGTTATGAGGACC					
	10	20	30	40	50	60
	70	80	90	100	110	120
DEF4	ATCGCCCTCCTCGCTGCTATTCTCTTGGTAGCCCTCCAGGTCCGGGCAGGCCCCACTCCA:					
	:	:	:	:	:	:
DEFX	CTCACCTCCTCTCTGCCTTTCTCCTGGTGGCCCTTCAGGCCTGGGCAGAGCCGCTCCAG					
	70	80	90	100	110	120
	130	140	150	160	170	180
DEF4	GCAAGAGGTGATGAGGCTCCAGGCCAGGAGCAGCGTGGGCCAGAAGACCAGGACATATCT					
	:	:	:	:	:	:
DEFX	GCAAGAGCTCATGAGATGCCAGCCCAGAAGCAGCCTCCAGCAGATGACCAGGATGTGGTC					
	130	140	150	160	170	180
	190	200	210	220	230	240
DEF4	ATTTCTTTTGCATGGGATAAAAGCTCTGCTCTTCAGGTTTCAGGCTCAACAAGGGGCATG					
	:	:	:	:	:	:
DEFX	ATTTACTTTTCAGGAGATGACAGCTGCTCTCTTCAGGTTCCAGGCTCAACAAAGGGCTTG					
	190	200	210	220	230	240
	250	260	270	280	290	300
DEF4	GTCTGCTCTTGCAGATTAGTATTCTGCCGGCGAACAGAACTTCGTGTTGGGAAC TGCTC					
	:	:	:	:	:	:
DEFX	ATCTGCCATTGCAGAGTACTATACTGCATTTTTGGAGAACATCTTGGTGGGACCTGCTTC					
	250	260	270	280	290	300
	310	320	330	340	350	360
DEF4	ATTGGTGGTGTGAGTTTCACATACTGCTGCACGCGTGTGATTAAAGTTCTGCTGTCAA					
	:	:	:	:	:	:
DEFX	ATCCTTGGTGAACGCTACCCAATCTGCTG-----CTACTAA-----					
	310	320	330	340	350	
	370	380	390	400	410	420
DEF4	GAGAATGTCATGCTGGGAACGCCATCATCGGTGGTGTAGCTTCACATGCTTCTGCAGCT					
DEFX	-----					
	360	370	380		390	
	430	440	450	460	470	480
DEF4	GAGCTTGCAGAATAGAGAAAAATGAGCTCATAATTTGCTTTGAGAGCTACAGGAAATGGT					
	:	:	:	:	:	:
DEFX	--GCTTGCAGACTAGAGAAAAA-GAGTTCATAATTTTCTTTGAGCATTAAAGGGAAT---					
	400	410	420	430	440	450
	490	500	510	520	530	
DEF4	TGTTTCTCCTATACTTTGTCCTTAACATCTT-TCTTGATCCTAAATATATATCTCGTAAC					
	:	:	:	:	:	:
DEFX	TGTTATTCTTATACCTTGTCTCGATTTCCTGTCCTCATCCCAAATAAATACTTGGTAAC					
	460	470	480	490	500	510
	540					
DEF4	AAG					
	:					
DEFX	ATG					

Figure 3

<----- Signal peptide -----><--
5 10 15 20
MetArgThrLeuThr LeuLeuSerAlaPhe LeuLeuValAlaLeu GlnAlaTrpAlaGlu

----- Propiece -----
25 30 35 40
ProLeuGlnAlaArg AlaHisGluMetPro AlaGlnLysGlnPro ProAlaAspAspGln

----- Propiece -----
45 50 55 60
AspValValIleTyr PheSerGlyAspAsp SerCysSerLeuGln ValProGlySerThr

-----><----- Mature peptide -----
65 70 75 80
LysGlyLeuIleCys HisCysArgValLeu TyrCysIlePheGly GluHisLeuGlyGly

----- Mature peptide ----->
85 90 94
ThrCysPheIleLeu GlyGluArgTyrPro IleCysCysTyr

Figure 4

	SIGNAL	PROPIECE
DEF4_HUMAN	MRIALLAAILLVALQVRA	GPLQAR-----GDEAPGQ-EQRPEDQDISISFAWDKSS
DEF5_HUMAN	MRTIAILAAILLVALQAQA	ESLQER-----ADEATTQ-KQSGEDNQDLAISFAGNGLS
DEF6_HUMAN	MRTLILTAVLLVALQAKA	EPLQAEDDPLQAKAYEADAQ-EQRGANDQDFAVSFAEDASS
DEF1_HUMAN	MRTLAILAAILLVALQAQA	EPLQAR-----ADEVAAAPEQIAADIPEVVVSLAWDESL
DEFX	MRTLTLLSAFLVALQAWA	EPLQAR-----AHEMPAQ-KQPPADDQDVVIYFSGDDSC
	** ...*.* ***** *	** * *

	PROPIECE	Mature PEPTIDE
DEF4_HUMAN	ALQVSGSTRGM	VCSCRLVFCRTELRVGNCLIGGVSFYTCCTRVD
DEF5_HUMAN	ALRTSGSQARA	TCYCRTGRCATRESLSGVCEISGRLYRLCCR---
DEF6_HUMAN	SLRALGSTRAF	TCHCRR-SCYSTEYSYGTCTVMGINHRFCCL---
DEF1_HUMAN	APKHPGSRKNM	ACYCRIPACIAGERRYGTCTYQGRWAFCC----
DEFX	SLQVPGSTKGL	ICHCRVLYCIFGEHLGGTCFILGERYPICCY---
	. . **	* * * * *
		^ ^ ^ ^ ^

Figure 5