

*French Specification of the EPAs filed**39 pages
2 abstract
17 drawings*

1

**ANIMAL TRANSGENIQUE OU CHIMERIQUE ET PROCEDE DE
CRIBLAGE DE PRODUITS PHARMACEUTIQUES**

L'invention concerne un procédé de remplacement spécifique d'une copie d'un gène présent dans le génome d'un organisme eucaryote receveur par l'intégration d'un gène différent du gène inactivé. De préférence, le gène receveur sera présent en au moins 2 exemplaires dans la cellule hôte transfectée. Le gène receveur est défini comme étant le gène ou se fera l'insertion du gène différent.

Plus particulièrement, l'invention concerne la production d'animaux transgéniques dans lesquels le gène étranger a été introduit d'une manière ciblée pour permettre, à la fois, le maintien des fonctions génétiques normales de l'animal et l'expression du gène étranger sous le contrôle de promoteurs endogènes.

Par "gène différent ou étranger" on entend toute séquence nucléotidique correspondant à la totalité ou à une partie d'un gène "étranger ou différent" du gène récepteur telle qu'elle est trouvée normalement dans le génome (ARN ou ADN), ou elle correspond également à une séquence modifiée artificiellement du gène normal ou encore à un fragment de cette séquence.

L'invention concerne également le procédé de production de ces animaux transgéniques.

Dans la production d'animaux transgéniques, les méthodes conventionnelles utilisées pour l'introduction de séquences d'ADN hétérologues dans la lignée cellulaire germinale, ne permettent pas de contrôler le site de l'intégration du gène étranger dans le génome, ni le nombre de copies ainsi introduit. L'intégration du gène étranger se fait au

2

hasard et, en général, plusieurs copies du gène s'intègrent en même temps, parfois sous forme de tandem tête à queue, le site de l'intégration et le nombre de copies intégrées variant d'un animal transgénique à l'autre.

Il peut donc arriver que des gènes cellulaires endogènes, situés au point d'insertion, soient ainsi inactivés, sans que cela soit facilement décélable en raison de nombreuses insertions au hasard. Si le produit de ces gènes est important pour le développement de l'animal, celui-ci sera sérieusement perturbé. D'ailleurs, l'insertion aléatoire du gène étranger peut se faire à un site qui n'est pas approprié pour l'expression du gène. De plus, le fait qu'il y ait variation du site et du nombre d'insertions d'animal en animal rend l'interprétation des études d'expression extrêmement difficile.

Un problème majeur rencontré dans la production d'animaux transgéniques, est l'obtention de l'expression du gène étranger. D'une manière générale, deux types d'expérience ont été réalisés chez les souris.

Les gènes introduits dans la lignée germinale sont :

- soit des gènes "complets", comprenant des séquences codantes flanquées de leurs propres séquences régulatrices;
- soit des gènes composites, formés de la séquence codante d'un gène fusionnée à la séquence promotrice d'un autre gène, les deux fragments appartenant même parfois à deux espèces animales différentes.

On a pu ainsi confirmer que la spécificité de l'expression des gènes dans tel ou tel tissu est déterminée par leur(s) séquence(s) régulatrices.

Le choix du promoteur approprié pour l'expression du gène étranger chez l'animal transgénique est donc d'une importance primordiale.

D'autre part, la mutagenèse dirigée de gènes murins dans des cellules souches embryonnaires a récemment été réalisée en faisant appel à une technique de "ciblage génétique" (gene targeting) (Thomas et al., 1987; Thompson et al., 1989).

Dans le premier cas, le gène murin HPRT a été muté par insertion et remplacement et, dans le deuxième cas, un gène HPRT muté a été corrigé. Thomson et al. ont étendu leurs expériences jusqu'à l'obtention de souris chimères et ont constaté le passage de la modification génétique dans la lignée cellulaire germinale.

Dans chacun des documents cités, le site précis de l'intégration a été ciblé par recombinaison homologue entre, d'une part, des séquences exogènes comportant la mutation ou correction incluses dans un vecteur, sous le contrôle d'un promoteur exogène, et, d'autre part, leur homologue endogène. Ceci étant, il faut remarquer que les auteurs antérieurs ont réalisé leurs expériences sur un gène spécifique (HPRT) dont l'activation par mutation s'accompagnait d'un phénotype décelable. La mutation ciblée décrite par Thomas et al., avait pour effet d'inactiver le gène HPRT et, par conséquent, de faire disparaître le phénotype décelable normalement associé avec le HPRT. Le gène de sélection Neo^R, sous le contrôle d'un promoteur TK, était donc incorporé dans l'ADN d'insertion afin de permettre la sélection des transformants. Il est à noter que les expériences décrites dans l'art antérieur impliquaient une sélection soit par le gène receveur (p. ex HPRT) soit par le gène d'insertion (p. ex Neo^R). Le site de l'insertion et/ou le type de gène inséré est donc

limité à des gènes conférant un caractère sélectable, une sélection directe.

En outre, dans l'art antérieur, les séquences exogènes sur le vecteur servent donc à la fois à cibler le site d'intégration et à introduire la modification. Suite à la recombinaison homologue, le gène modifié se trouve toujours dans son environnement génétique normal.

Rappelons qu'un problème qui se pose au cours de la production d'animaux transgéniques est le danger d'inactiver un gène cellulaire endogène qui se trouve au point d'insertion du gène étranger.

Selon la fonction du produit du gène inactivé, une telle inactivation peut conduire à des désordres physiologiques ou morphologiques importants chez l'animal transgénique, ou pourrait même empêcher sa survie.

En revanche, l'inactivation d'un gène pourrait être considéré comme avantageux si le gène en question codait pour un récepteur de virus ou autre agent infectieux.

Les inventeurs ont étudié la possibilité d'éviter les inconvénients décrits plus haut, et associés, dans certains cas, à l'inactivation possible d'un ou plusieurs gènes cellulaires endogènes de fonction importante au cours de la production d'animaux transgéniques.

L'invention a pour objet un procédé de remplacement ou d'insertion spécifique d'un gène, notamment par ciblage d'un ADN, dit ADN d'insertion constitué par la totalité d'un gène, ou par une partie d'un gène, susceptible d'être rendu fonctionnel, ou dont le fonctionnement peut être rendu plus efficace, lorsqu'il est recombinaisonné avec un ADN de complément pour alors fournir un gène

5

recombinant complet dans le génome d'une cellule eucaryote, caractérisé en ce que

- le site d'insertion se trouve dans un gène choisi, dit gène receveur, et contenant l'ADN de complément, et en ce que

- l'on transfecte des cellules eucaryotes avec un vecteur contenant un insert comprenant lui-même l'ADN d'insertion et deux séquences dites "flanquantes" de part et d'autre de l'ADN d'insertion, respectivement homologues à deux séquences génomiques qui jouxtent le site d'insertion souhaité dans le gène receveur,

- l'ADN d'insertion étant hétérologue vis-à-vis du gène receveur, et

- les séquences flanquantes, étant choisies parmi celles qui constituent le susdit ADN de complément et qui autorisent, par recombinaison homologue avec des séquences correspondantes du gène receveur, la reconstitution d'un gène recombinant complet dans le génome de la cellule eucaryote.

Selon le procédé de l'invention, ledit ADN d'insertion peut contenir soit une séquence codante soit une séquence régulatrice, le procédé étant alors caractérisé en ce que

- le site d'insertion se trouve dans un gène choisi dit gène receveur et en ce que

- l'on transfecte des cellules eucaryotes avec un vecteur contenant un insérat comprenant lui-même l'ADN d'insertion et deux séquences dites "flanquantes" de part et d'autre de l'ADN d'insertion, respectivement homologues à deux séquences génomiques qui jouxtent le site d'insertion souhaité dans le gène receveur,

- l'ADN d'insertion étant hétérologue vis-à-vis du gène receveur et,

- les séquences flanquantes étant choisies afin de permettre par recombinaison homologue selon le cas,

6

soit l'expression de la séquence codante de l'ADN d'insertion entier sous le contrôle des séquences régulatrices du gène receveur, soit l'expression d'une séquence codante du gène receveur sous le contrôle de séquences régulatrices de l'ADN d'insertion.

Par exemple, l'ADN d'insertion peut contenir une séquence codante dépourvue d'élément de régulation, notamment d'un promoteur qui lui est propre.

L'invention concerne aussi un procédé de production d'animaux transgéniques, caractérisé en ce que des cellules E.S. sont transfectées dans les conditions sus-décrites et sélectionnées pour l'évènement de recombinaison homologue, à savoir l'intégration correcte du gène étranger, les cellules transfectées sont injectées dans des embryons à un stade où ils sont aptes à intégrer les cellules transfectées (par exemple au stade blastocyste), ceux-ci sont ensuite réimplantés dans une mère porteuse et les individus chimères obtenus au terme de la gestation, et chez lesquels est constatée la colonisation par les cellules E.S. de la lignée germinale, sont accouplés. Si les cellules E.S. ont colonisé la lignée germinale de l'animal chimère, des animaux transgéniques hétérozygotes pour le gène remplacé seront obtenus par accouplement (F1) dans la descendance.

Il est également possible d'insérer l'ADN d'insertion, porté par le vecteur de l'invention, dans l'oeuf peu de temps après la fécondation (c'est-à-dire moins de 24 heures). De cette manière, l'insertion est effectuée pendant que l'oeuf est à l'état unicellulaire.

L'invention concerne aussi un plasmide apte à effectuer l'insertion ciblée d'un gène recombinant dit gène d'insertion dans le génome d'une cellule

eucaryote, caractérisé en ce qu'il contient un insérat comprenant lui-même le gène d'insertion et deux séquences dites "flanquantes" de part et d'autre du gène d'insertion, respectivement homologues aux deux séquences génomiques qui jouxtent le site d'insertion souhaité dans le gène receveur.

L'ADN d'insertion peut être un gène, ou une partie d'un gène, hétérologue à l'espèce transfectée.

L'invention concerne également des animaux transgéniques dans lesquels au moins un gène endogène, qui de préférence est présent dans le génome à au moins deux exemplaires, a été inactivé par l'insertion d'un gène qui est différent du gène inactivé, le gène d'insertion étant inséré dans une position qui permet l'expression de ce gène sous le contrôle des séquences régulatrices du gène endogène inactivé.

Le procédé de l'invention permet, donc, grâce au phénomène de recombinaison homologue, d'insérer d'une manière ciblée des gènes étrangers, en particulier des séquences codantes des gènes du promoteur qui leur est normalement associée, dans le génome d'un organisme eucaryote à un site qui permet leur expression sous le contrôle du promoteur endogène du gène où se fait l'insertion, et par conséquent, d'inactiver le gène endogène ciblé.

Selon un mode de réalisation préféré de l'invention, le gène receveur ciblé est un gène qui est présent dans le génome en au moins deux exemplaires. L'utilisation de la technique d'électroporation (Ref. 11) assure l'introduction d'une copie seulement du gène étranger.

Selon cette variante de l'invention, l'insertion ciblée du gène d'intérêt (c'est-à-dire le gène dit d'insertion) a pour effet d'inactiver la seule copie du gène cellulaire endogène où se fait l'insertion et

laisse intacte et fonctionnelle la ou les autre(s) copie(s) de ce gène.

De cette façon, le fonctionnement génétique de l'animal transgénique n'est pas ou peu perturbé par l'introduction du gène étranger, même si l'insertion inactive une seule copie d'un gène essentiel receveur pour le développement de l'animal. Soit son développement ne serait donc pas affecté par l'insertion du gène étranger, soit les perturbations mineures possibles dans le cas de l'inactivation d'un gène critique ne seraient probablement pas léthales pour l'animal. Les effets de l'insertion du gène étranger à l'état homozygote pourraient être de toute nature et seraient observés en 2ème génération (F2) après croisements d'individus hétérozygotes (F1) entre eux.

Si, par contre, l'inactivation de toutes les copies d'un gène est souhaitée, par exemple, dans le cas où le gène code pour un récepteur d'agent infectieux, de multiple copies du gène étranger sont introduites. Le contrôle de la quantité introduite peut être assuré en faisant appel à des techniques connues.

L'insertion ciblée du gène étranger permet donc son introduction dans un site où son expression est sous le contrôle des séquences régulatrices du gène endogène où se fait l'insertion.

Le procédé de l'invention permet ainsi d'insérer le gène étranger derrière un promoteur endogène qui a les fonctions désirées (par exemple, spécificité d'expression dans tel ou tel tissu), et cela, le cas échéant, sans inactiver les autres copies du gène receveur.

Selon un mode de réalisation préféré, l'ADN d'insertion, ou le vecteur, contient des séquences intercalées entre l'ADN d'insertion et les séquences

flanquantes. Ces séquences peuvent contenir par exemple une séquence codant pour un agent sélectif permettant la sélection des transformants et éventuellement un gène marqueur par exemple le LacZ.

Selon un mode de réalisation particulièrement préféré de l'invention, l'ADN d'insertion comporté entre les séquences flanquantes, d'une part une séquence d'ADN destinée à être recombinée avec l'ADN de complément dans le gène receveur pour fournir un gène recombinant, et, d'autre part, une séquence codant pour un agent sélectif permettant la sélection des transformants et un promoteur autorisant l'expression de l'agent sélectif, le gène receveur et le gène recombinant codant pour des produits d'expression ne conférant pas de phénotype sélectionnable. Selon cette variante, l'insérat du plasmide de l'invention comprend, entre les séquences flanquantes, d'une part une séquence d'ADN destinée à être recombinée avec l'ADN de complément dans le gène receveur, et, d'autre part, une séquence codant pour un agent sélectif permettant la sélection des transformants et un promoteur autorisant l'expression de l'agent sélectif, la séquence d'ADN destinée à être recombinée avec l'ADN de complément étant autre qu'un gène codant pour un agent sélectif.

De cette manière, la sélection des transformants est entièrement indépendante de la nature du gène receveur et du gène inséré, contrairement aux procédés décrits jusqu'à ce jour dans lesquels le gène inséré ou le gène receveur devait par nécessité coder pour un produit d'expression permettant la sélection des transformants. Le système développé par les inventeurs permet une flexibilité totale en ce qui concerne la nature du gène receveur et du gène inséré ou du gène formé par la recombinaison homologue. Les inventeurs ont constaté d'une manière

surprenante que l'insertion de séquences de taille importante (par exemple d'environ 7.5 kb) n'affecte pas la fréquence de recombinaison homologue.

L'effet que peut avoir l'insertion de la séquence d'ADN selon cet aspect de l'invention inclut, selon le type de séquence insérée par exemple le remplacement d'une séquence codante, le remplacement d'une séquence régulatrice, l'inactivation ou la réactivation d'un gène par mutation ou l'amélioration du taux d'expression d'un gène. Il est possible, selon l'invention, de remplacer une phase codante ou une partie d'une phase codante par une séquence hétérologue qui commence au codon d'initiation du gène remplacé afin que l'expression du gène inséré remplace entièrement l'expression du gène remplacé. Ceci évite la formation de protéines de fusion qui pourrait être indésirable chez un animal transgénique.

Selon ce mode de réalisation de l'invention, l'ADN d'insertion peut comporter entre les séquences flanquantes une séquence codante hétérologue dépourvue de promoteur, la séquence codante étant autre qu'un gène codant pour un agent de sélection. L'ADN d'insertion peut comporter en outre, en aval de la séquence codante et toujours entre les séquences flanquantes, un gène codant pour un agent de sélection, associé à un promoteur permettant son expression dans la cellule cible.

De cette manière, la séquence codante hétérologue peut être insérée derrière un promoteur endogène qui a les propriétés souhaitées, par exemple une certaine spécificité d'expression, ou grille de transcription etc, la sélectabilité des cellules transformées étant entièrement indépendante de l'expression de la séquence codante hétérologue. Ce type de construction permet, par exemple, de

sélectionner les transformants même si le gène remplacé par la séquence codante hétérologue n'est pas normalement exprimé dans les cellules cibles. Ceci est particulièrement important dans la production d'animaux transgéniques à partir de cellule E.S. ("Embryonic Stem Cells") puisqu'une proportion importante des gènes reste inactive jusqu'à un stade plus avancé du développement de l'animal. Le gène HOXC8 (également connu sous le nom Hox-3.1) est un exemple de ce type de gène. D'autre part, si la séquence codante code pour une protéine facilement décelable, par exemple, la β -Gal, le développement de la grille de transcription du gène endogène remplacé peut être suivi. Le vecteur pGN est un exemple de ce type de construction.

Selon un autre mode de réalisation de l'invention, l'ADN d'insertion peut comporter une séquence régulatrice étrangère. Le site d'insertion et, par conséquence, les séquences flanquantes sont choisies en fonction du but désiré, à savoir soit l'insertion de la séquence régulatrice étrangère pour donner un effet de "double promoteur" avec la séquence régulatrice endogène, ou soit le remplacement d'un promoteur endogène par le promoteur étranger. La séquence codante qui se trouve sous le contrôle de la séquence régulatrice peut être endogène.

Une autre possibilité serait l'insertion ciblée d'un ADN étranger qui comporte à la fois une séquence régulatrice et une séquence codante. Il est possible que la séquence régulatrice soit celle qui est naturellement associée avec la séquence codante.

Le procédé de l'invention met en oeuvre un vecteur contenant deux séquences "flanquantes" de part et d'autre du gène étranger. Ces séquences flanquantes ont au moins 150 paires de bases et sont

de préférence inférieures à la longueur du gène receveur. Il est essentiel que ces deux séquences flanquantes soient homologues aux deux séquences génomiques qui jouxtent le site d'insertion souhaité. La séquence flanquante du vecteur qui se trouve en amont du gène étranger à introduire, est normalement homologue à la séquence génomique qui est située du côté 5' du site d'insertion. De la même manière, la séquence flanquante du vecteur qui se trouve en aval du gène étranger, est normalement homologue à la séquence génomique qui est située au côté 3' du site d'insertion.

Il est possible d'introduire des séquences "intercalantes" entre l'une ou l'autre des séquences flanquantes et le gène étranger, par exemple des séquences permettant la sélection des transformants, des marqueurs, des séquences permettant le clonage du vecteur, etc...

La position de ces séquences intercalantes vis-à-vis du gène étranger doit pourtant être choisie afin de ne pas empêcher l'expression du gène étranger, en particulier de la séquence d'ADN codante étrangère sous le contrôle du promoteur endogène ou, à l'inverse, la séquence codante d'ADN endogène sous le contrôle d'éléments de régulation étrangers apportés par la séquence d'insertion.

Malgré la présence des séquences flanquantes, qui encourage une recombinaison homologue, il est possible qu'un certain nombre d'intégrations se fasse au hasard. Afin de vérifier que l'insertion ciblée a bien eu lieu dans le site ciblé et non pas dans un autre endroit, la technique du "Polymerase Chain Reaction" (P.C.R.) (voir Ref. 10) est utilisée pour amplifier la séquence d'ADN du locus où l'insertion se fait, ou aurait dû se faire. De cette façon, seuls

les clones transformés à la suite d'une recombinaison homologue sont sélectionnés.

Les séquences flanquantes du vecteur sont bien évidemment choisies en fonction du site d'insertion désiré pour que la recombinaison homologue puisse avoir lieu. Le cas échéant, les séquences flanquantes peuvent comporter des séquences répliques du promoteur endogène et/ou des modifications aux séquences qui précèdent le codon d'initiation pour améliorer le taux de traduction (séquences amont) et séquences répliques des séquences de terminaison notamment des sites de polyadénylation (séquences aval).

Le gène d'insertion peut être n'importe quel gène d'intérêt. On citera comme exemples non-limitatifs, le gène lacZ (comme dans le modèle décrit plus loin), les gènes codant pour l'interleukine ou l'interféron, les gènes de récepteur, par exemple de l'acide rétinolique ou beta-3 adrénergique ou de H.I.V., et des gènes connus comme étant liés à certaines maladies, par exemple la myopathie, etc...

Les cellules eucaryotes transformées par le procédé de l'invention sont aussi englobées par la présente demande.

Selon une variante préférée de l'invention, les cellules eucaryotes sont des cellules souches embryonnaires (voir Ref. 14 et 15).

En effet, une cellule E.S. mutée peut être injectée dans un embryon précoce qui, après réimplantation, pourra naître sous une forme chimère. Si la lignée germinale est colonisée par la cellule mutée, l'animal chimère transmettra la mutation à sa descendance. Par la suite, on pourra observer les effets de cette mutation, à l'état homozygote chez certains individus, sur leur développement, leur

comportement, leur métabolisme, leur pathologie, etc..

La figure 1 montre le plasmide pGN, qui fait partie de l'invention.

Les figures 2 a et b montrent les molécules pGMA et pGMD respectivement construites à partir du plasmide pGN par rapport au gène HOXC8 (Hox-3.1). Ces plasmides sont des plasmides de mutagenèse. Les deux parties de la phase codante du gène HOXC8 (Hox-3.1) sont représentées, sur le chromosome 15, avec la boîte "homéo" en noir. Les séquences correspondantes de HOXC8 (Hox-3.1) ont été clonées dans le plasmide pGN.

(A : signal de polyadénylation; Enh/Pro : enhancer-promoteur).

07 et 08 figurent les deux oligonucléotides utilisés dans la PCR.

Les figures 3 à 6 montrent les plasmides utilisés dans la construction du pGN.

La figure 7 illustre la détection de recombinaison homologue avec la technique Réaction de Polymérase en Chaîne (P.C.R.) sur des cellules E.S. transfectées.

La figure 8 (a) et (b) montre des analyses de Southern de clones individus positifs (L5 et F2) et cellules E.S. (C.C.E.).

Le procédé de l'invention est d'application industrielle très large et peut varier selon la nature du gène étranger introduit, par exemple pour la production d'animaux transgéniques.

La génétique des mammifères va progresser de manière considérable grâce à la possibilité récente de mutagéniser spécifiquement n'importe quel gène, permettant ainsi de mieux définir son rôle. Par cette technologie qui fait intervenir recombinaison homologue et cellules E.S., des informations

précieuses seront apportées sur des oncogènes, des facteurs de croissance, des facteurs de transcription, etc..., gènes qui concernent des sujets très actuels de la recherche fondamentale ou la recherche appliquée. Un débouché important pour la recherche médicale est la possibilité de reproduire une maladie humaine dont la détermination génétique est connue (certaines maladies humaines à pathologie, telle la myopathie de Duchesne) ceci afin de mieux en étudier les mécanismes et de rechercher une thérapeutique.

En appliquant le procédé de l'invention, un gène connu comme étant responsable d'une certaine maladie est inséré d'une manière ciblée dans le génome d'une cellule E.S. L'animal transgénique qui est produit à la suite présente un modèle utile de cette maladie.

Si nécessaire, et comme décrit plus haut, les fonctionnements génétiques normaux peuvent être sensiblement maintenus, malgré l'insertion du gène étranger.

Une autre application : le procédé de l'invention consiste en le marquage génétique d'animaux. Cette application consiste à insérer un gène d'insertion qui est facilement décélé e.g. le gène lacZ et qui peut donc jouer le rôle d'un marqueur cellulaire. De cette manière, des études de filiation e.g. chez des animaux de concours sont facilitées, et la race peut être suivie.

L'insertion du gène lacZ comme gène d'insertion rend aussi possible des études de promoteur. Grâce à la possibilité de déceler l'activité β -galactosidase, l'activité et la spécificité de différents promoteurs endogènes peuvent être étudiées en ciblant différents sites dans le même type ou différents types de cellules. Les mêmes études pourront être effectuées sur un organisme entier, au cours du développement ou

à l'état adulte, en utilisant les techniques d'animaux chimères ou transgéniques.

Les inventeurs ont constaté d'une manière surprenante que la fréquence de recombinaison homologue n'est pas affectée par l'insertion de fragments de taille importante, par exemple le Lac Z. Cette observation a suggéré aux inventeurs que la technique de recombinaison homologue serait bien adaptée à l'insertion d'autres gènes hétérologues qui sont de taille importante.

L'invention vise également l'application du procédé de remplacement ou d'insertion spécifique pour la thérapie génique.

Grâce à la possibilité de modifier le génome d'un animal, le procédé de l'invention peut donc également être utilisé en tant que "thérapie génique". Les utilisations les plus évidentes consisteraient à inactiver les gènes de récepteurs d'agents infectieux (virus ou bactéries) ou toxiques. Si cette mutagenèse s'avérait létale, il faudrait rétablir la fonction perdue sans rétablir la sensibilité aux agents nuisibles. Un gène modifié codant pour un tel récepteur pourrait être réintroduit dans la cellule mutée, à moins que la modification puisse être provoquée par la recombinaison homologue. Cette modification du patrimoine génétique conférerait à l'animal une immunité contre la maladie considérée.

Ce protocole peut aussi intervenir dans le cadre d'auto-greffe. Des cellules, malades ou saines, prélevées sur un patient, pourraient être soignées et immunisées, puis réimplantées chez le même individu.

La technique de l'invention se prête aussi aux études d'activité de produits pharmaceutiques présumés avoir une activité à l'égard des produits d'expression d'un gène pathologique lié à une

maladie. Dans ce cas, le gène d'insertion est constitué par le gène pathologique ou un fragment de celui-ci et on administre à l'animal transgénique le produit pharmaceutique à tester, en vue d'évaluer son activité sur la maladie.

L'invention va être illustrée en faisant référence au plasmide pGN et son utilisation dans l'insertion ciblée d'un gène étranger (lacZ, codant pour l'enzyme β -galactosidase d'E.Coli) dans le génome d'une cellule d'E.S. de souris. Le gène lacZ a été choisi en raison du fait que son expression peut être facilement décelée et est simplement à titre illustratif.

La phase codante de l'enzyme β -galactosidase d'E.Coli (lacZ; 1-3057), fusionnée avec une séquence génomique (7292-3) du gène murin HOXC8 (Hox.3-1) (Ref. 1), débute par le codon d'initiation de ce gène. En effet, la séquence qui précède le codon d'initiation de HOXC8 (Hox-3.1) est identique à la séquence consensus observée chez les vertébrés (Ref. 2), permettant ainsi un meilleur taux de traduction de la β -galactosidase dans les cellules de vertébrés. Le gène lacZ est suivi d'un signal de polyadénylation par exemple du virus SV 40, comme la plupart des gènes eucaryotes, afin de stabiliser les ARN_m messagers.

L'activité de la β -galactosidase d'E.Coli, qui est fonctionnelle dans les cellules eucaryotes, peut être décelée de différentes manières. Des cellules exprimant le gène lacZ prennent une coloration bleue, après fixation, en présence de X-Gal, qui est un substrat de la β -galactosidase (Ref. 3). Un nouveau substrat, le FDG (flúrorosceine di- β -galactopyranoside), permet de déceler et de doser l'activité β -gal. tout en gardant les cellules vivantes (Ref. 4). Les cellules exprimant lacZ

accumulent un produit fluorescent et peuvent être isolés à l'aide d'un trieur de cellules ou FACS (fluorescence-activated cell sorter).

L'unité de transcription du gène de résistance à la néomycine provient, en grande partie, du plasmide pRSV neo (Ref. 5). Le LTR (long terminal repeat) du virus du sarcome de Rous procure des séquences activatrice et promotrice très puissantes dans de nombreuses cellules eucaryotes (Ref. 6). Du transposon bactérien Tn5, viennent un promoteur actif dans E.Coli et la phase codante de l'enzyme phosphotransférase (Ref. 7), qui est suivie du signal de polyadénylation du virus SV40. Le même gène sous le double contrôle des promoteurs RSV et Tn5 peut conférer la résistance à la néomycine ou la kanamycine aux bactéries et la résistance au G418 aux cellules eucaryotes.

Par l'effet d'une simple mutation ponctuelle, l'unité B des séquences activatrices (enhancer) de la souche PyEC F9.1 du virus du Polyome est devenue beaucoup plus active dans différents types de cellules, et en particulier dans les cellules de carcinome embryonnaire (EC) (Ref. 8). Deux copies de cet enhancer Py F9.1 ont été insérées en tandem dans le plasmide pGN, en amont du LTR-RSV, et dans l'orientation "promoteur tardif" de la région régulatrice du Polyome.

Afin d'améliorer le taux de traduction de la phosphotransférase, la séquence précédant le codon d'initiation a été modifiée lors d'une mutagenèse par oligonucleotide. Ainsi la séquence T T C G C A U G est devenue G C A C C A U G, correspondant beaucoup mieux à la séquence consensus d'initiation de la traduction chez les vertébrés (Ref. 2).

Les améliorations apportées à l'unité de transcription du gène de résistance à la néomycine

ont pu être estimées en transfectant des cellules souches embryonnaires (ES) de souris. A molarité égale en plasmide, une construction avec les enhancer Py F9.1 a produit 7,5x plus de clones résistants au G418 que le pRSV neo et 2 à 3x plus que le pMC1 Neo décrit par Capecchi et al (réf. 13). A nouveau, le nombre de clones a été augmenté 60x, soit 450x par rapport au pRSV neo, en modifiant la séquence d'initiation de la traduction. La recombinaison homologue peut être un événement assez rare, selon les conditions expérimentales appliquées (p. ex 1/1000 pour HPRT, réf. 13). Un vecteur présentant une efficacité de sélection élevée est donc très utile, d'autant plus que les conditions d'électroporation donnent lieu principalement à l'intégration d'une seule copie.

Le plasmide pGN contient, en outre, une origine de répllication bactérienne de type colE1, pBR322, qui permet les clonages et les préparations dans E.Coli.

Enfin, un site de clonage multiple (M.C.S.), synthétisé in vitro, qui ne contient que des sites de coupure uniques dans pGN, a été inséré en amont de lacZ, afin de faciliter les utilisations de ce plasmide.

Les séquences "flanquantes" plasmidiques qui provoquent la recombinaison homologue sont ajoutées aux extrémités du plasmide pGN après linéarisation du plasmide en amont de lacZ, par un site du MCS (voir fig. 2). En l'occurrence, les séquences flanquantes choisies sont homologues aux séquences chromosomiques issues de HOXC8 (Hox-3.1) devant ultérieurement intervenir dans la recombinaison homologue.

La figure 2 situe la molécule construite à partir du plasmide pGN par rapport au gène HOXC8 (Hox-3.1). Dans ce cas, une recombinaison entre les séquences plasmidiques et chromosomales de HOXC8

(Hox-3.1) resulterait en une insertion au début de la phase codante de ce gène, donc à son inactivation totale.

Le plasmide pGN rassemble plusieurs avantages pour cette méthodologie, qui est applicable à n'importe quel gène. L'évènement de recombinaison homologue pouvant être assez rare (de l'ordre de 1 pour 1000 intégrations non-homologues), il est nécessaire de pouvoir analyser un grand nombre de clones dont la résistance au G418 soit suffisamment forte pour s'exprimer dans n'importe quelle partie du génome. Les modifications apportées à l'unité de transcription de la phosphotransférase répondent parfaitement à ces problèmes. La méthode de mutagenèse par recombinaison homologue équivaut à inactiver un gène par une insertion ou une substitution, mais le plasmide pGN présente l'avantage supplémentaire de pouvoir substituer l'expression de la β -galactosidase à celle du gène muté. Enfin, le MCS facilite les clonages de fragments génomiques.

EXEMPLES :

I - Construction du plasmide pGN

Les plasmides intermédiaires sont numérotés selon leur étape.

1^{re} étape :

Insertion d'un site Xho I dans le site Bgl I de pRSV neo

Insertion d'un linker Xho I dans le site Bgl I de pRSV neo, rempli par le fragment Klenow de l'ADN polymérase d'E.Coli.

2^e étape :

Insertion d'un site Cla I dans le site Nde I du plasmide p1

Insertion d'un linker Cla I dans le site Nde I de p1, rempli par la polymérase Klenow.

3^e étape :

Insertion d'enhancer Py F9.1 dans le site Cla I du plasmide p2

Insertion d'enhancer Py F9.1 Pvu II-Pvu II isolé par un site unique, Acc I, dans le site Cla I de p2. Sélection d'un clone contenant deux enhancers orienté dans le sens "promoteur tardif".

4^e étape :

Délétion Sma I-Hpa I du plasmide p3

Les deux enzymes, donnant des extrémités "bouts-francs", peuvent être ligués directement. Cette délétion enlève l'intron de l'antigène t de SV 40, qui n'est pas très utile, et diminue de manière appréciable la taille de l'unité de transcription de la phosphotransférase.

5^e étape :

Insertion d'un site Xho I dans le site Bam HI de pCH110

Insertion d'un linker Xho I dans le site Bam HI du plasmide pCH 110 (Pharmacia), rempli par la polymérase Klenow.

6^e étape :

Insertion du 3' lacZ-polyA SV 40 dans le plasmide p4

La partie 3' de la phase codante de la β -galactosidase, suivie du signal de polyadénylation du virus SV 40 est isolée du plasmide p5 par les sites Xho I-Aat II et clonée dans le plasmide p4 par les mêmes sites.

7^e étape :

Insertion du 5' lacZ dans le vecteur KS-

La partie 5' de la phase codante de la β -galactosidase est isolée du plasmide pMC 1871 (Pharmacia) par les sites Pst I-Sac I et clonée dans le vecteur KS- (Stratagene) par les mêmes sites.

8^e étape :

Fusion d'une séquence génomique HOXC8 (Hox-3.1) avec le 5' lacZ

Une séquence génomique du gène HOXC8 (Hox-3.1), clonée dans le vecteur KS-, est purifiée par digestions successives par l'enzyme Sac I, puis par la nucléase Mung bean et enfin par l'enzyme Apa I. Cet insert est fusionné avec la partie 5' de la phase codante de la β -galactosidase par clonage dans le plasmide p7 digéré par Apa I-Sma I. La protéine ainsi fusionnée contient le codon d'initiation de la traduction du gène HOXC8 (Hox-3.1) suivi de la phase codante de la β -galactosidase (vérifiée ensuite par séquençage).

Met Ser Ser	Ile Pro Gly Asp Pro
CCAGC ATG AGC TCC	ATT CCC GGG GAT CCC
GGTCG TAC TCG AGG	TAA GGG CCC CTA GGG
↓ Sac I	
CCAGC ATG AGC T	↓ Sma I
GGTCG TAC	
↓ nucléase Mung bean	
CCAGC ATG	GGG GAT CCC
GGTCG TAC	CCC CTG GGG
	⇒
Met Gly Asp Pro	
CCAGC ATG GGG GAT CCC	
GGTCG TAC CCC CTA GGG	

9^e étape :

Insertion de HOXC8 (Hox-3.1)-5' lacZ dans le plasmide p6

La fusion HOXC8 (Hox-3.1)-5' lacZ est isolée du plasmide p8 par les sites Apa I-Sac I et clonée dans le plasmide p6 par les mêmes sites. Ce clonage a pour effet de reconstituer la phase codante de la β -galactosidase dans sa totalité.

10^e étape :

Insertion du gène Neo^R dans le vecteur KS+

23

Le gène de résistance à la néomycine (promoteur bactérien et phase codante de la phosphotransférase) est isolée du pRSV neo par les sites Hind III-Eco RI et clonée dans le vecteur KS+ (Stratagene).

11^e étape :

Mutagenèse de la séquence d'initiation de Neo^R dans p10

La séquence d'initiation de la traduction de la phosphotransférase est modifiée pour être identique à la séquence consensus observée chez les vertébrés et permettre ainsi un taux supérieur d'initiation de la traduction donc une résistance accrue au G418 pour les cellules de mammifères. La modification crée également un site Apa LI qui permet de contrôler l'efficacité de la mutagenèse.

Apa LI
GTTTCGCAIG ⇒ GTGCACCAIG

Un oligonucléotide (CTTGTTC AATCATGGTGCACGATCCTCA) comportant une région de misappariement avec la séquence du pRSV neo (souligné) est synthétisé (Gene Assembler, Pharmacia) puis phosphorylée par la polynucléotide kinase du bactériophage T4. Une matrice simple brin du plasmide p10 est préparée grâce à l'origine fl du plasmide KS+ et hybridée avec l'oligonucléotide de mutagenèse. Le deuxième brin est synthétisé et réparé par la polymérase Klenow et l'ADN ligase du bactériophage T4. Après transformation de bactéries, les clones mutés sont criblés à l'aide de l'oligonucléotide marqué au ³²P. La mutagenèse a été vérifiée en digérant par Apa LI ainsi que par séquençage.

12^e étape :

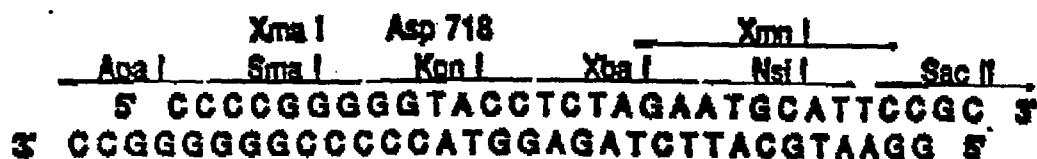
Remplacement de la séquence d'initiation dans le plasmide p9

Un fragment contenant la séquence modifiée d'initiation de la traduction du gène de résistance à la néomycine est isolée du plasmide p11 par les enzymes Hind III-Eag I et clonée dans le plasmide p9 par les mêmes sites.

13^e étape :

Insertion du site de clonage multiple dans le plasmide p12

Deux oligonucléotides complémentaires sont synthétisés (Gene Assembler, Pharmacia) puis phosphorylés. Après appariement, le MCS est cloné dans les sites Apa I- Sac II du plasmide p12 grâce à ses extrémités cohésives



Le site de clonage multiple a également été vérifié par séquençage.

II - Addition des séquences "flanquantes" aux extrémités du plasmide pGN linéarisé en amont de lacZ' par un site du M.C.S

Les séquences flanquantes utilisées ont été choisies en fonction du site d'insertion souhaité (par exemple, HOXC8 (Hox-3.1), voir Fig. 2 a et b pGMA et pGMD).

Dans la construction du plasmide de mutagenèse pGMD, deux bras d'ADN homologue au locus HOXC8 (Hox-3.1) ont été clonés aux sites Apa I-Nsi I et Nsi I-Sac II du vecteur pGN. Le bras 5' commence au site Sac II (CCGCGG) au nucléotide 219 de l'ADNc c21 de HOXC8 (Hox-3.1). Ce fragment s'étend sur 6.8 kb en 5' jusqu'au premier site BamHI. Le bras 3' commence au

site Apa 1 (GGGCCC) au nucléotide 885 de l'ADNc c21. Ce fragment s'étend sur 1.5 kb en 3' jusqu'au premier site PstI. Un linker NsiI a été inséré dans le site BamHI du fragment 5' et dans le site PstI du fragment 3'. Les bras 5' et 3' ont été clonés dans le vecteur pGN dans les sites Nsi I-Sac II et Apa I-Nsi I, respectivement. La séquence de l'ADNc de HOXC8 (Hox-3.1) c21 est publiée (réf. 1).

Le plasmide de mutagenèse est linéarisé par digestion avec Nsi I avant électroporation de cellules E.S. Ses extrémités sont formées des deux bras génomiques clonés aux sites Apa I-Nsi I et Nsi I-Sac II du vecteur pGN.

Le plasmide pGMD ne présente pas de signal de polyadénylation après le gène de résistance mais, en revanche, présente une séquence riche en AU responsable d'une dégradation sélective de mRNA, insérée dans la séquence de l'intron du HOXC8 (Hox-3.1) du plasmide.

Un autre plasmide de mutagenèse, pGMA, présente la même structure que le pGMD mais contient les signaux de polyadénylation et de terminaison de transcription du SV40 et ne présente pas la séquence AU de dégradation de mRNA en aval du gène Neo^r. Ces modifications avaient pour but de réduire le taux de transcrits de Neo^r dans des clones issus d'intégration au hasard. En revanche, des clones issus d'événements de recombinaison homologue entre pGMD et un locus HOXC8 (Hox-3.1) devrait avoir une croissance inaltérée pendant la sélection au G418, la séquence AT de dégradation de mRNA étant éliminée par le procédé de recombinaison même, ou épissée avec l'intron HOXC8 (Hox-3.1).

Dans les étapes expérimentales qui suivent, le protocole décrit par Thompson et al. 1989, a été suivi pour la production d'animaux chimères.

III - Transfection de cellules embryonnaires de souris

La méthode décrite par Thompson et al. 1989, a été utilisée pour transfecter des cellules embryonnaires de souris. L'utilisation de la technique de l'électroporation assure l'introduction d'une seule copie du gène étranger (lacZ) par cellule. Après transfection, plusieurs clones exprimant la β -galactosidase ont été isolés.

Les plasmides de mutagenèse pGMD et pGMA ont été linéarisés et introduits par électroporation dans des cellules E.S. afin de favoriser l'insertion d'une copie seulement dans le génome (réf. 11).

Les transfections initiales ont été effectuées pour comparer l'efficacité de ciblage du HOXC8 (Hox-3.1) des plasmides pGMA et pGMD (voir tableau I).

Recombinaison homologue dans le gène HOXC8 (Hox-3.1)

Exp.	Plasmide de mutagenèse	N° de l'ensemble analysé	Nb de clones formant l'ensemble	Nb de résultats P.C.R. positifs
I	pGMA	3	600	0(2)
II	pGMD	5	250	3(5)
III	pGMD	84	2-3	5(5)

Tableau I

La lignée cellulaire E.S. "C.C.E." (réf. 16) a été maintenue d'une manière continue sur des couches nourricières fibroblastiques (réf. 17). Pour les expériences I et II, 1.5×10^7 cellules E.S. dans 1.5 ml HeBS ont été électroporées (réf. 11) à 200 V, avec 40 mg de plasmide linéarisé, puis étalées sur quatre boîtes de cultures (diamètre 100 mm). Pour

l'expérience III, le choc a été effectué dans les mêmes conditions mais un quart des cellules ont été étalées sur quatre plaques à 24 puits. Le lendemain, 250 $\mu\text{g ml}^{-1}$ G418 ont été ajoutés. Chaque transfection a donné lieu à environ 2400 clones avec pGMA et environ 1000 clones avec pGMD.

Le nombre moyen de clones de cellules E.S. résistantes au G418 dans chaque ensemble est indiqué dans le tableau I, ainsi que le nombre d'ensembles donnant un résultat positif avec la technique P.C.R. Un résultat positif signifie qu'une bande de 1.6 Kb a pu être observée sur un gel d'agarose coloré de bromure d'ethidium (voir fig. 7). Le nombre d'ensembles donnant un signal positif après une analyse de Southern du mélange P.C.R. et hybridation d'une sonde spécifique qui ne contenait pas les séquences des amorces est indiqué entre parenthèses (fig. 8).

Détection de la recombinaison homologue avec la P.C.R.

P.C.R. a été effectuée sur les cellules d'un ensemble de 250 clones de la transfection II (voir voie D de la fig. 7). Dans les autres voies, quatre ensembles de la transfection III ont été analysés ensemble en mélangeant environ 4 x 5000 cellules. Les amorces 07 et 08 utilisées dans la P.C.R. entourent la séquence 3'-HOXC8 (Hox-3.1) du plasmide de mutagenèse (fig. 2). Le fragment de 1.6 Kb recouvrant cette séquence 3' ne peut être amplifié que dans le cas d'une recombinaison homologue. Les voies 2, 3 et D illustrent des résultats positifs.

L'ADN de clones E.S. a été préparé au moment de la réplique sur filtre en utilisant la méthode "boiling-proteinase K digestion boiling" (réf. 18). 40 cycles d'amplification (40 secondes à 94°C, 1 minute à 60°C, 7 minutes à 72°C) ont été effectués

dans un mélange réactionnel de 100 μ l, contenant 67 mM Tris-HCL (pH 8.6), 16.7 mM $(\text{NH}_4)_2\text{SO}_4$, 6.7 mM MgCl_2 , 10 mM 2-mercaptoethanol, 0.01 % (p/v) gélatine, 200 μ M dATP, dTTP et dCTP, 100 μ M dGTP, 100 μ M 7-deaza dGT, 600 ng de chaque amorce (07 : AACTTCCCTCTCTGCTATTC et 08 : CAGCAGAAACATACAAGCTG) et 3U polymérase Taq (Perkin Elmer Cetus), couvert de 100 μ l paraffin. La moitié du mélange de réaction a été appliquée sur un gel d'agarose 0.7 % coloré au bromure d'ethidium. Le marqueur de taille est un digest Eco RI + Hind III d'ADN lambda.

Analyses de Southern

Trois clones indépendants de cellules E.S. contenant le HOXC8 (Hox-3.1) muté (identifié par P.C.R.) ont été isolés des ensembles positifs en utilisant des pipettes. Leur ADN a été examiné par analyse de Southern après digestion avec les enzymes de restriction indiqués dans la figure 8, afin de confirmer le ciblage spécifique et de faire la distinction entre les loci recombines et sauvages. Deux sondes différentes ont été utilisées dans l'analyse de l'extrémité 3' des loci HOXC8 (Hox-3.1) dans les clones mutés et dans les cellules E.S. non-mutées agissant comme témoin (fig. 8 c). La première sonde "a" était contenue dans les séquences HOXC8 (Hox-3.1) du plasmide de mutagenèse et démontrait le nombre d'intégrations de vecteur et leurs liaisons physiques. Un des trois clones recombines contenait en outre une copie du plasmide intégrée au hasard (fig. 8 a, clone F2). La deuxième sonde "b" qui n'était pas contenue dans le vecteur de mutagenèse faisait la distinction entre les allèles HOXC8 (Hox-3.1) recombines et sauvages (fig. 8 b). Le locus recombine HOXC8 (Hox-3.1) présentait, avec les deux sondes, l'image d'hybridation attendue à partir des cartes de restriction du vecteur de mutagenèse et

du locus intact. En outre, l'existence de deux domaines de recombinaison dans le bras 3' du vecteur a été confirmée par la présence ou l'absence de la séquence AT dans le locus HOXC8 (Hox-3.1) recombiné (par exemple fig. 8, clone L5). L'extrémité 5' du locus HOXC8 (Hox-3.1) a également été analysée pour l'événement de recombinaison homologue. Des enzymes de restriction ne présentant pas de sites dans la séquence HOXC8 (Hox-3.1) 5' de 6.8 Kb du vecteur de mutagenèse ont été utilisés dans la digestion des ADNs des clones recombinés. Ces ADNs ont ensuite été soumis à une électrophorèse dans un champ pulsé pour différencier les fragments de poids moléculaire élevé. Une analyse de Southern de ce gel a également indiqué les allèles recombinés correctement et les allèles HOXC8 (Hox-3.1) sauvages, en utilisant une sonde présentant une séquence en amont du plasmide de mutagenèse.

Les analyses de Southern ont démontré qu'une allèle du gène HOXC8 (Hox-3.1) a été recombinée comme prévu. La recombinaison était équivalente à un double "crossing-over" entre les bras génomiques du plasmide de mutagenèse et les séquences chromosomales homologues (Fig. 2).

Dans les clones recombinants, le gène lac Z a été placé sous le contrôle des séquences promotrices et régulatrices du HOXC8 (Hox-3.1) en amont du codon AUG, mais les signaux 3' de maturation de mRNA provenaient du SV40. Dans ces clones recombinés, l'expression de lac Z n'était pas décelable par coloration avec β -Gal, ce qui est cohérent avec l'absence de transcription de HOXC8 (Hox-3.1) dans des cellules E.S. déterminée par analyse de protection de RNase. L'activité de β -Gal pouvait être induite dans certaines cellules après 3 ou 4 jours de culture en présence de $5 \cdot 10^{-7} M$ acide rétinolique,

conditions connues comme induisant la transcription de HOXC8 (Hox-3.1) (réf. 19).

En utilisant le vecteur de mutagénèse pGMA, qui présente une homologie totale de 8.3 Kb ADN avec le locus HOXC8 (Hox-3.1), un fragment de 120 pb a été remplacé par une insertion de 7.2 Kb. La fréquence de ce remplacement ciblé (1/900) est comparable à celle obtenue récemment (1/1000) avec HPRT (réf. 13) ou avec En-2 (1/260) (réf. 20), le fragment hétérologue inséré étant cependant dans ces derniers cas d'une taille beaucoup moins importante (1.1 et 1.5 Kb respectivement). D'une manière surprenante, il a été constaté qu'une fréquence de recombinaison homologue très élevée (1/40) a pu être obtenue avec le vecteur pGMD. L'élimination des signaux 3' de maturation de mRNA et l'addition de la séquence de dégradation de mRNA au gène de résistance à la néomycine a eu pour effet de réduire le nombre total de clones résistants au G418 par 2.4 (tableau I). Le rapport de ciblage spécifique était presque 10 fois plus élevé (900/40). Le mécanisme de recombinaison homologue même a dû être affecté dans les expériences avec pGMD. Une explication possible de ces résultats serait qu'une séquence AT de 51 pb pourrait fournir, in vivo, une boucle ouverte dans le plasmide de mutagénèse à cause de sa température de fusion plus basse. Si les séquences HOXC8 (Hox-3.1) voisines du pGMD peuvent être influencées par cette ouverture, de chaque côté de la région AT, elles pourraient réagir d'une manière plus efficace, à l'état simple-brin, avec le locus chromosomal HOXC8 (Hox-3.1). Le modèle de recombinaison mitotique chez la levure suggère qu'il serait initié par un tel échange de brins, bien que le mécanisme de recombinaison homologue reste inconnu chez les eukaryotes plus complexes.

La figure 8 montre les résultats de l'analyse de Southern effectuée sur des clones individus positifs (L5 et F2) et des cellules E.S. (C.C.E.).

Les sondes utilisées n'hybrident qu'avec des séquences HOXC8 (Hox-3.1) incluses dans le vecteur (a) ou exclues du vecteur de mutagenèse (b). L'image d'hybridation du locus HOXC8 (Hox-3.1) recombiné (triangles ouverts) se distingue clairement du locus sauvage (triangles noirs). Les étoiles indiquent les bandes d'hybridation d'une copie du plasmide qui s'est intégrée au hasard. Le marqueur de taille est un digest Eco RI + Hind III d'ADN lambda.

La figure 8(c) montre les cartes de restriction des allèles HOXC8 (Hox-3.1) recombinées (rec.) et sauvages (wt). Les parties du vecteur de mutagenèse et du locus HOXC8 (Hox-3.1) sont indiquées avec les mêmes symboles que ceux utilisés dans la figure 2. Dans ce cas, la séquence AT a été intégrée par recombinaison homologue. La flèche verticale indique l'extrémité 3' du plasmide de mutagenèse. La localisation des sondes "a" et "b" utilisées dans l'analyse de Southern est également indiquée. Les abréviations utilisées dans la figure 8 sont les suivantes : B, Bam HI ; D, Dra I, E, Eco RI ; H, Hind III ; S, Sal I ; X, Xho I.

IV - Production d'embryons chimères

Une microinjection dans des blastocystes a été effectuée avec deux clones E.S. recombinants contenant un allèle HOXC8 (Hox-3.1) intact et un allèle recombiné, ces clones ne contenant aucune autre copie du plasmide de mutagenèse. Les karyotypes de ces cellules étaient normaux.

Dix à quinze cellules mutées ont été microinjectées par blastocyste. Après réimplantation dans des mères porteuses, les embryons ont été recueillis à 9.5, 10.5 et 12.5 jours p.c. et analysés

pour expression de lac Z. La grille de transcription de HOXC8 (Hox-3.1) à ces stades avait été déterminée au préalable par analyse d'hybridation in situ (réf. 1). Les transcrits HOXC8 (Hox-3.1) sont détectables pour la première fois au stade de gastrulation tardive et sont répartis dans tous les tissus de la partie postérieure de l'animal. Plus tard, la répartition devient progressivement limitée dans l'espace et spécifique au tissu. Au stade de 12.5 jours p.c., la transcription est localisée dans la région cervicale du tube neural, au niveau du coeur. Au cours de l'embryogenèse, la répartition de la transcription de HOXC8 (Hox-3.1) subit donc des modifications. Le stade 10.5 jours p.c. semble être une période de transition, la transcription ayant lieu à la fois dans les deux régions arrières et dans le tube neural cervical.

Dans des embryons chimériques de 9.5 et 10.5 jours p.c., la partie caudale au bourgeon postérieur présentait une activité β -Gal intense, tandis que le marqueur n'a jamais été détecté dans la région thoracique antérieure ou la tête (Fig. 9a). Dans la région postérieure, des cellules colorées par le β -Gal ont été observées dans tous les tissus et de toutes les couches embryonnaires. Entre les deux bourgeons qui donnent les membres, des cellules colorées étaient réparties dans des zones restreintes, dans l'ectoderme superficiel (Fig. 9b), comme dans les régions arrières (Fig. 9c) et, en forme de lignes étroites ou rayures, dans le tube neural (Fig. 9b). Ces rayures présentaient une répartition irrégulière et asymétrique sur la paroi du tube neural. La transcription de HOXC8 (Hox-3.1) n'a pas été détectée dans la couche mince de cellules vers la fermeture du tube neural. Ces cellules n'ont peut-être pas résisté aux traitements appliqués lors

de l'hybridation *in situ*. Il a été observé que les cellules de l'ectoderme neural font partie, très tôt, de parties différentes du système nerveux et migrent dans une direction radiale, suivant des mouvements latéraux étroits (réf. 21). Ces résultats sont donc cohérents avec cette observation.

L'expression de Lac Z a donc illustré correctement la première partie de la transcription de l'homéogène HOXC8 (Hox-3.1), c'est-à-dire dans tous les tissus des régions caudales des embryons de 9.5 et 10.5 jours p.c., et a fourni de nouvelles informations concernant le mode de transcription de HOXC8 (Hox-3.1).

En revanche, l'expression de Lac Z n'a pas été observée dans les régions cervicales du tube neural d'embryons chimères de 12.5 jours, ni dans la région antérieure d'embryons de 10.5 jours ; ceci n'était pas le résultat attendu à partir des études d'hybridation *in situ*. La phase ultérieure de transcription de HOXC8 (Hox-3.1) observée à partir du jour 10.5 dans les zones latérales localisées du tube neural n'était pas mise en évidence par l'activité de β -Gal. Une explication possible pour ce résultat serait que, bien que l'expression de Lac Z soit sous le contrôle du promoteur HOXC8 (Hox-3.1), les séquences 3' du HOXC8 (Hox-3.1) sont absentes dans le gène reporteur. Il est possible que des séquences 3' du codon d'initiation AUG du HOXC8 (Hox-3.1) aient une influence sur l'expression tardive de HOXC8 (Hox-3.1) dans le domaine antérieur. Un effet de "dosage de gène" pourrait aussi expliquer ce résultat. L'autoactivation de plusieurs homéogènes chez *Drosophila* a été démontrée génétiquement ou suggérée par la formation de complexes entre l'ADN et les protéines des homeobox.

Si le composant tardif de la transcription de HOXC8 (Hox-3.1) dans le tube neural est maintenu par un mécanisme semblable, l'inactivation d'un allèle pourrait avoir un effet dominant dans les cellules de l'ectoderme neural. Puisqu'un allèle seulement produirait la protéine HOXC8 (Hox-3.1), le signal d'activation serait dilué sur les deux promoteurs. La réduction d'autoinactivation dans les deux loci pourrait ainsi conduire à un arrêt total de l'initiation de transcription. Ceci pourrait expliquer pourquoi aucune expression de Lac Z n'a été détectée dans la région cervicale du tube neural d'embryons de 10.5 et 12.5 jours.

V - Passage de la modification dans la lignée cellulaire germinale : production d'animaux transgéniques

Les effets en F₁ et en F₂ de la modification apportée par l'insertion ciblée ont été observés après reproduction des chimères. Le passage de la modification dans la lignée cellulaire germinale a été constaté.

BIBLIOGRAPHIE

1. Le Mouellic, H., Condamine, H. et Brûlet, P. (1988). Genes & Development. 2, 125-135.
2. Cavenir, D.R. (1987). Nucleic Acids Res. 15, 1353-1361.
3. Sanes, J.R. Rubenstein, J.L.R. et Nicolas, J.F. (1986) EMBO J. 5, 3133-3142.
4. Nolan, G.P., Fiering, S., Nicolas, J.F. et Herzenberg, L.A. (1988) Proc. Natl. Acad. Sci. USA 85, 2603-2607.
5. Gorman, C., Padmanabhan, R. et Howard, B.H. (1983). Science. 221, 661-553.
6. Gormann, C.M., Merlino, G.T. Willingham, M.C. Pastan, I. et Howard, B. H. (1982) Proc. Natl. Acad. Sci. USA 79, 6777-6781.
7. Southern, P.J. et Berg, P. (1982) J. Mol. Appl. Genet 1, 327- 341.
8. Herbomel, P., Bourachot, B. et Yaniv, M. (1984). Cell. 39, 653-662.
9. Robertson, E.J. (1987) Teratocarcinomas and embryonic stem cells: A critical approach. IRL Press, Oxford.
10. Kahn, A. (1988). Médecine/Sciences 4, 515-518.
11. G. Chu, Hayakawa H., Berg. P. (1987) Nucleic Acid Research 15, Nr. 3, 1311-1326.
12. Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., Melton, D.W., (1989) Cell, 56, 313-321.
13. Thomas, K.R., Capecchi, M.R., (1987) Cell, 51 503-512.
14. Evans, M.J., Kaufmann, M.H. (1981) Nature, 292, 154-155,
15. Robertson, E.J., (1986) Trends in Genetics, 9-13
16. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Nature 323, 445-448 (1986).

36

17. Robertson, E.J. in Teratocarcinomas and Embryonic Stem Cells (ed. Robertson, E.J.) 71-112 (IRL, Oxford, 1987).
18. Kim, H.S. & Smithies, O. Nucleic Acids Res. 16, 8887-8903 (1988).
19. Breier, G., Bucan, M., Francke, U., Colberg-Poley, A.M. & Gruss, P. EMBO J. 5, 2209-2215 (1986).
20. Joyner, A.L., Skarnes, W.C. & Rossant, J. Nature 338, 153-156 (1989).
21. McKay, R. D. G. Cell 58, 815-821 (1989).



REVENDICATIONS

1. Animal transgénique ou chimérique comportant des cellules dont le génome comprend :

- un ADN hétérologue, dit ADN d'insertion, et étant constitué par la totalité d'un gène ou par une partie d'un gène ;
- l'ADN d'insertion étant positionné dans le génome par rapport à un gène endogène, dit gène receveur, comprenant un ADN de complément, de façon à ce que l'ADN d'insertion ou l'ADN de complément est rendu fonctionnel, ou son fonctionnement est rendu plus efficace grâce à l'association de l'ADN d'insertion et l'ADN de complément.

2. Animal selon la revendication 1, caractérisé en ce que l'ADN d'insertion contient soit une séquence codante, dont l'expression a lieu sous le contrôle des séquences régulatrices du gène receveur, soit une séquence régulatrice permettant l'expression d'une séquence codante du gène receveur.

3. Animal selon la revendication 2, caractérisé en ce que l'ADN d'insertion est constitué d'une séquence codante autre que la séquence codante d'un gène codant pour un agent de sélection.

4. Animal selon la revendication 1, caractérisé en ce que l'ADN d'insertion comprend la totalité d'un gène.

5. Animal selon l'une quelconque des revendications précédentes caractérisé en ce que le gène receveur est un gène qui normalement n'est pas transcrit dans la cellule.

6. Animal selon l'une quelconque des revendications précédentes caractérisé en ce que le gène receveur est inactivé suite à l'insertion de l'ADN d'insertion, l'expression de l'ADN d'insertion se substituant à celle du gène receveur.



7. Animal selon l'une quelconque des revendications précédentes caractérisé en ce que l'ADN d'insertion remplace une partie du gène receveur.

8. Animal selon l'une quelconque des revendications précédentes caractérisé en ce que le gène receveur est, hormis la présence de l'ADN d'insertion, dans son environnement génétique naturel.

9. Animal selon l'une quelconque des revendications précédentes caractérisé en ce que le génome contient, adjacent à l'ADN d'insertion, une séquence dite "intercalante" comprenant par exemple une séquence codant pour un agent de sélection avec un promoteur autorisant l'expression de l'agent de sélection, ou un marqueur.

10. Animal transgénique dans lequel une seule copie d'un gène qui est présent dans le génome à au moins deux exemplaires, a été inactivé par l'insertion d'un gène qui est différent du gène inactivé, le gène d'insertion étant inséré dans une position qui permet l'expression de ce gène sous le contrôle des séquences régulatrices du gène endogène inactivé.

11. Animal selon l'une quelconque des revendications précédentes utilisé pour un marquage génétique.

12. Procédé de criblage de produits pharmaceutiques présumés avoir une activité à l'égard des produits d'expression d'un gène pathologique lié à une maladie, caractérisé par l'administration du produit à un animal transgénique selon l'une quelconque des revendications 1 à 11, l'ADN d'insertion étant constitué par le gène pathologique ou un fragment de celui-ci.



13. Acide nucléique issu de la recombinaison homologue entre un ADN d'insertion et un gène receveur, ledit acide nucléique comprenant :

- un ADN d'insertion, hétérologue vis-à-vis du gène receveur, l'ADN d'insertion étant constitué par la totalité d'un gène ou par une partie d'un gène ;
- un gène receveur ;
- l'ADN d'insertion étant positionné par rapport au gène receveur comprenant un ADN de complément de façon à ce que l'ADN d'insertion ou l'ADN de complément est rendu fonctionnel, ou son fonctionnement est rendu plus efficace grâce à l'association de l'ADN d'insertion et l'ADN de complément.

14. Cellule d'animal transgénique selon l'une quelconque des revendications 1 à 10.

15. Lignée cellulaire issue d'un animal transgénique selon l'une quelconque des revendications 1 à 10.

ABREGE DESCRIPTIF

L'invention concerne un animal transgénique ou chimérique comportant des cellules dont le génome comprend :

- un ADN hétérologue, dit ADN d'insertion, et étant constitué par la totalité d'un gène ou par une partie d'un gène ;
- l'ADN d'insertion étant positionné dans le génome par rapport à un gène endogène, dit gène receveur, comprenant un ADN de complément, de façon à ce que l'ADN d'insertion ou l'ADN de complément est rendu fonctionnel, ou son fonctionnement est rendu plus efficace grâce à l'association de l'ADN d'insertion et l'ADN de complément.

1/12

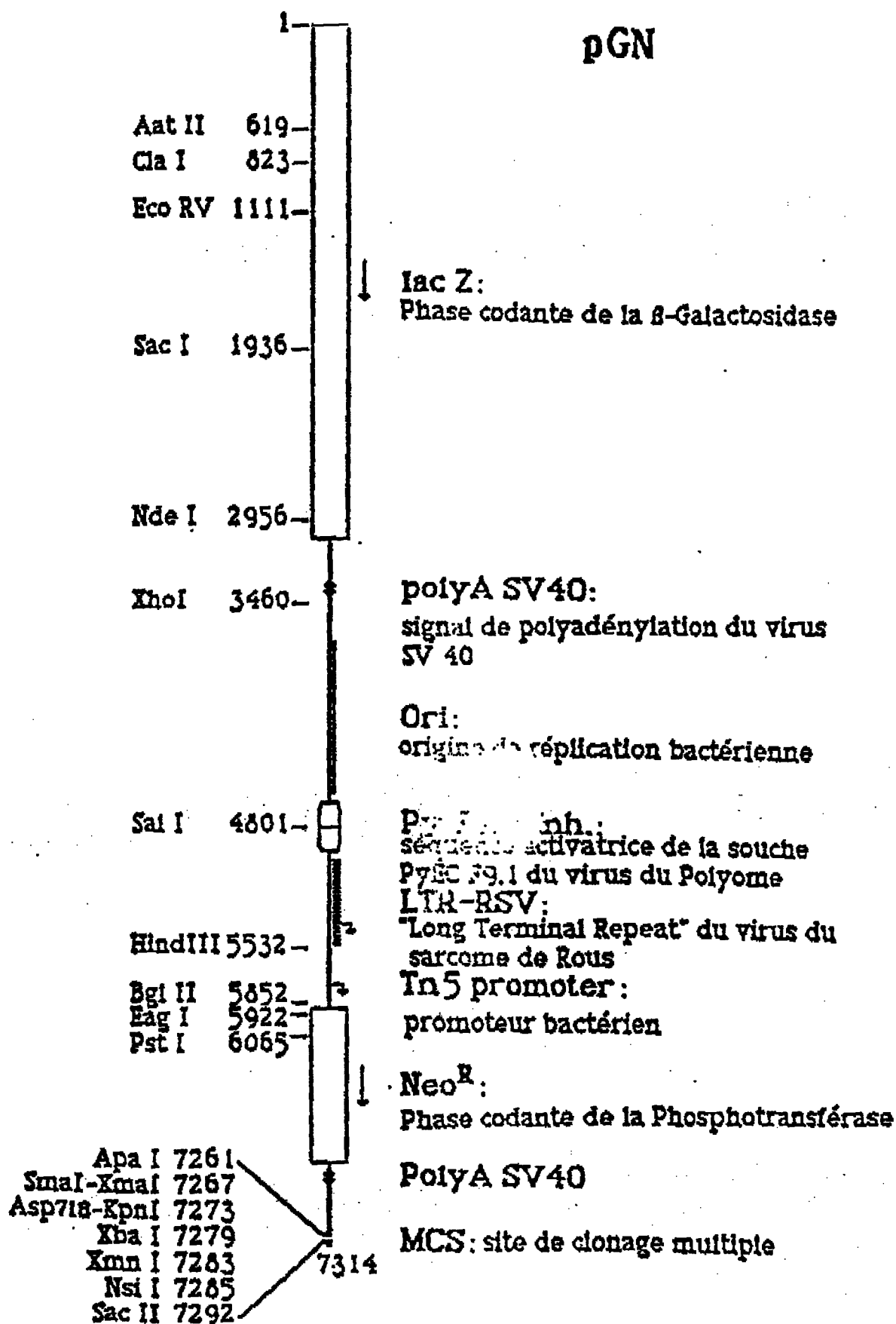


Figure 1: plasmide linéaire.

2/12

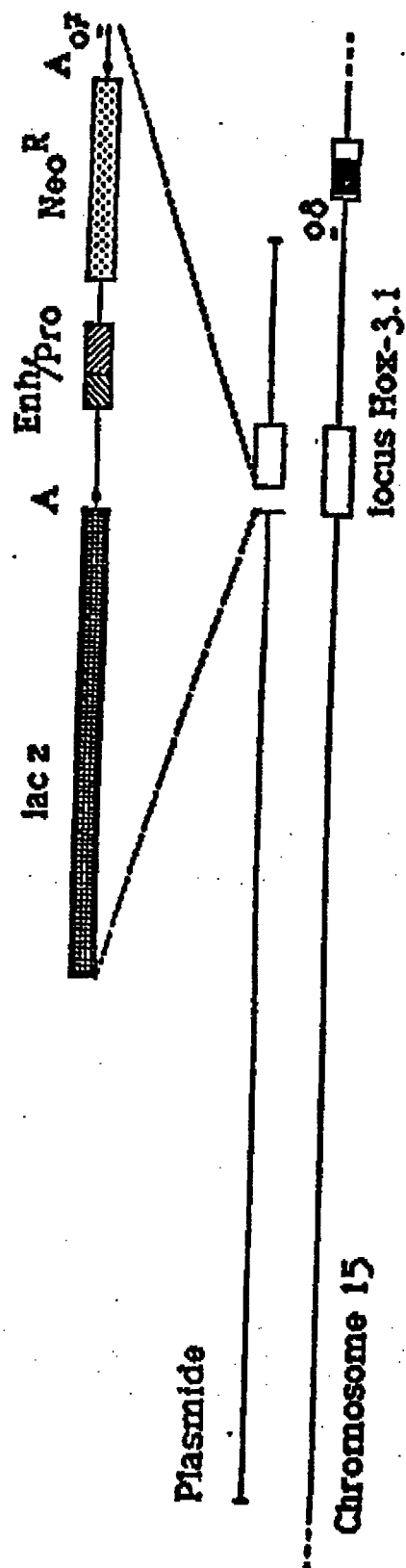


FIGURE 2a

3/12

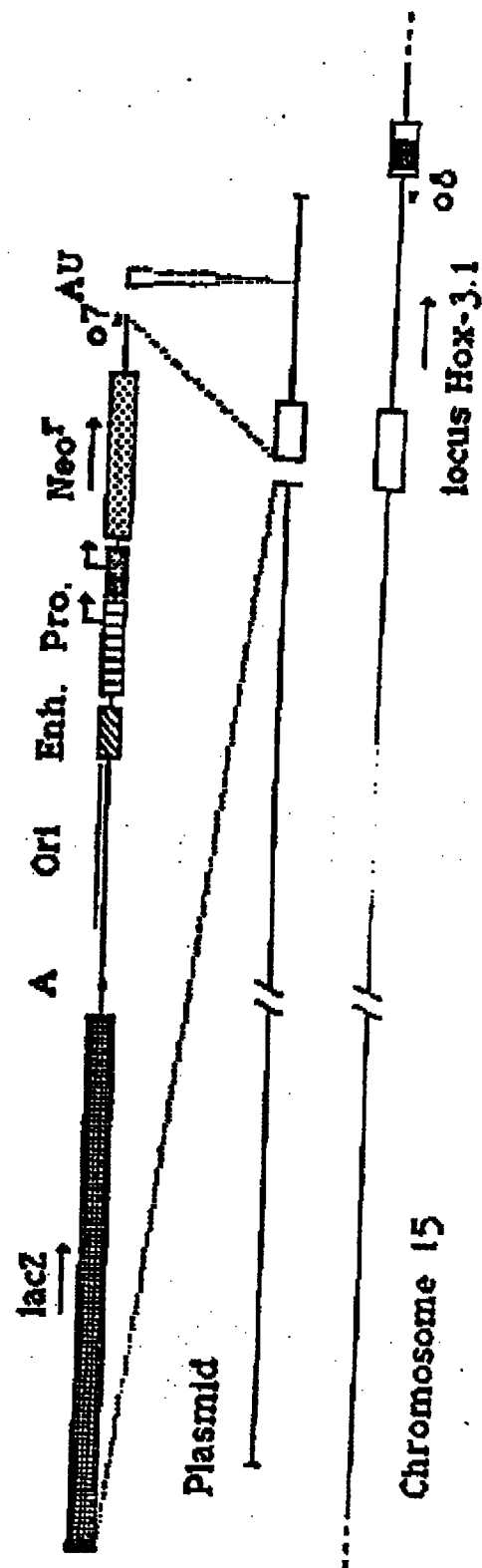


FIGURE 2b

4/12

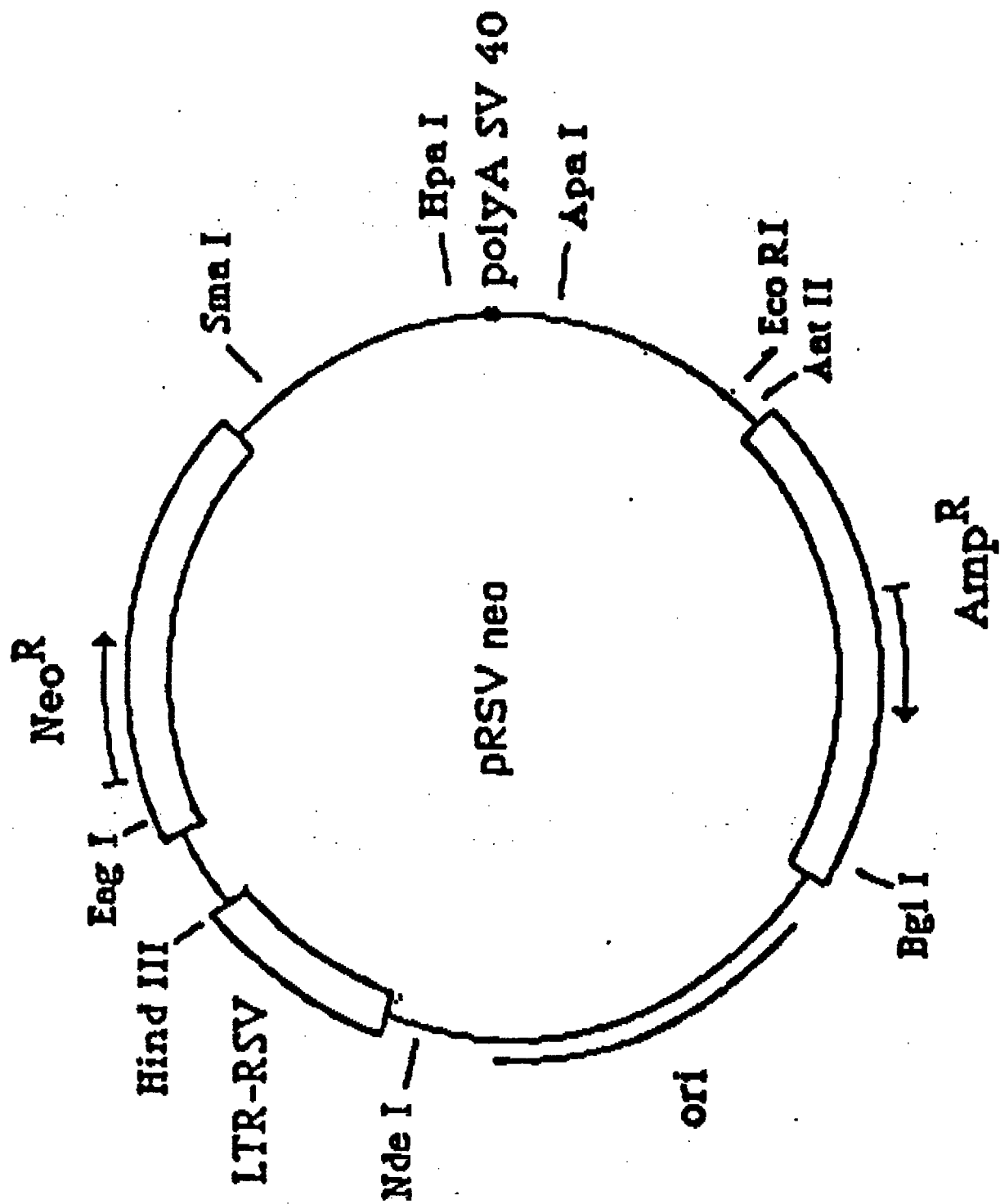
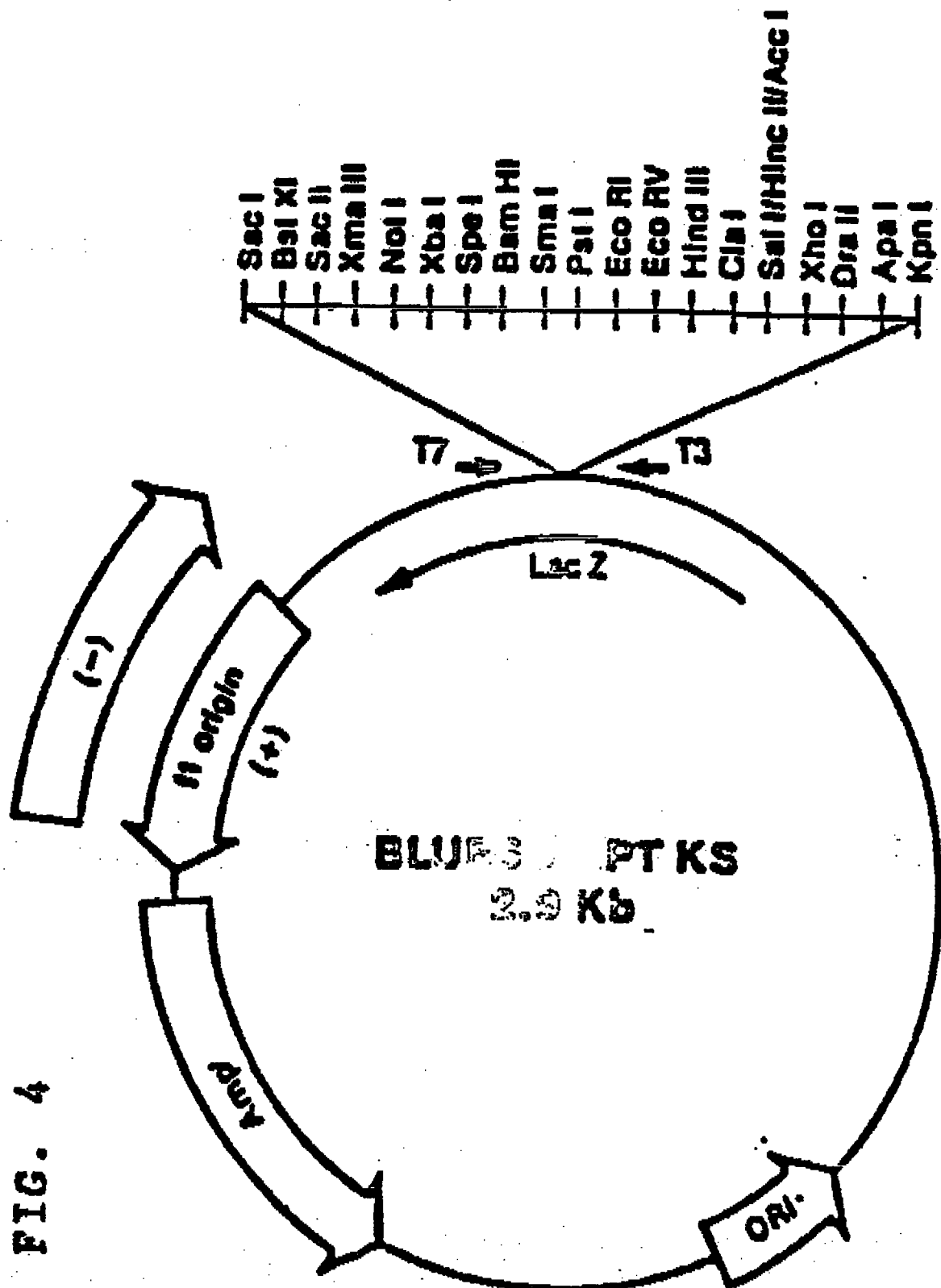


FIG. 3

5/12



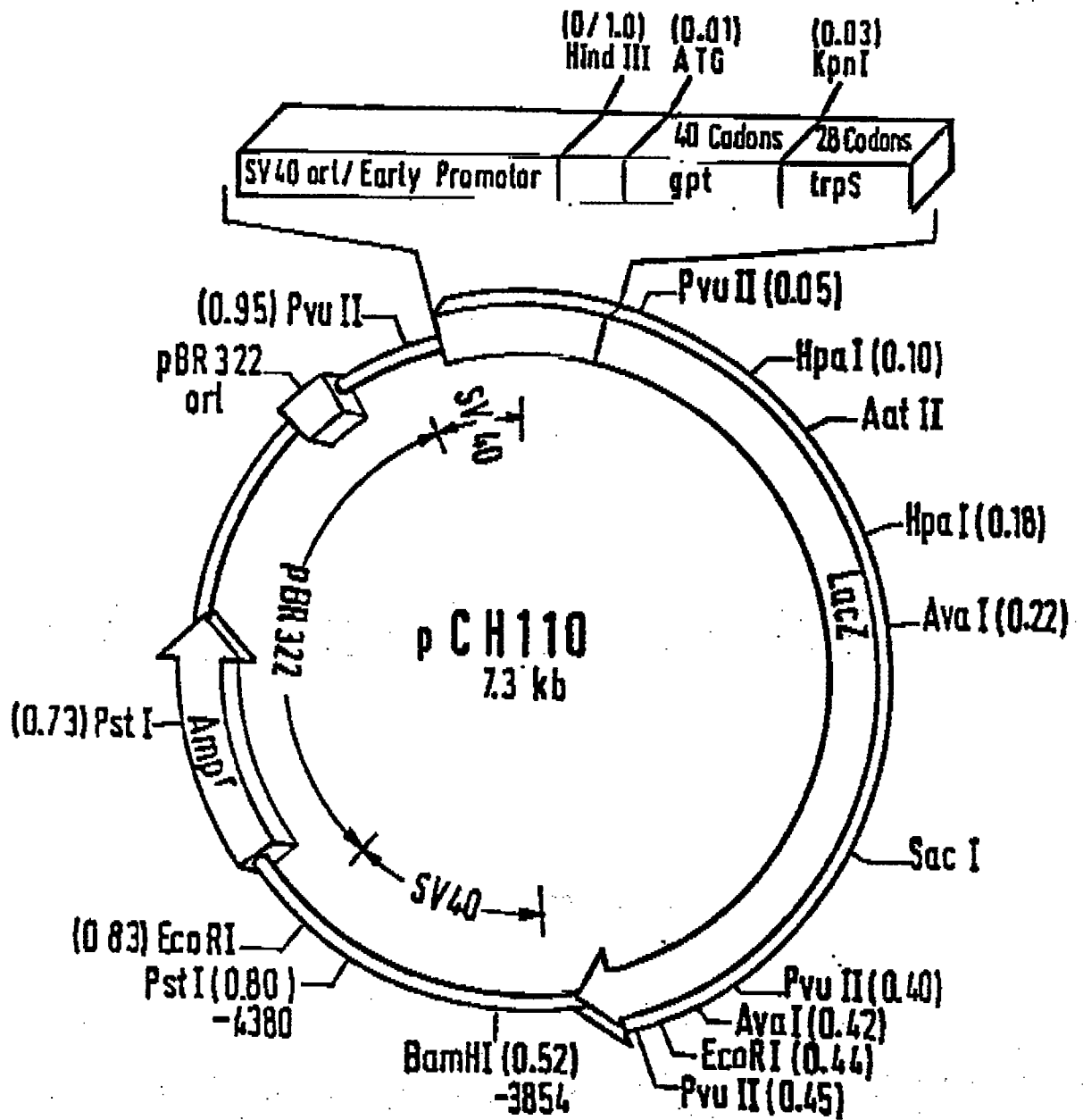
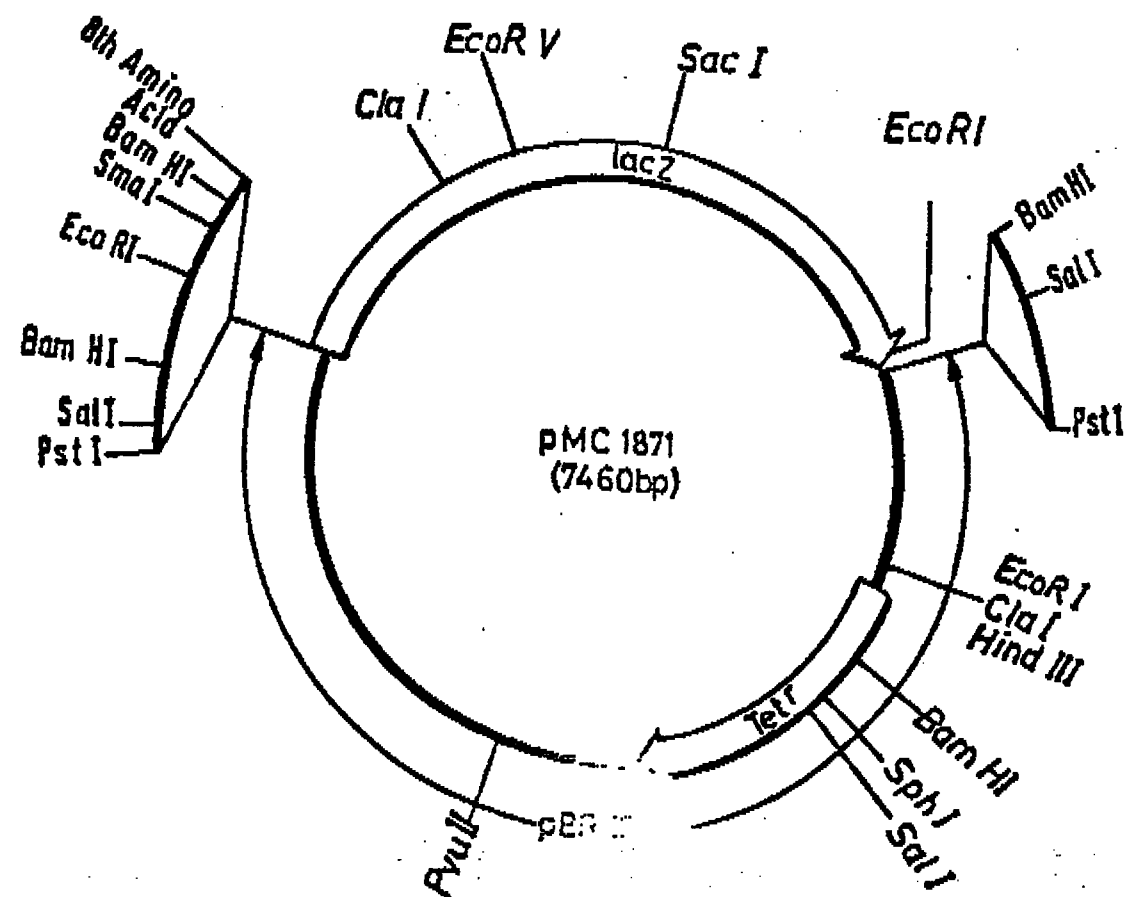


FIG. 5

7/12



pBR322 — CT GCA GGT CGA CCG ATC CCG GGAATT CCC

Pst I Sal I BamHI EcoRI Sma I

BamHI 8 9 1021

pro val lys stop stop stop

GGG GAT CCC GTC — lac Z — AAA TAA TAA TAA CCG GGC

AGG GGG GAT CCG TCG ACC TGC AG — pBR322

BamHI Sal I Pst I

FIG.6

8/12

1 2 3 4 5 6 7 8 9 10 M D



FIG. 7

9/12

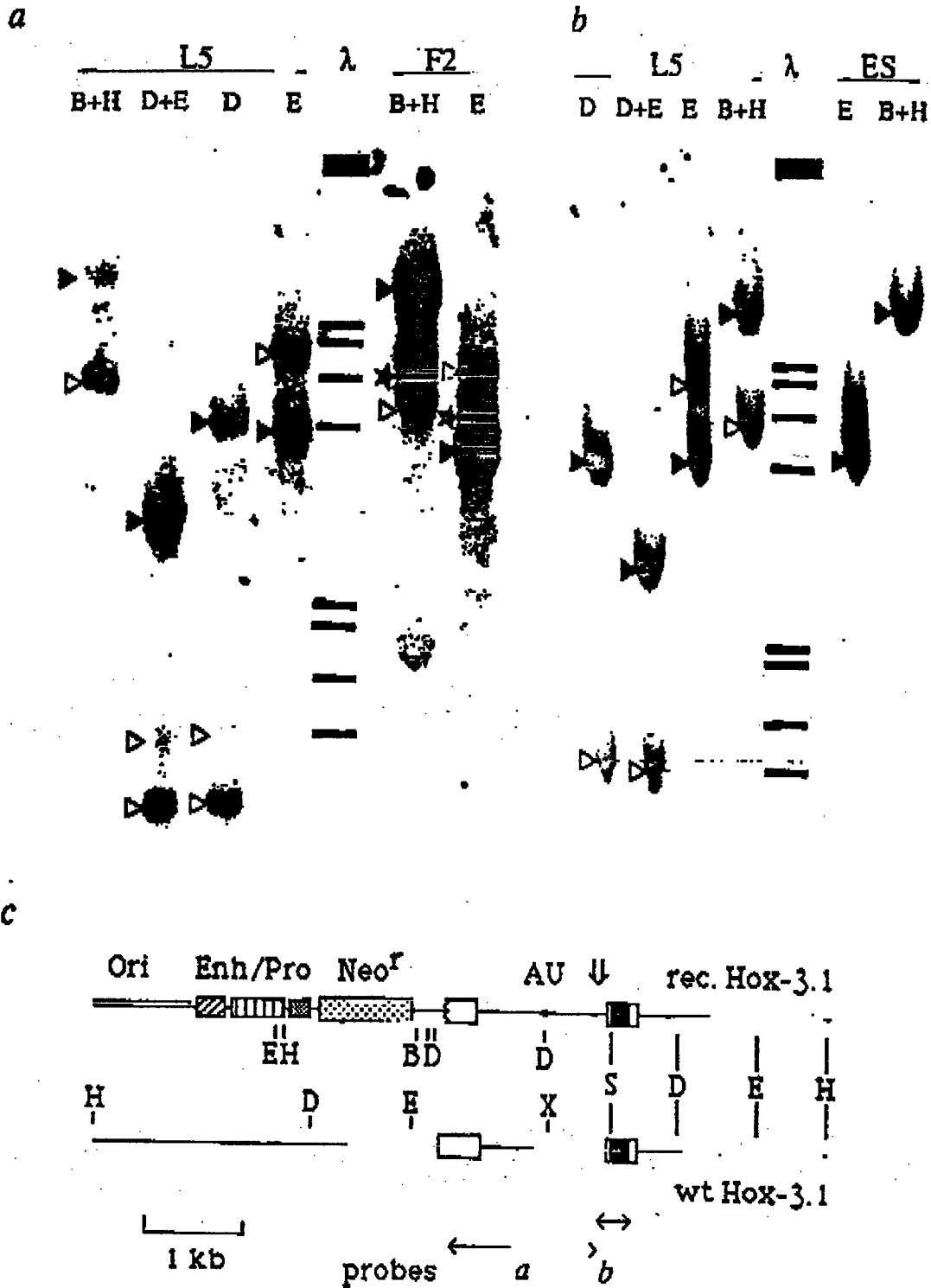


FIG. 8

10/12



Fig. 9A

11/12



Fig. 9B

12/12



Fig. 9C

THIS SHEET HAS BEEN
REPLACED

TRANSGENIC OR CHIMERIC ANIMAL AND PROCESS FOR SCREENING
PHARMACEUTICAL PRODUCTS

The present invention relates to a process for specific replacement of a copy of a gene present in the genome of a receiver eucaryotic organism by the integration of a gene different from the inactivated gene. The receiver gene will preferably be present in at least 2 examples in the transfected host cell. The receiver gene is defined as being the gene where the insertion of the different gene will be made.

More particularly, the invention relates to the production of transgenic animals in which the foreign gene has been introduced in a targeted manner to allow both the maintenance of the normal genetic functions of the animal and the expression of the foreign gene under the control of endogenous promoters.

By "different or foreign gene" is meant any nucleotide sequence corresponding to the totality or to a part of a "foreign or different" gene of the receptor gene as it is found normally in the genome (RNA or DNA), where it corresponds also to an artificially modified sequence of the normal gene or again to a fragment of this sequence.

The invention also concerns the process of production of these transgenic animals.

In the production of transgenic animals, the conventional methods used for the introduction of heterologous DNA sequences in the germ cell line do not allow control of the site of integration of the foreign gene in the genome, nor the number of copies thus introduced. The integration of the foreign gene is carried out at random and, in general, several copies of the gene combine at the same time, sometimes under the form of a head-to-tail tandem, the site of



THIS SHEET HAS BEEN
REPLACED

integration and the number of copies integrated varying from one transgenic animal to another.

It can thus happen that the endogenous cellular genes, situated at the point of insertion, may be inactivated, without this being easily detectable because of the numerous insertions made at random. If the product of these genes is important for the development of the animal, this development will be seriously disturbed. Moreover, the aleatory insertion of the foreign gene may take place at a site which is not appropriate for the expression of the gene. In addition, the fact that there is a variation in the site and in the number of insertions from animal to animal makes interpretation of studies of expression extremely difficult.

A major problem encountered in the production of transgenic animals is obtaining the expression of the foreign gene. Generally speaking, two types of experiment have been made on mice.

The genes introduced in the germ line are:

- either "complete" genes, including coding sequences flanked by their own regulatory sequences;
- or composite genes, formed from the coding sequence of a gene fused with the promoter sequence of another gene, the two fragments sometimes even belonging to two different animal species.

It was thus possible to confirm that the specificity of the expression of genes in such and such a tissue is determined by their regulatory sequence(s).

The choice of the appropriate promoter for the expression of the foreign gene in the transgenic animal is thus of primordial importance.

Besides, the controlled mutagenesis of murine genes in embryonic cell lines has recently been carried out using a technique of gene targeting (Thomas et al., 1987; Thompson et al., 1989).



THIS SHEET HAS BEEN
REPLACED

In the first case, the HPRT murine gene was mutated by insertion and replacement and, in the second case, a mutated HPRT gene was corrected. Thompson et al. widened their experiments as far as obtaining
5 chimeric mice and recorded the passage of the genetic modification into the germ cell line.

In each of the documents quoted, the precise site of integration was targeted by homologous recombination between, on the one hand, exogenous sequences carrying
10 the mutation or correction included in the vector, under the control of an exogenous promoter and, on the other hand, their genomic homologue. This being so, it must be pointed out that earlier authors have carried out their experiments on a specific gene (HPRT) whose
15 activation by mutation was accompanied by a detectable phenotype. The targeted mutation described by Thomas et al. had the effect of inactivating the HPRT gene and, consequently, making the detectable phenotype normally associated with the HPRT disappear. The Neo^R selection
20 gene, under the control of a TK promoter, was thus incorporated into the insertion DNA in order to permit the selection of transformants. It should be noted that the experiments described in the prior art implied a selection either by the receiver gene (e.g. HPRT) or by
25 the insertion gene (e.g. Neo^R). The site of insertion and/or the type of gene inserted is thus limited to genes conferring a selectable character, a direct selection.

Besides, in the prior art, the exogenous
30 sequences on the vector thus serve both to target the site of integration and to introduce the modification. Following homologous recombination, the modified gene always finds itself in its normal genetic environment.

We should remember that a problem posed during
35 the production of transgenic animals is the danger of

This Sheet Has Been Replaced



inactivating an endogenous cellular gene existing at the insertion point of the foreign gene.

According to the function of the product of the inactivated gene, such an inactivation can lead to
5 significant physiological or morphological disorders in the transgenic animal, or could even prevent its survival.

On the other hand, the inactivation of the gene could be considered as advantageous if the gene in
10 question coded for a virus receptor or other infectious agent.

The inventors have studied the possibility of avoiding the disadvantages described above and associated, in certain cases, to the possible
15 inactivation of one or several endogenous cell genes with an important function during the production of transgenic animals.

The object of the invention is a process for replacement or specific insertion of a gene, in
20 particularly by targeting a DNA, called insertion DNA constituted by the totality of a gene, or by a part of a gene, susceptible to being rendered functional, where this functioning can be made more efficient, when it is recombined with a complementary DNA to then provide a
25 complete recombinant gene in the genome of a eucaryotic cell, characterized in that

- the insertion site is in a chosen gene, called a receiver gene, and containing the complementary DNA, and in that

- 30 - eucaryotic cells are transfected with a vector containing an insert which comprises the insertion DNA and two so-called "flanking" sequences on either side of the insertion DNA, respectively homologous to two genomic sequences which are immediately adjacent to the
35 desired site of insertion in the receiver gene,



THIS SHEET HAS BEEN
REPLACED

5

- the insertion DNA being heterologous vis-à-vis the receiver gene, and

- the flanking sequences, being chosen from those which make up the above-mentioned complementary DNA and which allow, through homologous recombination with corresponding sequences of the receiver gene, the reconstitution of a complete recombinant gene in the genome of the eucaryotic cell.

According to the invention, the so-called insertion DNA can contain either a coding sequence or a regulatory sequence, the process then being characterized in that

- the insertion site is in a chosen gene, called a receiver gene, and in that

- eucaryotic cells are transfected with a vector containing an insert which comprises the insertion DNA and two so-called "flanking" sequences on either side of the insertion DNA, respectively homologous to two genomic sequences which are immediately adjacent to the desired site of insertion in the receiver gene,

- the insertion DNA being heterologous with respect to the receiver gene, and

- the flanking sequences being chosen so as to permit by homologous recombination depending on the case, either the expression of the coding sequence of the entire insertion DNA under the control of the regulatory sequences of the receiver gene, or the expression of a coding sequence of the receiver gene under the control of regulatory sequences of the insertion DNA.

For example, the insertion DNA may contain a coding sequence without the regulation element, notably without its own promoter.

The invention also relates to a process for the production of transgenic animals, characterized in that E.S. cells are transfected in the conditions described

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

6

above and selected for the occurrence of homologous recombination, that is the correct integration of the foreign gene, the transfected cells are injected into embryos at a stage where they are able to integrate the transfected cells (for example, at the blastocyst stage), and these are then reimplanted in a surrogate mother and the chimeric individuals obtained at the end of gestation, and in which is noted the colonization through E.S. cells of the germ line of the chimeric animal, are mated. If the E.S. cells have colonized the germinal line of the chimeric animal, heterozygous transgenic animals for the replaced gene will be obtained by mating (F1) the progeny.

It is also possible to insert insertion DNA, carried by the vector of the invention, in the egg shortly after fertilization (that is less than 24 hours). In this way, the insertion takes place while the egg is still in the unicellular state.

The invention also concerns a plasmid able to effect the targeted insertion of a recombinant gene called the insertion gene in the genome of a eucaryotic cell, characterized in that it contains an insert which itself comprises the insertion gene and two so-called "flanking" sequences on either side of the insertion gene, respectively homologous to the two genomic sequences which are immediately adjacent to the desired site of insertion in the receiver gene.

The insertion DNA can be a gene, or a part of a gene, heterologous to the transfected species.

The invention also concerns transgenic animals in which at least one endogenous gene, which is preferably present in the genome in at least two copies, has been inactivated by the insertion of a gene which is different from the inactivated gene, the insertion gene being inserted in a position which allows the

This Sheet Has Been Replaced



expression of this gene under the control of regulatory sequences of the inactivated endogenous gene.

Thus, the process of the invention allows, thanks to the phenomenon of homologous recombination, the insertion in a targeted manner of foreign genes, in particular coded sequences without the promoter with which they are usually associated, in the genome of a eucaryotic organism at a site which allows their expression under the control of the endogenous promoter of the gene where the insertion is made, and consequently, the inactivation of the targeted endogenous gene.

In a preferred embodiment of the invention, the targeted receiver gene is a gene which is present in the genome in at least two copies. Utilization of the electroporation technique (ref. 11) ensures the introduction of only one copy of the foreign gene.

In this embodiment of the invention, the targeted insertion of the gene of interest (i.e. the so-called insertion gene) has the effect of inactivating the single copy of the endogenous cellular gene where the insertion takes place and leaves intact and operational the other copy (copies) of this gene.

In this way, the genetic functioning of the transgenic animal is not or is little disturbed by the introduction of the foreign gene, even if the insertion inactivates a single copy of a receiving gene essential for the development of the animal. Either its development is not affected by the insertion of the foreign gene, or the possible minor disturbances in the case of inactivation of a critical gene will probably not be lethal for the animal. The effects of the insertion of the foreign gene in the homozygous state could be of any kind and would be observed in the second generation (F2) after crossing heterozygous individuals (F1) among themselves.



THIS SHEET HAS BEEN
REPLACED

8

If, on the other hand, the inactivation of all the copies of a gene is desired, for example in the case the gene codes for a receptor of an infectious agent, multiple copies of the foreign gene are introduced. Control of the quantity introduced can be ensured by using known techniques.

The targeted insertion of the foreign gene thus allows its introduction in a site where its expression is under the control of regulatory sequences of the endogenous gene where the insertion takes place.

The process of the invention also allows insertion of the foreign gene behind an endogenous promoter which has the desired functions (for example, specificity of expression in a particular tissue) and this, in the relevant case, without inactivating the other copies of the receiver gene.

In a preferred embodiment, the insertion DNA, or the vector, contains sequences inserted between the insertion DNA and the flanking sequences. These sequences can for example contain a sequence coding for a selective agent allowing the selection of transformants and eventually a marker gene, for example LacZ.

In a highly preferred embodiment of the invention, the insertion DNA carries, between the flanking sequences, on the one hand a DNA sequence intended to be recombined with the complementary DNA in the receiver gene to provide a recombinant gene and, on the other hand, a sequence coding for a selective agent permitting the selection of transformants and a promoter authorizing the expression of the selective agent, the receiver gene and the recombinant gene coding for expression products not conferring a selectable phenotype. According to this embodiment, the insert of the plasmid of the invention comprises, between the flanking sequences, on the one hand a DNA

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

9

sequence intended to be recombined with the complementary DNA in the receiver gene and, on the other hand, a sequence coding for a selective agent allowing the selection of transformants and a promoter
5 authorizing the expression of the selective agent, the DNA sequence intended to be recombined with the complementary DNA being other than a gene coding for a selective agent.

In this way, the selection of transformants is
10 entirely independent from the nature of the receiver gene and the inserted gene, contrary to the processes described until now, in which the inserted gene or the receiver gene necessarily had to code for a product of expression allowing the selection of transformants. The
15 system developed by the inventors gives total flexibility regarding the nature of the receiver gene and the inserted gene, or the gene formed by homologous recombination. The inventors have noted that, surprisingly, the insertion of sequences of a
20 significant size (for example about 7.5 kb) does not affect the frequency of homologous recombination.

The possible effect of the insertion of the DNA sequence according to this aspect of the invention includes, depending on the type of sequence inserted
25 for example the replacement of a coding sequence, the replacement of a regulatory sequence, the inactivation or reactivation of a gene by mutation or improvement of the expression level of a gene. It is possible, according to the invention, to replace a coding frame
30 or part of a coding frame by a heterologous sequence which begins at the initiation codon of the replaced gene so that the expression of the inserted gene entirely replaces the expression of the replaced gene. This avoids the formation of fusion proteins which
35 could be undesirable in a transgenic animal.

This Sheet Has Been Replaced



In this embodiment of the invention, the insertion DNA can carry between the flanking sequences a heterologous coding sequence without promoter, the coding sequence being other than a gene coding for a selection agent. The insertion DNA can also carry, downstream of the coding sequence and still between the flanking sequences, a gene coding for a selection agent, associated with a promoter allowing its expression in the target cell.

In this way, the heterologous coding sequence can be inserted after an endogenous promoter which has the desired properties, for example a certain specificity of expression, or transcription pattern etc., the selectability of the transformed cells being entirely independent from the expression of the heterologous coding sequence. This type of construction makes it possible, for example, to select the transformants even if the gene replaced by the heterologous coding sequence is not normally expressed in the target cells. This is particularly important in the production of transgenic animals from E.S. cells ("Embryonic Stem Cells") since a significant proportion of genes remains inactive until a more advanced stage of the development of the animal. The HOXC8 gene (also known under the name of Hox-3.1) is an example of this type of gene. Moreover, if the coding sequence codes for an easily detectable protein, for example β -Gal, the development of the transcription template of the replaced endogenous gene can be followed. The pGN vector is an example of this type of construction.

In another embodiment of the invention, the insertion DNA can carry a foreign regulatory sequence. The site of insertion and, consequently, the flanking sequences are chosen in relation with the desired aim, that is either the insertion of the foreign regulatory



sequence to give an effect of "double promoter" with the endogenous regulatory sequence, or the replacement of an endogenous promoter by the foreign promoter. The coding sequence which comes under the control of the regulatory sequence can be endogenous.

Another possibility can be the targeted insertion of a foreign DNA which carries both a regulatory sequence and a coding sequence. It is possible that the regulatory sequence can be that which is naturally associated with the coding sequence.

The process of the invention activates a vector containing two "flanking" sequences on either side of the foreign gene. These flanking sequences have at least 150 base pairs and are preferably inferior to the length of the receiver gene. It is essential that these two flanking sequences be homologous to the two genomic sequences which are immediately adjacent to the desired insertion site. The flanking sequence of the vector which is located upstream of the foreign gene to be introduced is normally homologous to the genomic sequence which is located at the 5' end of the insertion site. In the same way, the flanking sequence of the vector located downstream of the foreign gene is normally homologous to the genomic sequence which is located at the 3' end of the site of insertion.

It is possible to introduce "intercalated" sequences between either one of the flanking sequences and the foreign gene, for example sequences permitting the selection of transformants, markers, sequences permitting cloning of the vector etc.

The position of these intercalated sequences with respect to the foreign gene must however be chosen so as not to hinder the expression of the foreign gene, in particular the coding sequence of the foreign DNA under the control of the endogenous promoter or, inversely, the coding sequence of the endogenous DNA under the



control of foreign elements of regulation carried by the insertion sequence.

Despite the presence of flanking sequences, which encourages homologous recombination, it is possible that a certain number of integrations take place at random. In order to verify that the targeted insertion has actually taken place in the targeted site and not in another place, the technique of "Polymerase Chain Reaction" (P.C.R.) (see Ref. 10) is used to amplify the DNA sequence of the locus where the insertion takes place, or should take place. In this way, only the clones transformed following a homologous recombination are selected.

The flanking sequences of the vector are evidently chosen in relation to the desired site of insertion so that homologous recombination can take place. Thus the flanking sequences can carry duplicated sequences of the endogenous promoter and/or modifications to the sequences which precede the initiation codon to raise the translation level (upstream sequences) and duplicated sequence of the termination sequences particularly of the sites of polyadenylation (downstream sequences).

The insertion gene can be any gene of interest. One will cite as non-limiting examples the gene LacZ (as in the model described below), the genes coding for interleukin or interferon, receptor genes for example of retinoic acid or beta-3 adrenergic or H.I.V., and genes known to be linked with certain diseases, for example myopathy etc.

The eucaryotic cells transformed by the process of the invention are also included in the present application.

In a preferred embodiment of the invention, the eucaryotic cells are embryonic cell lines (see Ref. 14 and 15).



In fact, a mutated E.S. cell can be injected into an early embryo which, after reimplantation, can be born under a chimeric form. If the germ line is colonized by the mutated cell, the chimeric animal will transmit the mutation to its progeny. After this, it will be possible to observe the effects of this mutation, in the homozygous state in some individuals, on their development, their behaviour, their metabolism, their pathology etc.

10 Figure 1 shows the pGN plasmid, which is part of the invention.

Figures 2a and 2b show the molecules pGMA and pGMD respectively constructed from the pGN plasmid in relation to the gene HOXC8 (Hox-3.1). These plasmids are mutagenesis plasmids. The two parts of the coding sequence of the gene HOXC8 (Hox-3.1) are represented, on chromosome 15, with the "homeo" box in black. The corresponding sequences of HOXC8 (Hox-3.1) have been cloned in the pGN plasmid.

20 (A: signal of polyadenylation; Enh/Pro: enhancer-promoter).

07 and 08 show the two oligonucleotides used in the PCR.

25 Figures 3 to 6 show the plasmids used in the construction of the pGN.

Figure 7 illustrates the detection of homologous recombination by the Polymerase Chain Reaction (P.C.R.) technique on the transfected E.S. cells.

30 Figures 8a and 8b show Southern analyses of positive individual clones (L5 and F2) and E.S. cells (C.C.E.).

The process of the invention has a very wide industrial applicability and can vary according to the nature of the foreign gene introduced, for example for the production of transgenic animals.



The genetics of mammals will progress considerably in view of the recent possibility of specifically mutating any gene, thus making it possible to define its role more accurately. Through this technology which involves homologous recombination and E.S. cells, valuable information will be obtained on oncogenes, growth factors, transcription factors etc., genes which concern very actual subjects both in fundamental research and in applied research. An important outcome for medical research is the possibility of reproducing a human disease whose genetic determination is known (certain pathological human diseases, such as the myopathy of Duchesne) so as to study its mechanisms better and search for a therapy.

By applying the process of the invention, a gene known to be responsible for a certain disease is inserted in a targeted manner into the genome of an E.S. cell. The transgenic animal which results from this process presents a useful model of the disease.

If necessary, and as described above, normal genetic functions can be significantly maintained, despite the insertion of the foreign gene.

Another application of the process of the invention is the genetic marking of animals. This application consists in inserting an insertion gene which is easily detectable e.g. the LacZ gene which can then play the role of a cell marker. In this way, filiation studies e.g. in competition animals are facilitated, and the breed can be followed up.

Insertion of the LacZ gene as insertion gene also makes promoter studies possible. In view of the possibility of identifying the activity of β -galactosidase, the activity and specificity of different endogenous promoters can be studied by



targeting different sites in the same type or different types of cells. The same studies could be carried out on an entire organism, during development or, at the adult stage, by using the techniques of chimeric or
5 transgenic animals.

The inventors have observed in a surprising manner that the frequency of homologous recombination is not affected by the insertion of fragments of significant size, for example LacZ. This observation
10 has suggested to the inventors that the technique of homologous recombination could be well adapted to the insertion of other heterologous genes which are of significant size.

The invention is also aimed at the application of
15 the process of replacement or specific insertion for genic therapy.

In view of the possibility of modifying the genome of an animal, the process of the invention can thus also be used as "gene therapy". The most obvious
20 utilizations would consist in inactivating the genes of receptors of toxic or infectious agents (viruses or bacteria). If this mutagenesis proved to be lethal, it would be necessary to re-establish the lost function without re-establishing the sensitivity to the harmful
25 agents. A modified gene coding for such a receptor could be reintroduced into the mutated cell, as long as the modification can be provoked by homologous recombination. This modification of the genetic character would confer on the animal an immunity
30 against the disease under consideration.

This protocol can also intervene in the framework of auto-grafting. Cells, sick or healthy, taken from a patient, could be treated and immunized, then reimplanted back into the same individual.

35 The technique of the invention is also suitable for studies of the activity of pharmaceutical products



presumed to have an activity with regard to products of expression of a pathological gene linked to a disease. In this case, the insertion gene is constituted by the pathological gene or a fragment of it, and one
5 administers to the transgenic animal the pharmaceutical product to be tested with a view to evaluating its action on the disease.

The invention will be illustrated by referring to the pGN plasmid and its utilization in the targeted
10 insertion of a foreign gene (LacZ, coding for the enzyme β -galactosidase of E. Coli) in the genome of a mouse E.S. cell. The gene LacZ was chosen because of the fact that its expression can easily be detected and is used simply to demonstrate the feasibility of the
15 technique.

The coding frame of the enzyme β -galactosidase of E. Coli (LacZ; 1-3057), fused with a genomic sequence (7292-3) of the murine gene HOXC8 (Hox.3-1) (Ref. 1) begins with the initiation codon of this gene. In fact,
20 the sequence which precedes the initiation codon of HOXC8 (Hox.-3.1) is identical to the consensus sequence observed in vertebrates (Ref.2), thus allowing a better translation rate of β -galactosidase into the cells of vertebrates. The gene LacZ is followed by a signal of
25 polyadenylation for example of the virus SV 40, like most of the eucaryotic genes, so as to stabilize the messenger RNAs.

The activity of β -galactosidase of E. Coli, which is functional in the eucaryotic cells, can be detected
30 in different manners. Cells expressing the LacZ gene take on a blue colouring, after fixation, in the presence of X-Gal, which is a substrate of β -galactosidase (Ref. 3). A new substrate, FDG (fluorescein di- β -galactopyranoside), makes it possible
35 to detect and to dose the β -gal activity, while still



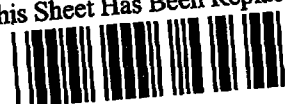
keeping the cells alive (Ref. 4). The cells expressing LacZ accumulate a fluorescent product and can be isolated with the aid of a cell sorter or FACS (fluorescence-activated cell sorter).

5 The transcription unit of the gene with resistance to neomycin comes, to a large extent, from the plasmid pRSV neo (Ref. 5). The LTR (long terminal repeat) of the virus of Rous-sarcoma provides very powerful activating and promoting sequences in numerous
10 eucaryotic cells (Ref. 6). From the bacterial transposon Tn5 comes a promoter active in E. Coli and the coding phase of the enzyme phosphotransferase (Ref. 7), which is followed by the polyadenylation signal of the virus SV 40. The same gene, under the double
15 control of promoters RSV and Tn5 can confer resistance to neomycin or kanamycin to bacteria and resistance to G418 to eucaryotic cells.

Through the effect of a simple single mutation, unit B of the activating (enhancer) sequences of the
20 strain PyEC F9.1 of the Polyoma virus has become much more active in different types of cells, and in particular in the cells of embryonic carcinoma (EC) (Ref.8). Two copies of this enhancer Py F9.1 have been inserted in tandem in the pGN plasmid, upstream of the
25 LTR-RSV, and in the "late promoter" orientation of the regulatory region of the Polyoma.

In order to improve the translation rate of phosphotransferase, the sequence preceding the initiation codon was modified during a mutagenesis by
30 oligonucleotide. Thus the sequence T T C G C A U G became G C A C C A U G, corresponding much better to the consensus sequence of the initiation of translation in vertebrates (Ref.2).

The improvements made to the transcription unit
35 of the gene of resistance to neomycin could be estimated by transfecting embryonic cell strains (ES)



of mice. With equal molarity in plasmid, a construction with the enhancer Py F9.1 produced 7.5 times more clones resistant to G418 than pRSV neo and 2 to 3 times more than the pMC1 neo described by Capecchi et al. (Ref. 13). Again, the number of clones was raised by 60x, that is 450x compared with pRSV neo, by modifying the sequence of initiation of translation. Homologous recombination may be a fairly rare event, according to the experimental conditions applied (e.g. 1/1000 for HPRT, Ref. 13). A vector presenting a high selection efficiency is thus very useful, especially since the conditions of electroporation result mainly in the integration of a single copy.

Moreover, the pGN plasmid contains an origin of bacterial replication of the type colE1, pBR322, which allows cloning and preparations in E. Coli.

Finally, a multiple cloning site (M.C.S.), synthesized in vitro, which only contains unique restriction sites in pGN, was inserted upstream of LacZ, so as to facilitate the utilizations of this plasmid.

The plasmid "flanking" sequences which provoke homologous recombination are added to the extremities of the pGN plasmid after linearization of the plasmid upstream from LacZ, by an MCS site (see Fig. 2). In this case, the flanking sequences chosen are homologous to the chromosome sequences taken from HOXC8 (Hox-3.1) which intervene later in the homologous recombination.

Figure 2 provides the location of the molecule built from the pGN plasmid with respect to the gene HOXC8 (Hox-3.1). In this case, recombination between the plasmid and chromosomal sequences of HOXC8 (Hox-3.1) would result in an insertion at the beginning of the coding phase of this gene, thus in its total inactivation.



The pGN plasmid comprises several advantages for this methodology, which is applicable to any gene. As the event of homologous recombination can be fairly rare (of the order of 1 per 1000 non-homologous integrations), it is necessary to be able to analyze a large number of clones whose resistance to G418 is sufficiently strong to be expressed in any part of the genome. The modifications made to the transcription unit of phosphotransferase answer these problems perfectly. The method of mutagenesis by homologous recombination is equivalent to inactivating a gene by an insertion or a substitution, but the pGN plasmid presents the supplementary advantage of being able to substitute the expression of β -galactosidase to that of the mutated gene. Finally, the MCS facilitates the cloning of genetic fragments.

EXAMPLES:

I - Construction of the pGN plasmid

The intermediate plasmids are numbered according to their stage.

1st stage:

Insertion of a Xho I site in the Bgl I site of pRSV neo

Insertion of a Xho I linker in the Bgl I site of pRSV neo, filled by the Klenow fragment of the DNA polymerase of E. Coli.

2nd stage:

Insertion of a Cla I site in the Nde I site of the p1 plasmid

Insertion of a Cla I linker in the Nde I site of p1, filled by Klenow polymerase.

3rd stage:

Insertion of Py F9.1 enhancer in the Cla I site of the p2 plasmid



Insertion of enhancer Py F9.1 Pvu II-Pvu II isolated by a unique Acc I site, in the Cla I site of p2. Selection of a clone containing two enhancers oriented in the direction "late promoter".

5 4th stage:

Deletion Sma I-Hpa I of the p3 plasmid

The two enzymes, presenting "blunt end" extremities can be bonded directly. This deletion removes the intron of the t antigen of SV 40, which is
10 not very useful, and reduces appreciably the size of the transcription unit of phosphotransferase.

5th stage:

Insertion of a Xho I site in the Bam HI site of pCH110

15 Insertion of a Xho I linker in the Bam HI site of plasmid pCH 110 (Pharmacia), filled by the Klenow polymerase.

6th stage:

Insertion of 3' LacZ-polyA SV 40 in the p4
20 plasmid

The 3' end of the coding phase of β -galactosidase, followed by the signal of polyadenylation of the virus SV 40 is isolated from the p5 plasmid through sites Xho I-Aat II and cloned in the
25 p4 plasmid through the same sites.

7th stage:

Insertion of 5' LacZ in the KS-vector

The 5' end of the coding phase of β -galactosidase is isolated from plasmid PMC 1871 (Pharmacia) through
30 the sites Pst I-Sac I and cloned in the vector KS- (Stratagene) through the same sites.

8th stage:

Fusion of a genomic sequence HOXC8 (Hox-3.1) with
5' LacZ



A genomic sequence of the gene HOXC8 (Hox-3.1), cloned in the vector KS-, is purified by successive digesting with the enzyme Sac I, then with the Mung bean nuclease and finally with the enzyme Apa I. This
 5 insert is fused with the 5' end of the coding phase of β -galactosidase by cloning in the p7 plasmid digested by Apa I-Sma I. The protein thus fused contains the initiation codon of the translation of the gene HOXC8 (Hox-3.1) followed by the coding phase of β -
 10 galactosidase (then verified by sequencing).

Met Ser Ser	Ile Pro Gly Asp Pro
CCAGC ATG AGC TCC	ATT CCC GGG GAT CCC
GGTCG TAC TCG AGG	TAA GGG CCC CTA GGG

↓ Sac I

15 CCAGC ATG AGC T
GGTCG TAC

↓ Sma I

↓ Mung bean nuclease

CCAGC ATG	GGG GAT CCC
GGTCG TAC	CCC CTG GGG

20 Met Gly Asp Pro
CCAGC ATG GGG GAT CCC
GGTCG TAC CCC CTA GGG

9th stage:

25 Insertion of HOXC8 (Hox-3.1)-5' LacZ in the p6 plasmid

The fusion HOXC8 (Hox-3.1)-5' LacZ is isolated from the p8 plasmid through the sites Apa I-SacI and cloned in the p6 plasmid through the same sites. The
 30 effect of this cloning is to reconstitute the coding phase of β -galactosidase in its totality.

10th stage:

Insertion of the gene Neo^R in the vector KS+



The gene of resistance to neomycin (bacterial promoter and coding phase of phosphotransferase) is isolated in the pRSV neo through the sites Hind III-Eco RI and cloned in the vector KS+ (Stratagene).

5. 11th stage:

Mutagenesis of the initiation sequence of Neo^R in p10

The sequence of the initiation of translation of phosphotransferase is modified to be identical to the consensus sequence observed in vertebrates and thus to allow a higher level of initiation of translation and therefore an increased resistance to G418 for the cells of mammals. The modification also creates a Apa LI site which makes it possible to control the efficiency of the mutagenesis.



20 An oligonucleotide (CTTGTTCAATCATGGTGCACGATCCTCA) comprising a region of mispairing with the sequence of pRSV neo (underlined) is synthesized (Gene Assembler, Pharmacia) then phosphorylated by the polynucleotide kinase of the T4 bacteriophage. A single strand
25 template of the p10 plasmid is prepared with the fl origin of the KS+ plasmid and hybridized with the mutagenesis oligonucleotide. The second strand is synthesized and repaired by the Klenow polymerase and the DNA ligase of the T4 bacteriophage. After
30 transformation of bacteria, the mutated clones are screened with the aid of the oligonucleotide marked with ³²P. The mutagenesis was verified by digesting with Apa LI and also by sequencing.

12th stage:



Replacement of the initiation sequence in the p9 plasmid

A fragment containing the modified sequence of initiation of translation of the gene of resistance to
5 neomycin is isolated from the p11 plasmid through the enzymes Hind III-Eag I and cloned in the p9 plasmid through the same sites.

13th stage:

Insertion of the multiple cloning site in the p12
10 plasmid

Two complementary oligonucleotides are synthesized (Gene Assembler, Pharmacia) then phosphorylated. After pairing, the MCS is cloned in the sites Apa I-Sac II of the p12 plasmid through its
15 sticky ends.

XmaI Asp 718 XmnI
ApaI SmaI KpnI XbaI NsiI SacII
5' CCCCAGGGGTACCTCTAGAATGCATTCCGC 3'
20 3' CCGGGGGGCCCCCATGGAGATCTTACGTAAGG 5'

The multiple cloning site was also verified by sequencing.

II - Addition of flanking sequences at the
25 extremities of the linearized pgN plasmid upstream from lacZ' by an M.C.S. site

The flanking sequences used were chosen in relation to the desired site of insertion (for example, HOXC8 (Hox-3.1), see Fig. 2a and 2b pGMA and pGMD).

30 In the construction of the mutagenesis plasmid pGMD, two fragments of DNA homologous to the locus HOXC8 (Hox-3.1) were cloned at the sites Apa I-Nsi I and Nsi I-Sac II of the pGN vector. The 5' fragment begins at the site Sac II (CCGCGG) at nucleotide 219 of
35 the DNAC c21 of HOXC8 (Hox-3.1). This fragment extends over 6.8 kb in 5' up to the first BamHI site. The 3'



fragment begins at the site Apa 1 (GGGCCC) at nucleotide 885 of the cDNA c21. This fragment extends over 1.5 kb in 3' up to the first PstI site. A NsiI linker was inserted in the BamHI site of the 5' fragment and in the PstI site of the 3' fragment. The 5' and 3' fragments were cloned in the pGN vector in the sites Nsi I-Sac II and Apa I-Nsi I, respectively. The sequence of the cDNA of HOXC8 (Hox-3.1) c21 has been published (Ref. 1).

10 The mutagenesis plasmid is linearized by digesting with Nsi I before electroporation of E.S. cells. Its extremities are formed by the two genomic fragments cloned at the sites Apa I-Nsi I and Nsi I-Sac II of the pGN vector.

15 The pGMD plasmid does not present any signal of polyadenylation after the gene of resistance but, on the other hand, presents a sequence rich in AU responsible for selective degradation of mRNA, inserted in the sequence of the intron of HOXC8 (Hox-3.1) of the plasmid.

20 Another mutagenesis plasmid, pGMA, presents the same structure as pGMD but contains the signals of polyadenylation and of termination of transcription of SV40 and does not present the AU sequence of degradation of mRNA downstream from the gene Neo^R. The aim of these modifications was to reduce the level of transcripts of Neo^R in clones resulting from random integration. On the other hand, clones resulting from homologous recombination events between pGMD and a locus HOXC8 (Hox-3.1) should have an unaltered growth during the selection at G418, the AT sequence of degradation of mRNA being eliminated by the process of recombination itself, or spliced with the intron HOXC8 (Hox-3.1).



In the experimental stages which follow, the protocol described by Thompson et al. 1989, was followed for the production of chimeric animals.

III - Transfection of embryonic mouse cells

5 The method described by Thompson et al. 1989, was used to transfect embryonic mouse cells. Utilization of the technique of electroporation ensures the introduction of a single copy of the foreign gene (LacZ) per cell. After transfection, several clones
10 expressing β -galactosidase were isolated.

The mutagenesis plasmids pGMD and pGMA were linearized and introduced by electroporation into E.S. cells in order to favor the insertion of only one copy in the genome (Ref. 11).

15 The initial transfections were carried out to compare the targeting efficiency of the HOXC8 (Hox-3.1) of the plasmids pGMA and pGMD (see table I).

Homologous recombination in the gene HOXC8

20 (Hox-3.1)

Exp.	Plasmid	No. of group	No. clones forming the group	No. of positive PCR results
	of mutagenesis	analysed		
25 I	pGMA	3	600	0(2)
II	pGMD	5	250	3(5)
III	pGMD	84	2-3	5(5)

30 Table I

The E.S. cell line "C.C.E." (Ref. 16) was maintained in a continuous manner on fibroblast feeding layers (Ref. 17). For experiments I and II, 1.5×10^7 E.S. cells in 1.5 ml HeBS were electroporated (Ref. 11)
35 at 200 V, with 40mg of linearized plasmid, and then



spread over four culture trays (diameter 100 mm). For experiment III, the shock was carried out under the same conditions but a quarter of the cells were spread over four 24-well plates. The next day, $250 \mu\text{g} \cdot \text{ml}^{-1}$ of G418 were added. Each transfection produced about 2400 clones with pGMA and about 1000 clones with pGND.

The average number of clones of E.S. cells resistant to G418 in each group is indicated in table I, as well as the number of groups giving a positive result with the P.C.R. technique. A positive result means that a band of 1.6 kb could be observed on an agarose gel coloured with ethidium bromide (see Fig. 7). The number of groups giving a positive signal after a Southern analysis of the P.C.R. mixture and hybridization of a specific probe which does not contain the primer sequences is indicated between brackets (Fig. 8).

Detection of homologous recombination with P.C.R.

P.C.R. was carried out on 10^5 cells of a group of 250 clones of transfection II (sec lane D of Figure 7). In the other lanes, four groups of transfection III were analysed together by mixing about 4×5000 cells. The primers 07 and 08 used in the P.C.R. surround the sequence 3'HOXC8 (Hox-3.1) of the mutagenesis plasmid (Fig. 2). The fragment of 1.6 Kb covering this 3' sequence cannot be amplified except in the case of homologous recombination. Lanes 2, 3 and D illustrate these positive results.

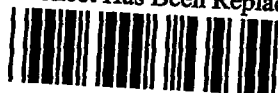
The DNA of E.S. clones was prepared at the moment of replica on filter by utilizing the method "boiling-proteinase K digestion boiling" (Ref. 18). 40 amplification cycles (40 seconds at 94°C , 1 minute at 60°C , 7 minutes at 72°C) were carried out in a reaction mixture of $100 \mu\text{l}$, containing 67 mM Tris-HCl (pH 8.6), 16.7 mM $(\text{NH}_4)_2\text{SO}_4$, 6.7 mM MgCl_2 , 10 mM 2-mercaptoethanol, 0.01% (p/v) gelatine, 200 μM dATP,



dTTP and dCTP, 100 μ M dGTP, 100 μ M 7-deaza dCT, 600 ng of each primer (07: AACTTCCCTCTCTGCTATTC and 08: CAGCAGAAACATACAAGCTG) and 3U polymerase Taq (Perkin Elmer Cetus), covered by 100 μ l paraffin. Half of the
5 reaction mixture was applied on an agarose gel 0.7% coloured with ethidium bromide. The main marker is an Eco RI + Hind III digest of lambda DNA.

Southern analyses

Three independent clones of E.S. cells containing
10 the mutated HOXC8 (Hox-3.1) (identified by P.C.R.) were isolated from positive groups using pipettes. Their DNA was examined by Southern analysis after digestion with the restriction enzymes indicated in Figure 8, so as to confirm the specific targeting and to make the
15 distinction between recombined and wild-type loci. Two different probes were used in the analysis of the 3' end of loci HOXC8 (Hox-3.1) in the mutated clones and in the non-mutated E.S. used as controls (Figure 8c). The first probe "a" was contained in the HOXC8 (Hox-
20 3.1) sequences of the mutagenesis plasmid and showed the number of vector integrations and their physical links. One of the three recombined clones also contained a copy of the plasmid integrated at random (Fig. 8a, clone F2). The second "b" probe which was
25 not in the mutagenesis vector made the distinction between the recombined and wild-type alleles HOXC8 (Hox-3.1) (Fig. 8b). The recombined locus HOXC8 (Hox-3.1) presented, with the two probes, the image of hybridization expected from the restriction maps of the
30 mutagenesis vector and of the intact locus. In addition, the existence of two fields of recombination in the 3' fragment of the vector was confirmed by the presence or the absence of the AT sequence in the recombined locus HOXC8 (Hox-3.1), (for example Fig. 8, clone L5). The 5' end of the locus HOXC8 (Hox-3.1) was
35 also analyzed for a homologous recombination event.



Restriction enzymes not presenting sites in the sequence HOXC8 (Hoc-3.1) 5' of 6.8 kb of the mutagenesis vector were utilized in the digestion of the DNAs of recombined clones. These DNAs were then submitted to electrophoresis in a pulsed field to differentiate the fragments of high molecular weight. A Southern analysis of this gel also indicated the correctly recombined alleles and the wild-type HOXC8 (Hoc-3.1) alleles, using a sensor presenting a sequence upstream from the mutagenesis plasmid.

Southern analyses have shown that an allele of the gene HOXC8 (Hoc-3.1) was recombined as predicted. The homologous recombination was equivalent to a double "crossing-over" between the genomic arms of the mutagenesis plasmid and the homologous chromosomal sequences (Fig. 2).

In the recombinant clones, the gene LacZ was placed under the control of promoter and regulatory sequences of HOXC8 (Hox-3.1) upstream from the AUG codon, but the 3' signals of maturation of mRNA came from the SV 40. In these recombined clones, the expression of LacZ was not detectable by colouring with β -Gal, which is coherent with the absence of transcription of HOXC8 (Hoc-3.1) in E.S. cells determined by analysis of RNase protection. The activity of β -Gal could be induced in certain cells after 3 or 4 days of culture in the presence of $5 \cdot 10^{-7}M$ retinoic acid, conditions known for inducing the transcription of HOXC8 (Hoc-3.1) (Ref. 19).

By using the mutagenesis vector pGMA, which presents a total homology of 8.3 Kb DNA with the locus HOXC8 (Hoc-3.1), a fragment of 120 pb was replaced by an insertion of 7.2 Kb. The frequency of this targeted replacement (1/900) is comparable to that obtained recently (1/1000) with HPRT (Ref. 13) or with En-2



(1/260) (Ref. 20), the heterologous fragment inserted being nonetheless in these latter cases of a much less significant size (1.1 and 1.5 Kb respectively). In a surprising manner, it was noted that a very high homologous recombination frequency (1/40) could be obtained with the pGMD vector. The elimination of 3' signals of maturation of mRNA and the addition of the sequence of degradation of mRNA to the gene of resistance to neomycin had the effect of reducing the total number of clones resistant to G418 by 2.4 (table I). The specific targeting result was almost ten times higher (900/40). The mechanism of homologous recombination was probably affected in the experiments with pGMD. A possible explanation for these results could be that an AT sequence of 51 pb could provide, *in vivo*, an open loop in the mutagenesis plasmid because of its lower fusion temperature. If the HOXC8 (Hoc-3.1) sequences close to the pGMD can be influenced by this opening, on each side of the AT region, they could react in a more efficient manner, in the single-strand state, with the chromosomal locus HOXC8 (Hoc-3.1). The model of mitotic recombination in yeast suggests that it could be initiated by such an exchange of strands, even though the mechanism of homologous recombination remains unknown in the more complex eucaryotes.

Figure 8 shows the results of Southern analysis carried out on positive individual clones (L5 and F2) and E.S. cells (C.C.E.).

The probes used only hybridize with HOXC8 (Hoc-3.1) sequences included in the vector (a) or excluded from the mutagenesis vector (b). The image of hybridization of the recombined locus HOXC8 (Hoc-3.1) (open triangles) is clearly distinguished from the wild-type locus (black triangles). The stars indicate the hybridization bands of a copy of the plasmid which



was integrated at random. The size marker is a digest Eco RI + Hind III of Dna lambda.

Figure 8(c) shows the restriction maps of recombined (rec.) and wild-type (wt) HOXC8 (Hoc-3.1) alleles. The parts of the mutagenesis vector and of the locus HOXC8 (Hoc-3.1) are indicated by the same symbols as those used in Figure 2. In this case, the sequence AT was integrated by homologous recombination. The vertical arrow indicates the 3' end of the mutagenesis plasmid. The localization of probes "a" and "b" used in the Southern analysis is also indicated. The abbreviations used in Figure 8 are the following: B, Bam HI; D, Dra I; E, Eco RI, H, Hind III; S, Sal I; X, Xho I.

15 IV - Production of chimeric embryos

A micro-injection was made in blastocysts with two recombinant E.S. clones containing an intact allele HOXC8 (Hoc-3.1) and a recombined allele, these clones did not contain any other copy of the mutagenesis plasmid. The karyotypes of these cells were normal.

Ten to fifteen mutated cells were micro-injected per blastocyst. After reimplantation in surrogate mothers, the embryos were taken at 9.5, 10.5 and 12.5 days p.c. and analyzed for LacZ expression. The transcription template of HOXC8 (Hoc-3.1) at these stages had been determined beforehand by hybridization analysis *in situ* (Ref. 1). The HOXC8 (Hoc-3.1) transcripts are detectable for the first time at the stage of late gastrulation and are spread throughout all the tissues of the posterior part of the animal. Later, the distribution becomes progressively limited in space and specific to the tissue. At the stage of 12.5 days p.c. the transcription is localized in the cervical region of the neural tube, at the level of the heart. During embryogenesis, the repartition of the transcription of HOXC8 (Hoc-3.1) thus undergoes



modifications. The stage 10.5 days p.c. seems to be a period of transition, the transcription taking place both in the two rear regions and in the cervical neural tube.

- 5 In chimeric embryos of 9.5 and 10.5 days p.c., the caudal part to the posterior bud presented intense β -Gal activity, while the marker was never detected in the anterior thoracic region or the head (Fig. 9a). In the posterior region, the cells coloured by β -Gal were
10 observed in all the tissues and all the embryonic layers. Between the two buds which form the members, coloured cells were distributed in the restricted zones, in the superficial ectoderm (Fig. 9b), as in the rear regions (Fig. 9c) and, in straight lines or rays,
15 in the neural tube (Fig. 9b). These rays showed an irregular and asymmetric distribution on the wall of the neural tube. The transcription of HOXC8 (Hoc-3.1) was not detected in the thin layer of cells near the closure of the neural tube. Perhaps these cells did not
20 resist the treatments applied during the hybridization in situ. It was observed that the cells of the neural ectoderm very soon become part of the different parts of the nervous system and migrate in a radial direction, following narrow lateral movements (Ref.
25 21). These results are thus in line with this observation.

The expression of LacZ thus illustrated correctly the first part of the transcription of the homeogene HOXC8 (Hoc-3.1), that is to say in all the tissues of
30 the caudal regions of embryos of 9.5 and 10.5 days p.c., and provided new information concerning the mode of transcription of HOXC8 (Hoc-3.1).

On the other hand, the expression of LacZ was not observed in the cervical regions of the neural tube of
35 chimeric embryos of 12.5 days, nor in the anterior



region of embryos of 10.5 days; this was not the result expected from studies of hybridization in situ. The ulterior phase of transcription of HOXC8 (Hoc-3.1) observed from day 10.5 in the very localized zones of the neural tube were not made evident by the activity of β -Gal. A possible explanation for this result could be that, even though the expression of LacZ was under the control of the promoter HOXC8 (Hoc-3.1), the 3' sequences of HOXC8 (Hoc-3.1) are absent in the reporter gene. It is possible that the 3' sequences of the AUG initiation codon of HOXC8 (Hoc-3.1) have an influence on the late expression of HOXC8 (Hoc-3.1) in the anterior domain. An effect of "gene dosing" could also explain this result. Auto-activation of several homeogenes of *Drosophila* was demonstrated genetically or suggested by the formation of complexes between the DNA and the proteins of the homeoboxes.

If the late component of transcription of HOXC8 (Hoc-3.1) in the neural tube is maintained by a similar mechanism, the inactivation of an allele could have a dominating effect in the cells of the neural ectoderm. Since one allele only would produce the protein HOXC8 (Hoc-3.1), the signal of activation would be diluted on the two promoters. The reduction of auto-inactivation in the two loci could thus lead to a total interruption of the transcription initiation. This could explain why no expression of LacZ was detected in the cervical region of the neural tube of embryos of 10.5 and 12.5 days.

V - Passage of the modification into the germ cell line: production of transgenic animals

The effects in F₁ and in F₂ of the modification brought by targeted insertion were observed after reproduction of the chimeric animals. Passage of the modification into the germ cell line was observed.



BIBLIOGRAPHY

1. Le Mouellic, H., Condamine, H. et Brûlet, P.
(1988), *Gènes & Développement*. 2, 125-135.
- 5 2. Cavenir, D.R. (1987). *Nucleic Acid Res.* 15,
1353-1361.
3. Sanes, J.R. Rubenstein, J.L.R. and Nicolas,
J.F. (1986) *EMBO J.* 5, 3133-3142.
4. Nolan, G.P., Fiering, S., Nicolas, J.F. and
10 Hertenberg, L.A. (1988) *Proc. Natl. Acad. Sci. USA*
85, 2603-2607.
5. Gorman, C., Padmanabhan, R. and Howard, B.H.
(1983). *Science*. 221, 661-553.
6. Gormann, C.M., Merlino, G.T., Willingham,
15 M.C., Pastan, I. and Howard, B.H. (1982) *Proc. Natl.*
Acad. Sci. USA 79, 6777-6781.
7. Southern, P.J. and Berg, P. (1982) *J. Mol.*
Appl. Genet. 1, 327-341.
8. Herbomel, P., Bourachot, B. and Yaniv, M.
20 (1984). *Cell*. 39, 653-662.
9. Robertson, E.J. (1987). *Teratocarcinomas and*
embryonic stem cells: A practical approach. IRL Press,
Oxford.
10. Khan, A. (1988). *Medecine/Sciences* 4, 515-
25 518.
11. G. Chu, Hayakawa, H., Berg, P. (1987).
Nucleic acid Research 15, No. 3, 1311-1326.
12. Thompson, S., Clarke, A.R., Pow, A.M.,
Hooper, M.L., Melton, D.W., (1989) *Cell*, 56, 313-321.
- 30 13. Thomas, K.R., Capecchi, M.R., (1987) *Cell*,
51 503-512.
14. Evans, M.J., Kaufmann, M.H. (1981) *Nature*,
292, 154-155.
15. Robertson, E.J., (1986) *Trends in Genetics*,
35 9-13.



16. Robertson, E., Bradley, A., Kuehn, M. &
Evans, M., Nature 323, 445-448 (1986).
17. Robertson, E.J. in Teratocarcinomas and
Embryonic Stem Cells (ed. Robertson, E.J.) *71-112
5 (IRL, Oxford, 1987).
18. Kim, H.S. & Smithies, O. Nucleic Acids. Res.
16, 8887-8903 (1988).
19. Breier, G., Bucan, M., Francke, U., Colberg-
Poley, A.M. & Gruss, P. EMBO J.5, 2209-2215 (1986).
- 10 20. Joyner, A.L., Skarnes, W.C. & Rossant, J.
Nature 338, 153-156 (1989).
21. McKay, R. D. G. Cell 58, 815-821 (1989).

This Sheet Has Been Replaced

CLAIMS

1. Transgenic or chimeric animal comprising cells whose genome comprises:

- a heterologous DNA, called insertion DNA, and being constituted by the totality of a gene or by a
5 part of a gene;

- the insertion DNA being positioned in the genome in relationship to an endogenous gene, called the receiver gene, comprising a complementary DNA, in such a way that the insertion DNA or the complementary
10 DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

2. Animal according to claim 1, characterized in that the insertion DNA contains either a coding
15 sequence, whose expression takes place under the control of regulatory sequences of the receiver gene, or a regulatory sequence permitting the expression of a coding sequence of the receiver gene.

3. Animal according to claim 2, characterized in that the insertion DNA is constituted by a coding
20 sequence other than the coding sequence of a gene coding for a selection agent.

4. Animal according to claim 1, characterized in that the insertion DNA comprises the totality of a
25 gene.

5. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is a gene which normally is not transcribed in the cell.

6. Animal as claimed in any one of the preceding
30 claims characterized in that the receiver gene is inactivated following the insertion of the insertion DNA, the expression of the insertion DNA substituting that of the receiver gene.





THIS SHEET HAS BEEN
REPLACED

7. Animal as claimed in any one of the preceding claims characterized in that the insertion DNA replaces a part of the receiver gene.

5 8. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is, apart from the presence of the insertion DNA, in its natural genetic environment.

9. Animal as claimed in any one of the preceding claims characterized in that the genome contains,
10 adjacent to the insertion DNA, a so-called "intercalated" sequence comprising for example a sequence coding for a selection agent with a promoter authorizing the expression of the agent of selection, or a marker.

15 10. Transgenic animal in which a single copy of a gene which is present in the genome in at least two copies, has been inactivated by the insertion of a gene which is different from the inactivated gene, the insertion gene being inserted in a position which
20 permits the expression of this gene under the control of regulatory sequences of the inactivated endogenous gene.

11. Animal as claimed in any one of the preceding claims utilized for a genetic marking.

25 12. Process for screening pharmaceutical products presumed to have an activity with regard to expression products of a pathological gene linked to a disease, characterized by the administration of the product to a transgenic animal according to any one of the claims 1
30 to 11, the insertion DNA being constituted by the pathological gene or a fragment of it.

13. Nucleic acid obtained from the homologous recombination between an insertion DNA and a receiver gene, said nucleic acid comprising:

This Sheet Has Been Replaced





THIS SHEET HAS BEEN
REPLACED

- an insertion DNA, heterologous with respect to the receiver gene, the insertion DNA being constituted by the totality of a gene or by a part of a gene;

- a receiver gene;

5 - the insertion DNA being positioned in relationship to the receiver gene comprising a complementary DNA in such a way that the insertion DNA or the complementary DNA is rendered functional, or its
10 functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

14. Cell of a transgenic animal as claimed in any one of claims 1 to 10.

15 15. Cell line issuing from a transgenic animal as claimed in any one of claims 1 to 10.

This Sheet Has Been Replaced



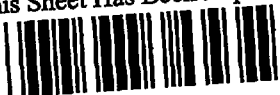
ABSTRACT OF THE DISCLOSURETRANSGENIC OR CHIMERIC ANIMAL AND PROCESS FOR
SCREENING PHARMACEUTICAL PRODUCTS

The invention relates to a transgenic or chimeric animal carrying cells whose genome comprises:

- a heterologous DNA, called insertion DNA, and being constituted by the totality of a gene or by a
5 part of a gene;
- the insertion DNA being positioned in the genome in relationship to an endogenous gene, called the receiver gene, comprising a complementary DNA, in such a way that the insertion DNA or the complementary
10 DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

Figure 1

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

1/12

pGN

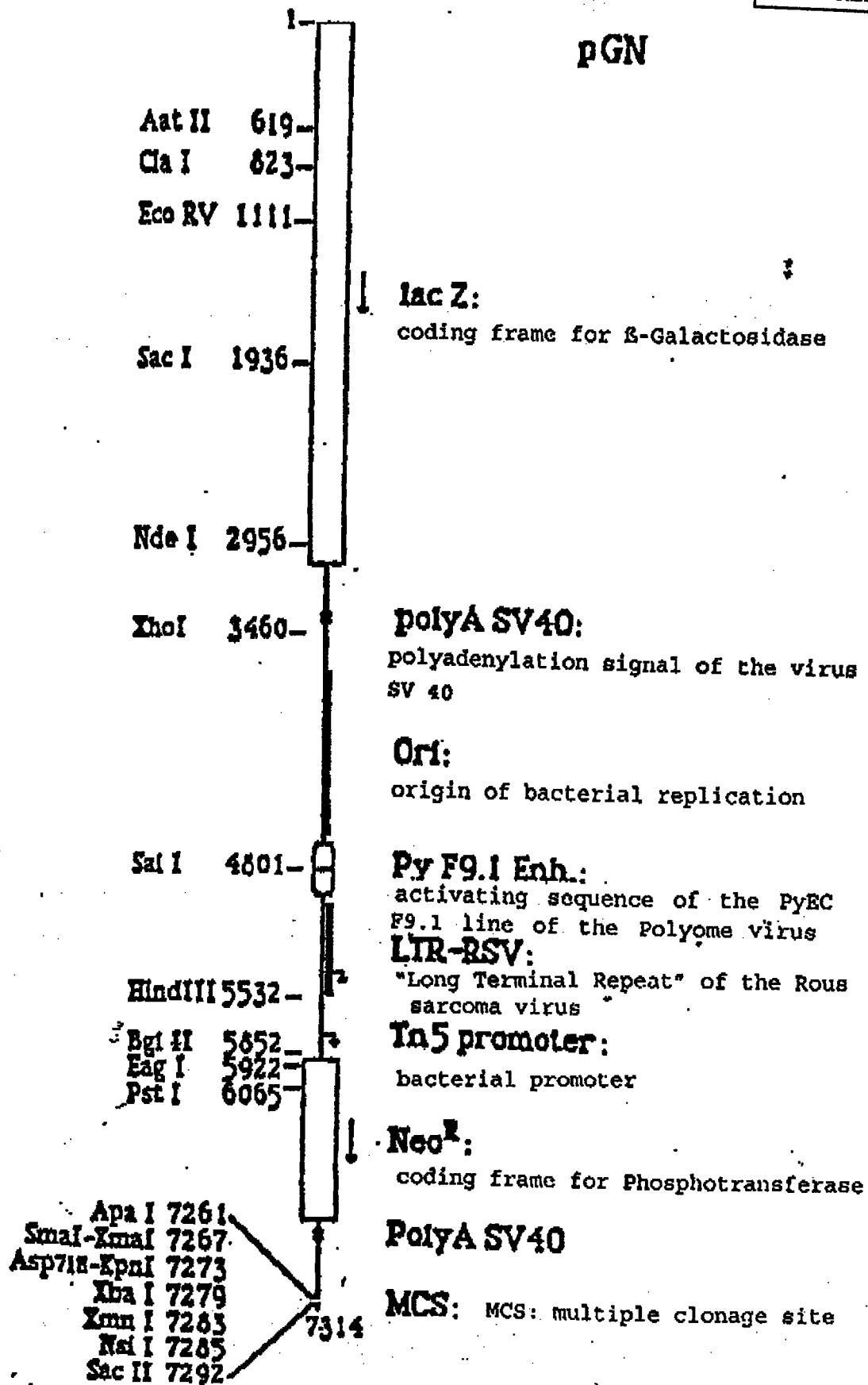


Figure 1: linear plasmid

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

2/12

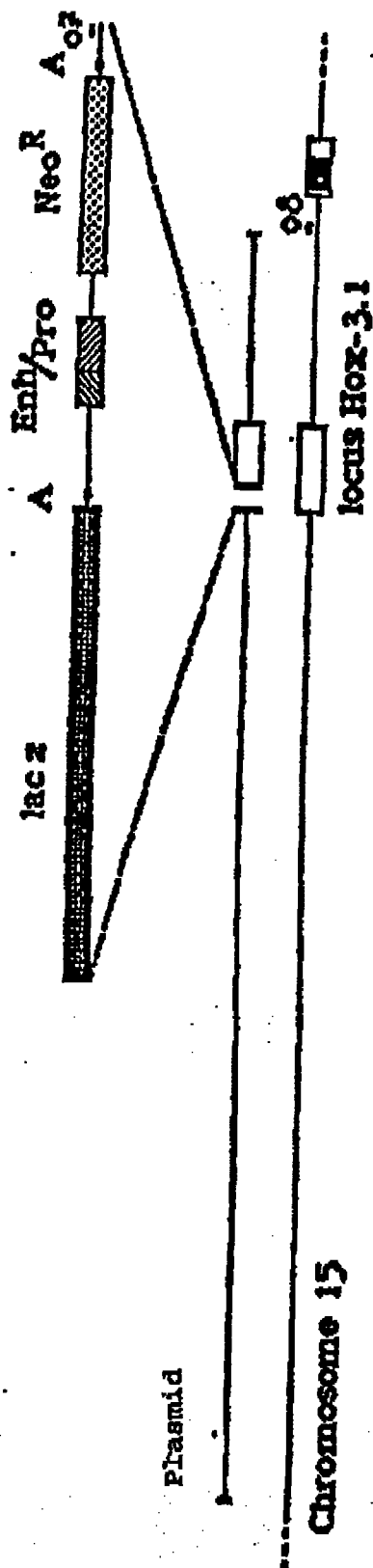


FIGURE 2a

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

3/12

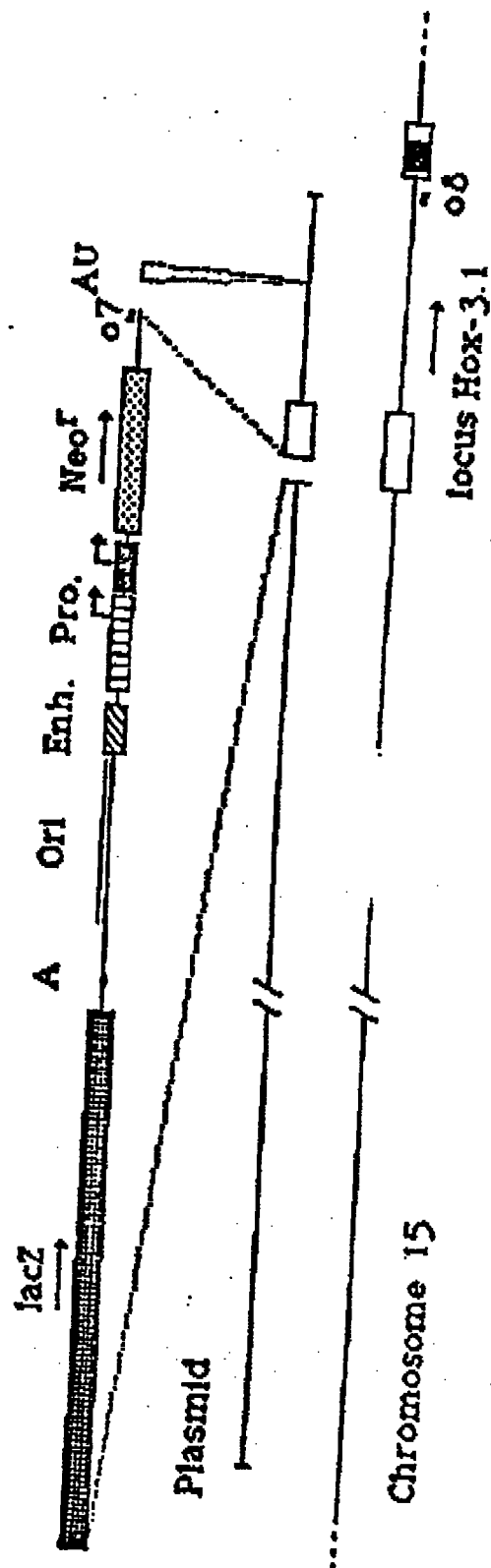


FIGURE 2b

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

4/12

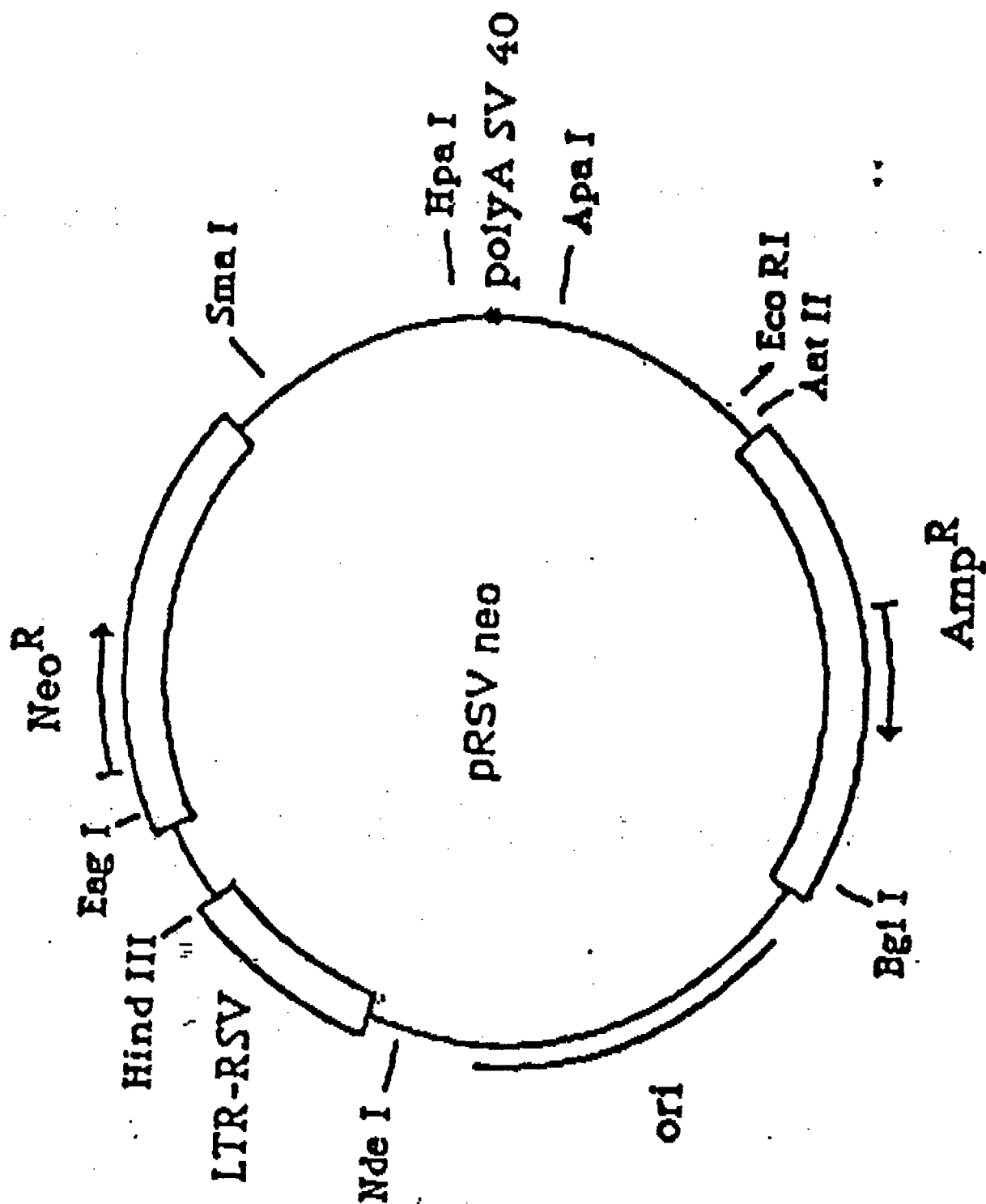


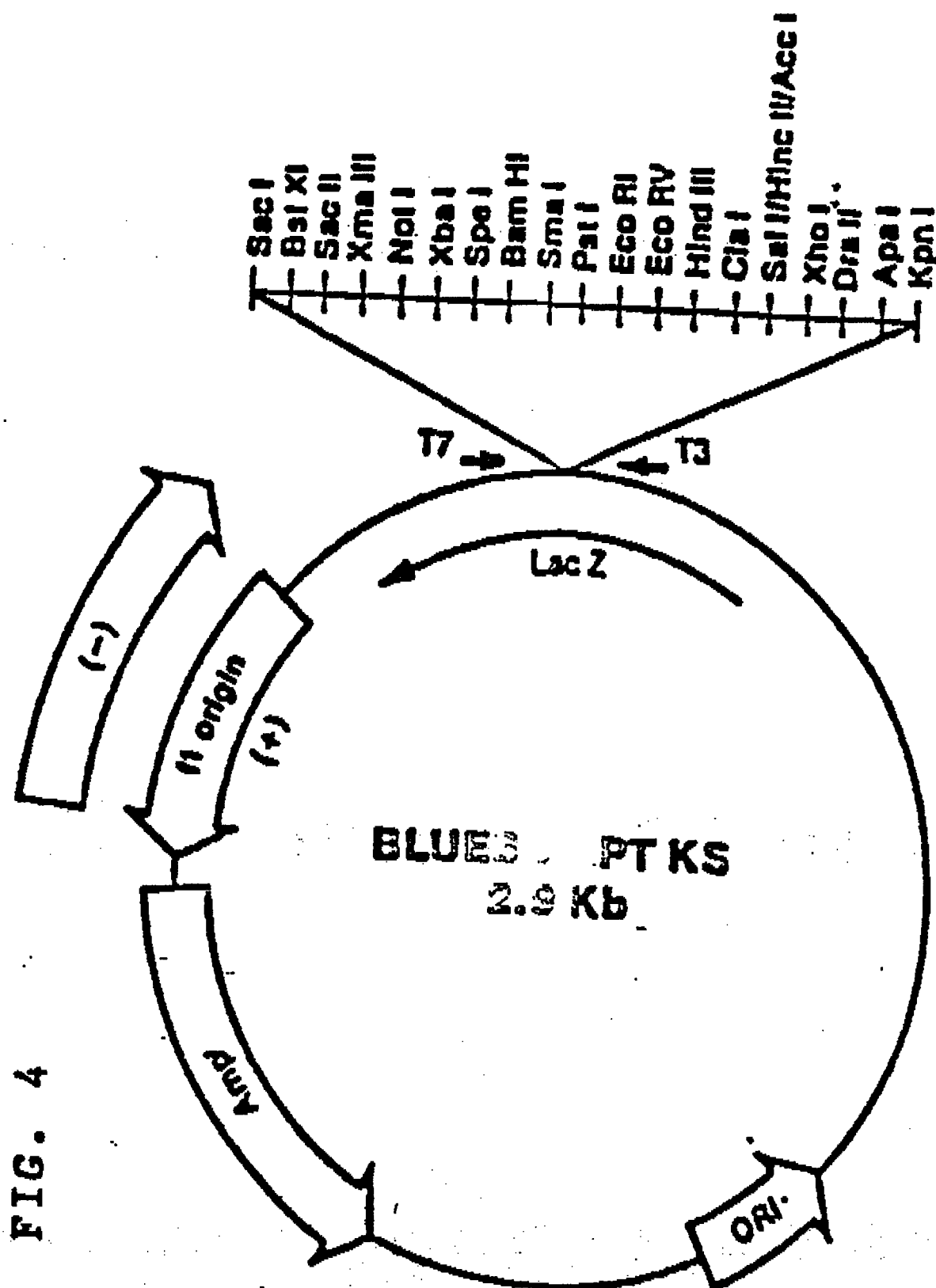
FIG. 3

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

5/12



This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

6/12

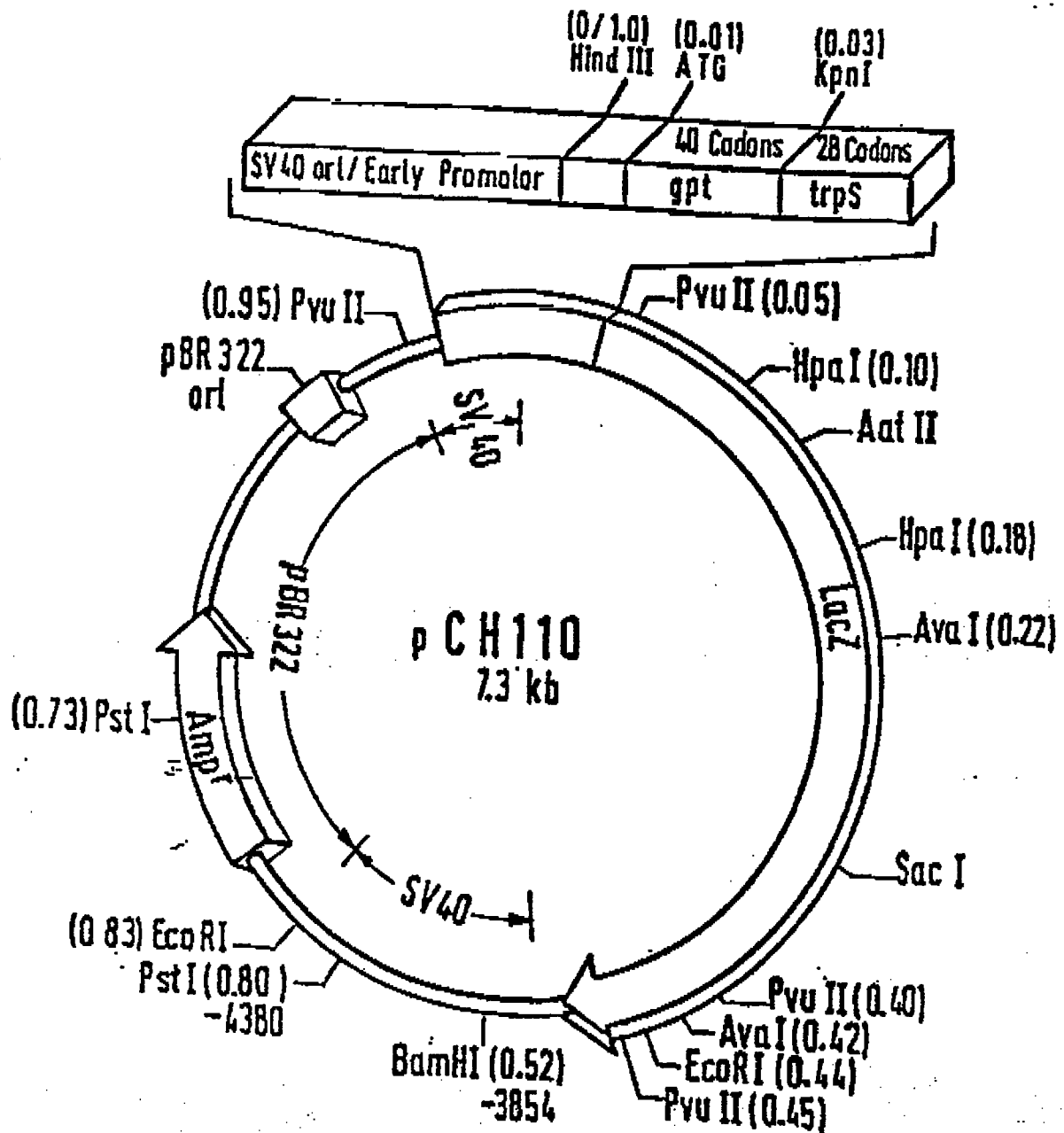


FIG. 5

This Sheet Has Been Replaced



7/12

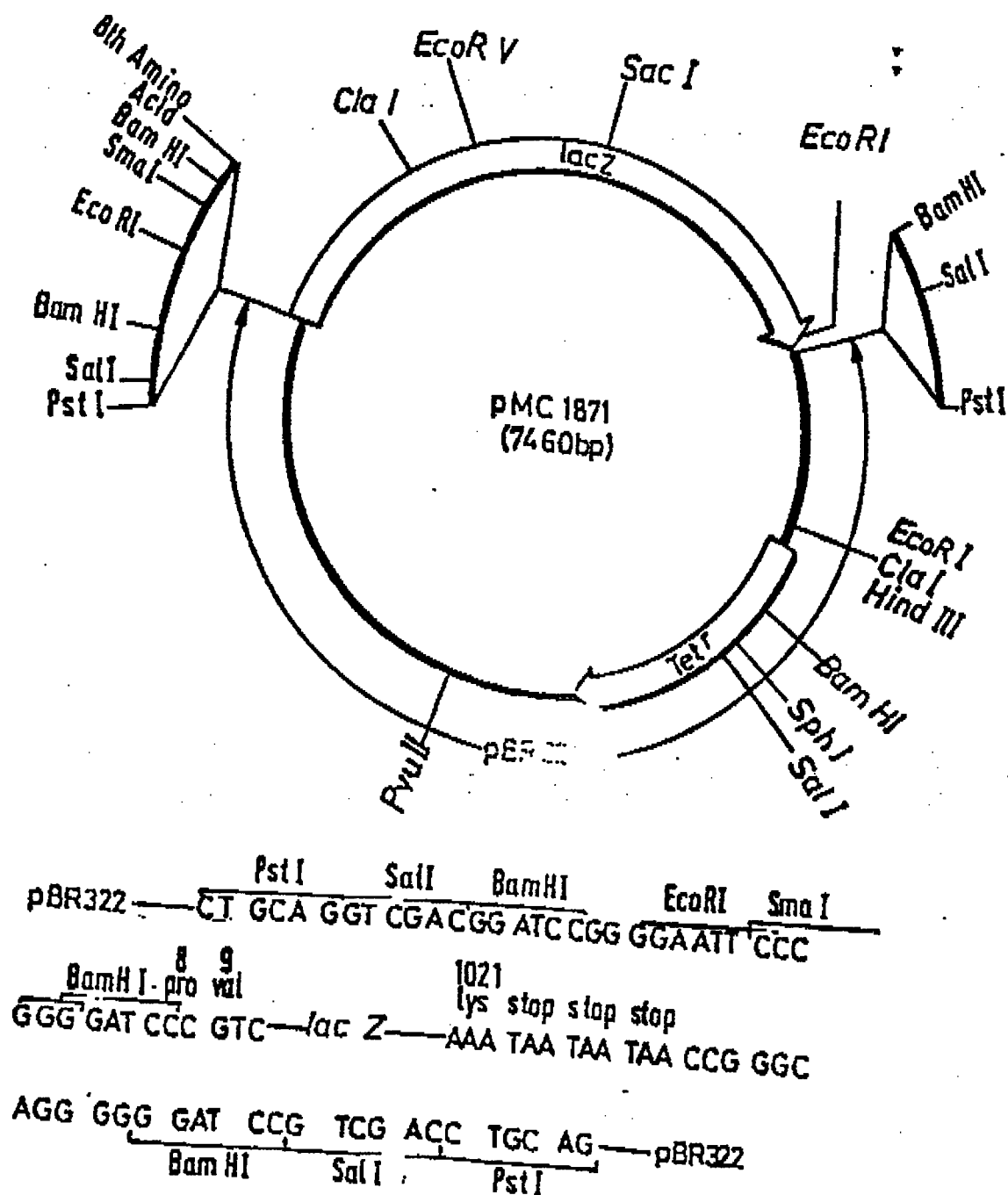


FIG.6

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

8/12

1 2 3 4 5 6 7 8 9 10 M D :



FIG. 7

This Sheet Has Been Replaced

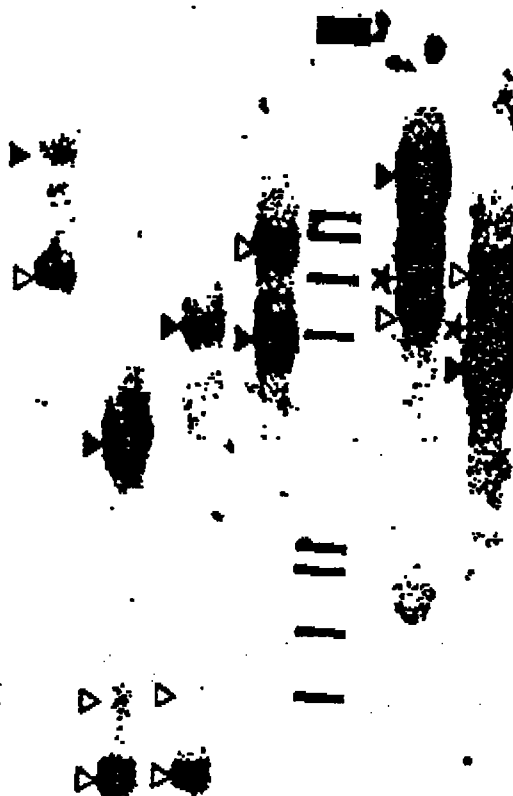


THIS SHEET HAS BEEN
REPLACED

9/12

a

$\xrightarrow{\text{L5}}$ $\xrightarrow{\lambda}$ $\xrightarrow{\text{F2}}$
 B+H D+E D E B+H E

*b*

$\xrightarrow{\text{L5}}$ $\xrightarrow{\lambda}$ $\xrightarrow{\text{ES}}$
 D D+E E B+H E B+H

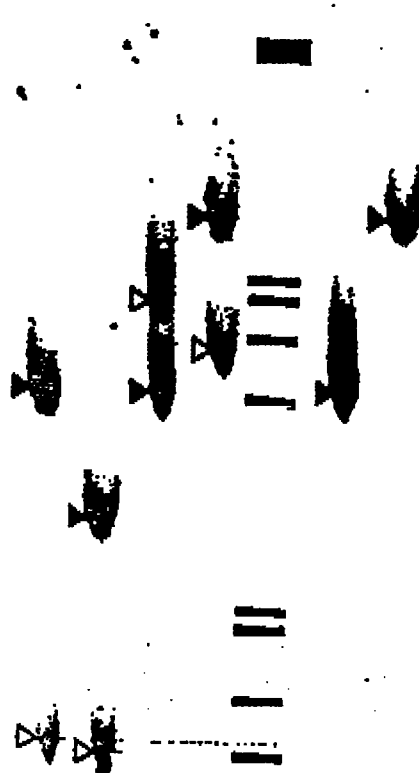
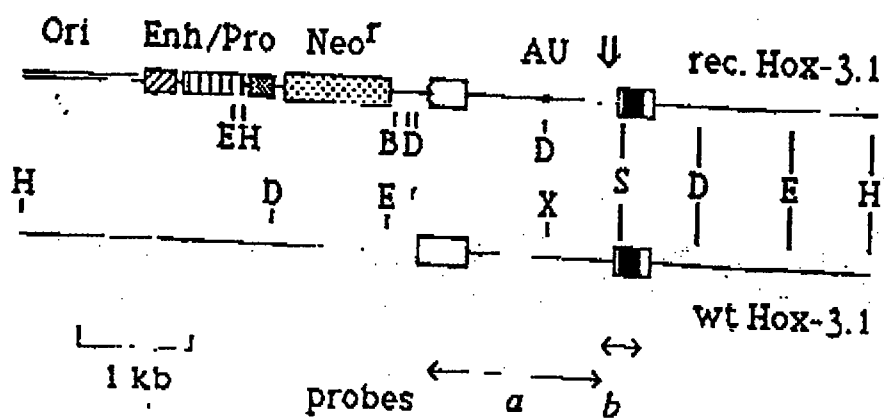
*c*

FIG. 8

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

10/12



Fig. 9A

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

11/12



Fig. 98

This Sheet Has Been Replaced



THIS SHEET HAS BEEN
REPLACED

12/12



Fig. 9C

This Sheet Has Been Replaced



09 NOV 1996

TRANSGENIC OR CHIMERIC ANIMAL AND PROCESS FOR SCREENING
PHARMACEUTICAL PRODUCTS

The present invention relates to a process for specific replacement of a copy of a gene present in the genome of a receiver eucaryotic organism by the integration of a gene different from the inactivated gene. The receiver gene will preferably be present in at least 2 examples in the transfected host cell. The receiver gene is defined as being the gene where the insertion of the different gene will be made.

More particularly, the invention relates to the production of transgenic animals in which the foreign gene has been introduced in a targeted manner to allow both the maintenance of the normal genetic functions of the animal and the expression of the foreign gene under the control of endogenous promoters.

By "different or foreign gene" is meant any nucleotide sequence corresponding to the totality or to a part of a "foreign or different" gene of the receptor gene as it is found normally in the genome (RNA or DNA), where it corresponds also to an artificially modified sequence of the normal gene or again to a fragment of this sequence.

The invention also concerns the process of production of these transgenic animals.

In the production of transgenic animals, the conventional methods used for the introduction of heterologous DNA sequences in the germ cell line do not allow control of the site of integration of the foreign gene in the genome, nor the number of copies thus introduced. The integration of the foreign gene is carried out at random and, in general, several copies of the gene combine at the same time, sometimes under the form of a head-to-tail tandem, the site of

09 NOV 1996

integration and the number of copies integrated varying from one transgenic animal to another.

It can thus happen that the endogenous cellular genes, situated at the point of insertion, may be inactivated, without this being easily detectable because of the numerous insertions made at random. If the product of these genes is important for the development of the animal, this development will be seriously disturbed. Moreover, the aleatory insertion of the foreign gene may take place at a site which is not appropriate for the expression of the gene. In addition, the fact that there is a variation in the site and in the number of insertions from animal to animal makes interpretation of studies of expression extremely difficult.

A major problem encountered in the production of transgenic animals is obtaining the expression of the foreign gene. Generally speaking, two types of experiment have been made on mice.

The genes introduced in the germ line are:

- either "complete" genes, including coding sequences flanked by their own regulatory sequences;
- or composite genes, formed from the coding sequence of a gene fused with the promoter sequence of another gene, the two fragments sometimes even belonging to two different animal species.

It was thus possible to confirm that the specificity of the expression of genes in such and such a tissue is determined by their regulatory sequence(s).

The choice of the appropriate promoter for the expression of the foreign gene in the transgenic animal is thus of primordial importance.

Besides, the controlled mutagenesis of murine genes in embryonic cell lines has recently been carried out using a technique of gene targeting (Thomas et al., 1987; Thompson et al., 1989).

09 NOV 1996

In the first case, the HPRT murine gene was mutated by insertion and replacement and, in the second case, a mutated HPRT gene was corrected. Thompson et al. widened their experiments as far as obtaining chimeric mice and recorded the passage of the genetic modification into the germ cell line.

In each of the documents quoted, the precise site of integration was targeted by homologous recombination between, on the one hand, exogenous sequences carrying the mutation or correction included in the vector, under the control of an exogenous promoter and, on the other hand, their genomic homologue. This being so, it must be pointed out that earlier authors have carried out their experiments on a specific gene (HPRT) whose activation by mutation was accompanied by a detectable phenotype. The targeted mutation described by Thomas et al. had the effect of inactivating the HPRT gene and, consequently, making the detectable phenotype normally associated with the HPRT disappear. The Neo^R selection gene, under the control of a TK promoter, was thus incorporated into the insertion DNA in order to permit the selection of transformants. It should be noted that the experiments described in the prior art implied a selection either by the receiver gene (e.g. HPRT) or by the insertion gene (e.g. Neo^R). The site of insertion and/or the type of gene inserted is thus limited to genes conferring a selectable character, a direct selection.

Besides, in the prior art, the exogenous sequences on the vector thus serve both to target the site of integration and to introduce the modification. Following homologous recombination, the modified gene always finds itself in its normal genetic environment.

We should remember that a problem posed during the production of transgenic animals is the danger of

09 NOV 1996

inactivating an endogenous cellular gene existing at the insertion point of the foreign gene.

According to the function of the product of the inactivated gene, such an inactivation can lead to significant physiological or morphological disorders in the transgenic animal, or could even prevent its survival.

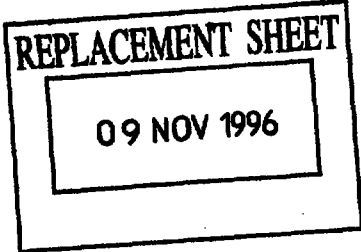
On the other hand, the inactivation of the gene could be considered as advantageous if the gene in question coded for a virus receptor or other infectious agent.

The inventors have studied the possibility of avoiding the disadvantages described above and associated, in certain cases, to the possible inactivation of one or several endogenous cell genes with an important function during the production of transgenic animals.

The object of the invention is a process for replacement or specific insertion of a gene, in particular by targeting a DNA, called insertion DNA constituted by the totality of a gene, or by a part of a gene, susceptible to being rendered functional, where this functioning can be made more efficient, when it is recombined with a complementary DNA to then provide a complete recombinant gene in the genome of a eucaryotic cell, characterized in that

- the insertion site is in a chosen gene, called a receiver gene, and containing the complementary DNA, and in that

- eucaryotic cells are transfected with a vector containing an insert which comprises the insertion DNA and two so-called "flanking" sequences on either side of the insertion DNA, respectively homologous to two genomic sequences which are immediately adjacent to the desired site of insertion in the receiver gene,



- the insertion DNA being heterologous vis-à-vis the receiver gene, and

- the flanking sequences, being chosen from those which make up the above-mentioned complementary DNA and which allow, through homologous recombination with corresponding sequences of the receiver gene, the reconstitution of a complete recombinant gene in the genome of the eucaryotic cell.

According to the invention, the so-called insertion DNA can contain either a coding sequence or a regulatory sequence, the process then being characterized in that

- the insertion site is in a chosen gene, called a receiver gene, and in that

- eucaryotic cells are transfected with a vector containing an insert which comprises the insertion DNA and two so-called "flanking" sequences on either side of the insertion DNA, respectively homologous to two genomic sequences which are immediately adjacent to the desired site of insertion in the receiver gene,

- the insertion DNA being heterologous with respect to the receiver gene, and

- the flanking sequences being chosen so as to permit by homologous recombination depending on the case, either the expression of the coding sequence of the entire insertion DNA under the control of the regulatory sequences of the receiver gene, or the expression of a coding sequence of the receiver gene under the control of regulatory sequences of the insertion DNA.

For example, the insertion DNA may contain a coding sequence without the regulation element, notably without its own promoter.

The invention also relates to a process for the production of transgenic animals, characterized in that E.S. cells are transfected in the conditions described

09 NOV 1996

above and selected for the occurrence of homologous recombination, that is the correct integration of the foreign gene, the transfected cells are injected into embryos at a stage where they are able to integrate the transfected cells (for example, at the blastocyst stage), and these are then reimplanted in a surrogate mother and the chimeric individuals obtained at the end of gestation, and in which is noted the colonization through E.S. cells of the germ line of the chimeric animal, are mated. If the E.S. cells have colonized the germinal line of the chimeric animal, heterozygous transgenic animals for the replaced gene will be obtained by mating (F1) the progeny.

It is also possible to insert insertion DNA, carried by the vector of the invention, in the egg shortly after fertilization (that is less than 24 hours). In this way, the insertion takes place while the egg is still in the unicellular state.

The invention also concerns a plasmid able to effect the targeted insertion of a recombinant gene called the insertion gene in the genome of a eucaryotic cell, characterized in that it contains an insert which itself comprises the insertion gene and two so-called "flanking" sequences on either side of the insertion gene, respectively homologous to the two genomic sequences which are immediately adjacent to the desired site of insertion in the receiver gene.

The insertion DNA can be a gene, or a part of a gene, heterologous to the transfected species.

The invention also concerns transgenic animals in which at least one endogenous gene, which is preferably present in the genome in at least two copies, has been inactivated by the insertion of a gene which is different from the inactivated gene, the insertion gene being inserted in a position which allows the

expression of this gene under the control of regulatory sequences of the inactivated endogenous gene.

Thus, the process of the invention allows, thanks to the phenomenon of homologous recombination, the insertion in a targeted manner of foreign genes, in particular coded sequences without the promoter with which they are usually associated, in the genome of a eucaryotic organism at a site which allows their expression under the control of the endogenous promoter of the gene where the insertion is made, and consequently, the inactivation of the targeted endogenous gene.

In a preferred embodiment of the invention, the targeted receiver gene is a gene which is present in the genome in at least two copies. Utilization of the electroporation technique (ref. 11) ensures the introduction of only one copy of the foreign gene.

In this embodiment of the invention, the targeted insertion of the gene of interest (i.e. the so-called insertion gene) has the effect of inactivating the single copy of the endogenous cellular gene where the insertion takes place and leaves intact and operational the other copy (copies) of this gene.

In this way, the genetic functioning of the transgenic animal is not or is little disturbed by the introduction of the foreign gene, even if the insertion inactivates a single copy of a receiving gene essential for the development of the animal. Either its development is not affected by the insertion of the foreign gene, or the possible minor disturbances in the case of inactivation of a critical gene will probably not be lethal for the animal. The effects of the insertion of the foreign gene in the homozygous state could be of any kind and would be observed in the second generation (F2) after crossing heterozygous individuals (F1) among themselves.

09 NOV 1996

If, on the other hand, the inactivation of all the copies of a gene is desired, for example in the case the gene codes for a receptor of an infectious agent, multiple copies of the foreign gene are introduced. Control of the quantity introduced can be ensured by using known techniques.

The targeted insertion of the foreign gene thus allows its introduction in a site where its expression is under the control of regulatory sequences of the endogenous gene where the insertion takes place.

The process of the invention also allows insertion of the foreign gene behind an endogenous promoter which has the desired functions (for example, specificity of expression in a particular tissue) and this, in the relevant case, without inactivating the other copies of the receiver gene.

In a preferred embodiment, the insertion DNA, or the vector, contains sequences inserted between the insertion DNA and the flanking sequences. These sequences can for example contain a sequence coding for a selective agent allowing the selection of transformants and eventually a marker gene, for example LacZ.

In a highly preferred embodiment of the invention, the insertion DNA carries, between the flanking sequences, on the one hand a DNA sequence intended to be recombined with the complementary DNA in the receiver gene to provide a recombinant gene and, on the other hand, a sequence coding for a selective agent permitting the selection of transformants and a promoter authorizing the expression of the selective agent, the receiver gene and the recombinant gene coding for expression products not conferring a selectable phenotype. According to this embodiment, the insert of the plasmid of the invention comprises, between the flanking sequences, on the one hand a DNA

09 NOV 1996

sequence intended to be recombined with the complementary DNA in the receiver gene and, on the other hand, a sequence coding for a selective agent allowing the selection of transformants and a promoter
5 authorizing the expression of the selective agent, the DNA sequence intended to be recombined with the complementary DNA being other than a gene coding for a selective agent.

In this way, the selection of transformants is
0 entirely independent from the nature of the receiver gene and the inserted gene, contrary to the processes described until now, in which the inserted gene or the receiver gene necessarily had to code for a product of expression allowing the selection of transformants. The system developed by the inventors gives total flexibility regarding the nature of the receiver gene and the inserted gene, or the gene formed by homologous recombination. The inventors have noted that, surprisingly, the insertion of sequences of a significant size (for example about 7.5 kb) does not affect the frequency of homologous recombination.

The possible effect of the insertion of the DNA sequence according to this aspect of the invention includes, depending on the type of sequence inserted for example the replacement of a coding sequence, the replacement of a regulatory sequence, the inactivation or reactivation of a gene by mutation or improvement of the expression level of a gene. It is possible, according to the invention, to replace a coding frame or part of a coding frame by a heterologous sequence which begins at the initiation codon of the replaced gene so that the expression of the inserted gene entirely replaces the expression of the replaced gene. This avoids the formation of fusion proteins which could be undesirable in a transgenic animal.

09 NOV 1996

In this embodiment of the invention, the insertion DNA can carry between the flanking sequences a heterologous coding sequence without promoter, the coding sequence being other than a gene coding for a selection agent. The insertion DNA can also carry, downstream of the coding sequence and still between the flanking sequences, a gene coding for a selection agent, associated with a promoter allowing its expression in the target cell.

In this way, the heterologous coding sequence can be inserted after an endogenous promoter which has the desired properties, for example a certain specificity of expression, or transcription pattern etc., the selectability of the transformed cells being entirely independent from the expression of the heterologous coding sequence. This type of construction makes it possible, for example, to select the transformants even if the gene replaced by the heterologous coding sequence is not normally expressed in the target cells. This is particularly important in the production of transgenic animals from E.S. cells ("Embryonic Stem Cells") since a significant proportion of genes remains inactive until a more advanced stage of the development of the animal. The HOXC8 gene (also known under the name of Hox-3.1) is an example of this type of gene. Moreover, if the coding sequence codes for an easily detectable protein, for example β -Gal, the development of the transcription template of the replaced endogenous gene can be followed. The pGN vector is an example of this type of construction.

In another embodiment of the invention, the insertion DNA can carry a foreign regulatory sequence. The site of insertion and, consequently, the flanking sequences are chosen in relation with the desired aim, that is either the insertion of the foreign regulatory

09 NOV 1996

sequence to give an effect of "double promoter" with the endogenous regulatory sequence, or the replacement of an endogenous promoter by the foreign promoter. The coding sequence which comes under the control of the regulatory sequence can be endogenous.

Another possibility can be the targeted insertion of a foreign DNA which carries both a regulatory sequence and a coding sequence. It is possible that the regulatory sequence can be that which is naturally associated with the coding sequence.

The process of the invention activates a vector containing two "flanking" sequences on either side of the foreign gene. These flanking sequences have at least 150 base pairs and are preferably inferior to the length of the receiver gene. It is essential that these two flanking sequences be homologous to the two genomic sequences which are immediately adjacent to the desired insertion site. The flanking sequence of the vector which is located upstream of the foreign gene to be introduced is normally homologous to the genomic sequence which is located at the 5' end of the insertion site. In the same way, the flanking sequence of the vector located downstream of the foreign gene is normally homologous to the genomic sequence which is located at the 3' end of the site of insertion.

It is possible to introduce "intercalated" sequences between either one of the flanking sequences and the foreign gene, for example sequences permitting the selection of transformants, markers, sequences permitting cloning of the vector etc.

The position of these intercalated sequences with respect to the foreign gene must however be chosen so as not to hinder the expression of the foreign gene, in particular the coding sequence of the foreign DNA under the control of the endogenous promoter or, inversely, the coding sequence of the endogenous DNA under the

09 NOV 1996

12

control of foreign elements of regulation carried by the insertion sequence.

Despite the presence of flanking sequences, which encourages homologous recombination, it is possible that a certain number of integrations take place at random. In order to verify that the targeted insertion has actually taken place in the targeted site and not in another place, the technique of "Polymerase Chain Reaction" (P.C.R.) (see Ref. 10) is used to amplify the DNA sequence of the locus where the insertion takes place, or should take place. In this way, only the clones transformed following a homologous recombination are selected.

The flanking sequences of the vector are evidently chosen in relation to the desired site of insertion so that homologous recombination can take place. Thus the flanking sequences can carry duplicated sequences of the endogenous promoter and/or modifications to the sequences which precede the initiation codon to raise the translation level (upstream sequences) and duplicated sequence of the termination sequences particularly of the sites of polyadenylation (downstream sequences).

The insertion gene can be any gene of interest. One will cite as non-limiting examples the gene LacZ (as in the model described below), the genes coding for interleukin or interferon, receptor genes for example of retinoic acid or beta-3 adrenergic or H.I.V., and genes known to be linked with certain diseases, for example myopathy etc.

The eucaryotic cells transformed by the process of the invention are also included in the present application.

In a preferred embodiment of the invention, the eucaryotic cells are embryonic cell lines (see Ref. 14 and 15).

09 NOV 1996

In fact, a mutated E.S. cell can be injected into an early embryo which, after reimplantation, can be born under a chimeric form. If the germ line is colonized by the mutated cell, the chimeric animal will transmit the mutation to its progeny. After this, it will be possible to observe the effects of this mutation, in the homozygous state in some individuals, on their development, their behaviour, their metabolism, their pathology etc.

Figure 1 shows the pGN plasmid, which is part of the invention.

Figures 2a and 2b show the molecules pGMA and pGMD respectively constructed from the pGN plasmid in relation to the gene HOXC8 (Hox-3.1). These plasmids are mutagenesis plasmids. The two parts of the coding sequence of the gene HOXC8 (Hox-3.1) are represented, on chromosome 15, with the "homeo" box in black. The corresponding sequences of HOXC8 (Hox-3.1) have been cloned in the pGN plasmid.

(A: signal of polyadenylation; Enh/Pro: enhancer-promoter).

07 and 08 show the two oligonucleotides used in the PCR.

Figures 3 to 6 show the plasmids used in the construction of the pGN.

Figure 7 illustrates the detection of homologous recombination by the Polymerase Chain Reaction (P.C.R.) technique on the transfected E.S. cells.

Figures 8a and 8b show Southern analyses of positive individual clones (L5 and F2) and E.S. cells (C.C.E.).

The process of the invention has a very wide industrial applicability and can vary according to the nature of the foreign gene introduced, for example for the production of transgenic animals.

The genetics of mammals will progress considerably in view of the recent possibility of specifically mutating any gene, thus making it possible to define its role more accurately. Through this
5 technology which involves homologous recombination and E.S. cells, valuable information will be obtained on oncogenes, growth factors, transcription factors etc., genes which concern very actual subjects both in fundamental research and in applied research. An
0 important outcome for medical research is the possibility of reproducing a human disease whose genetic determination is known (certain pathological human diseases, such as the myopathy of Duchesne) so as to study its mechanisms better and search for a therapy.

By applying the process of the invention, a gene known to be responsible for a certain disease is inserted in a targeted manner into the genome of an E.S. cell. The transgenic animal which results from this process presents a useful model of the disease.

If necessary, and as described above, normal genetic functions can be significantly maintained, despite the insertion of the foreign gene.

Another application of the process of the invention is the genetic marking of animals. This application consists in inserting an insertion gene which is easily detectable e.g. the LacZ gene which can then play the role of a cell marker. In this way, filiation studies e.g. in competition animals are facilitated, and the breed can be followed up.

Insertion of the LacZ gene as insertion gene also makes promoter studies possible. In view of the possibility of identifying the activity of β -galactosidase, the activity and specificity of
3 different endogenous promoters can be studied by

09 NOV 1996

15

targeting different sites in the same type or different types of cells. The same studies could be carried out on an entire organism, during development or, at the adult stage, by using the techniques of chimeric or transgenic animals.

The inventors have observed in a surprising manner that the frequency of homologous recombination is not affected by the insertion of fragments of significant size, for example LacZ. This observation has suggested to the inventors that the technique of homologous recombination could be well adapted to the insertion of other heterologous genes which are of significant size.

The invention is also aimed at the application of the process of replacement or specific insertion for genic therapy.

In view of the possibility of modifying the genome of an animal, the process of the invention can thus also be used as "gene therapy". The most obvious utilizations would consist in inactivating the genes of receptors of toxic or infectious agents (viruses or bacteria). If this mutagenesis proved to be lethal, it would be necessary to re-establish the lost function without re-establishing the sensitivity to the harmful agents. A modified gene coding for such a receptor could be reintroduced into the mutated cell, as long as the modification can be provoked by homologous recombination. This modification of the genetic character would confer on the animal an immunity against the disease under consideration.

This protocol can also intervene in the framework of auto-grafting. Cells, sick or healthy, taken from a patient, could be treated and immunized, then reimplanted back into the same individual.

The technique of the invention is also suitable for studies of the activity of pharmaceutical products

09 NOV 1996

presumed to have an activity with regard to products of expression of a pathological gene linked to a disease. In this case, the insertion gene is constituted by the pathological gene or a fragment of it, and one
5 administers to the transgenic animal the pharmaceutical product to be tested with a view to evaluating its action on the disease.

The invention will be illustrated by referring to the pGN plasmid and its utilization in the targeted
0 insertion of a foreign gene (LacZ, coding for the enzyme β -galactosidase of E. Coli) in the genome of a mouse E.S. cell. The gene LacZ was chosen because of the fact that its expression can easily be detected and is used simply to demonstrate the feasibility of the
3 technique.

The coding frame of the enzyme β -galactosidase of E. Coli (LacZ; 1-3057), fused with a genomic sequence (7292-3) of the murine gene HOXC8 (Hox.3-1) (Ref. 1) begins with the initiation codon of this gene. In fact,
) the sequence which precedes the initiation codon of HOXC8 (Hox.-3.1) is identical to the consensus sequence observed in vertebrates (Ref.2), thus allowing a better translation rate of β -galactosidase into the cells of vertebrates. The gene LacZ is followed by a signal of polyadenylation for example of the virus SV 40, like most of the eucaryotic genes, so as to stabilize the messenger RNAs.

The activity of β -galactosidase of E. Coli, which is functional in the eucaryotic cells, can be detected in different manners. Cells expressing the LacZ gene take on a blue colouring, after fixation, in the presence of X-Gal, which is a substrate of β -galactosidase (Ref. 3). A new substrate, FDG (fluorescein di- β -galactopyranoside), makes it possible to detect and to dose the β -gal activity, while still

09 NOV 1996

17

keeping the cells alive (Ref. 4). The cells expressing LacZ accumulate a fluorescent product and can be isolated with the aid of a cell sorter or FACS (fluorescence-activated cell sorter).

5 The transcription unit of the gene with resistance to neomycin comes, to a large extent, from the plasmid pRSV neo (Ref. 5). The LTR (long terminal repeat) of the virus of Rous-sarcoma provides very powerful activating and promoting sequences in numerous
.0 eucaryotic cells (Ref. 6). From the bacterial transposon Tn5 comes a promoter active in E. Coli and the coding phase of the enzyme phosphotransferase (Ref. 7), which is followed by the polyadenylation signal of the virus SV 40. The same gene, under the double
5 control of promoters RSV and Tn5 can confer resistance to neomycin or kanamycin to bacteria and resistance to G418 to eucaryotic cells.

Through the effect of a simple single mutation, unit B of the activating (enhancer) sequences of the
2 strain PyEC F9.1 of the Polyoma virus has become much more active in different types of cells, and in particular in the cells of embryonic carcinoma (EC) (Ref.8). Two copies of this enhancer Py F9.1 have been inserted in tandem in the pGN plasmid, upstream of the LTR-RSV, and in the "late promoter" orientation of the regulatory region of the Polyoma.

In order to improve the translation rate of phosphotransferase, the sequence preceding the initiation codon was modified during a mutagenesis by oligonucleotide. Thus the sequence T T C G C A U G became G C A C C A U G, corresponding much better to the consensus sequence of the initiation of translation in vertebrates (Ref.2).

The improvements made to the transcription unit of the gene of resistance to neomycin could be estimated by transfecting embryonic cell strains (ES)

09 NOV 1996

of mice. With equal molarity in plasmid, a construction with the enhancer Py F9.1 produced 7.5 times more clones resistant to G418 than pRSV neo and 2 to 3 times more than the pMC1 neo described by Capecchi et al. (Ref. 13). Again, the number of clones was raised by 60x, that is 450x compared with pRSV neo, by modifying the sequence of initiation of translation. Homologous recombination may be a fairly rare event, according to the experimental conditions applied (e.g. 1/1000 for HPRT, Ref. 13). A vector presenting a high selection efficiency is thus very useful, especially since the conditions of electroporation result mainly in the integration of a single copy.

Moreover, the pGN plasmid contains an origin of bacterial replication of the type colE1, pBR322, which allows cloning and preparations in E. Coli.

Finally, a multiple cloning site (M.C.S.), synthesized in vitro, which only contains unique restriction sites in pGN, was inserted upstream of LacZ, so as to facilitate the utilizations of this plasmid.

The plasmid "flanking" sequences which provoke homologous recombination are added to the extremities of the pGN plasmid after linearization of the plasmid upstream from LacZ, by an MCS site (see Fig. 2). In this case, the flanking sequences chosen are homologous to the chromosome sequences taken from HOXC8 (Hox-3.1) which intervene later in the homologous recombination.

Figure 2 provides the location of the molecule built from the pGN plasmid with respect to the gene HOXC8 (Hox-3.1). In this case, recombination between the plasmid and chromosomal sequences of HOXC8 (Hox-3.1) would result in an insertion at the beginning of the coding phase of this gene, thus in its total inactivation.

09 NOV 1996

The pGN plasmid comprises several advantages for this methodology, which is applicable to any gene. As the event of homologous recombination can be fairly rare (of the order of 1 per 1000 non-homologous integrations), it is necessary to be able to analyze a large number of clones whose resistance to G418 is sufficiently strong to be expressed in any part of the genome. The modifications made to the transcription unit of phosphotransferase answer these problems perfectly. The method of mutagenesis by homologous recombination is equivalent to inactivating a gene by an insertion or a substitution, but the pGN plasmid presents the supplementary advantage of being able to substitute the expression of β -galactosidase to that of the mutated gene. Finally, the MCS facilitates the cloning of genetic fragments.

EXAMPLES:

I - Construction of the pGN plasmid

The intermediate plasmids are numbered according to their stage.

1st stage:

Insertion of a Xho I site in the Bgl I site of pRSV neo

Insertion of a Xho I linker in the Bgl I site of pRSV neo, filled by the Klenow fragment of the DNA polymerase of E. Coli.

2nd stage:

Insertion of a Cla I site in the Nde I site of the p1 plasmid

Insertion of a Cla I linker in the Nde I site of p1, filled by Klenow polymerase.

3rd stage:

Insertion of Py F9.1 enhancer in the Cla I site of the p2 plasmid

09 NOV 1996

20

Insertion of enhancer Py F9.1 Pvu II-Pvu II isolated by a unique Acc I site, in the Cla I site of p2. Selection of a clone containing two enhancers oriented in the direction "late promoter".

5 4th stage:

Deletion Sma I-Hpa I of the p3 plasmid

The two enzymes, presenting "blunt end" extremities can be bonded directly. This deletion removes the intron of the t antigen of SV 40, which is
0 not very useful, and reduces appreciably the size of the transcription unit of phosphotransferase.

5th stage:

Insertion of a Xho I site in the Bam HI site of pCH110

5 Insertion of a Xho I linker in the Bam HI site of plasmid pCH 110 (Pharmacia), filled by the Klenow polymerase.

6th stage:

Insertion of 3' LacZ-polyA SV 40 in the p4 plasmid

The 3' end of the coding phase of β -galactosidase, followed by the signal of polyadenylation of the virus SV 40 is isolated from the p5 plasmid through sites Xho I-Aat II and cloned in the p4 plasmid through the same sites.

7th stage:

Insertion of 5' LacZ in the KS-vector

The 5' end of the coding phase of β -galactosidase is isolated from plasmid PMC 1871 (Pharmacia) through the sites Pst I-Sac I and cloned in the vector KS- (Stratagene) through the same sites.

8th stage:

Fusion of a genomic sequence HOXC8 (Hox-3.1) with 5' LacZ

09 NOV 1996

21

A genomic sequence of the gene HOXC8 (Hox-3.1), cloned in the vector KS-, is purified by successive digesting with the enzyme Sac I, then with the Mung bean nuclease and finally with the enzyme Apa I. This insert is fused with the 5' end of the coding phase of β -galactosidase by cloning in the p7 plasmid digested by Apa I-Sma I. The protein thus fused contains the initiation codon of the translation of the gene HOXC8 (Hox-3.1) followed by the coding phase of β -galactosidase (then verified by sequencing).

Met Ser Ser	Ile Pro Gly Asp Pro
CCAGC ATG AGC TCC	ATT CCC GGG GAT CCC
GGTCG TAC TCG AGG	TAA GGG CCC CTA GGG

↓ Sac I

CCAGC ATG AGC T
GGTCG TAC

↓ Sma I

↓ Mung bean nuclease

CCAGC ATG
GGTCG TAC

GGG GAT CCC
CCC CTG GGG

↔

Met Gly Asp Pro
CCAGC ATG GGG GAT CCC
GGTCG TAC CCC CTA GGG

9th stage:

Insertion of HOXC8 (Hox-3.1)-5' LacZ in the p6 plasmid

The fusion HOXC8 (Hox-3.1)-5' LacZ is isolated from the p8 plasmid through the sites Apa I-SacI and cloned in the p6 plasmid through the same sites. The effect of this cloning is to reconstitute the coding phase of β -galactosidase in its totality.

10th stage:

Insertion of the gene Neo^R in the vector KS+

09 NOV 1996

The gene of resistance to neomycin (bacterial promoter and coding phase of phosphotransferase) is isolated in the pRSV neo through the sites Hind III-Eco RI and cloned in the vector KS+ (Stratagene).

5 11th stage:

Mutagenesis of the initiation sequence of Neo^R in p10

0 The sequence of the initiation of translation of phosphotransferase is modified to be identical to the consensus sequence observed in vertebrates and thus to allow a higher level of initiation of translation and therefore an increased resistance to G418 for the cells of mammals. The modification also creates a Apa LI site which makes it possible to control the efficiency of
5 the mutagenesis.

Apa LI

GTTTCGCATG

⇒

GTGCACCATG

1 An oligonucleotide (CTGTTC AATCATG GGTGCACGATCCTCA) comprising a region of mispairing with the sequence of pRSV neo (underlined) is synthesized (Gene Assembler, Pharmacia) then phosphorylated by the polynucleotide kinase of the T4 bacteriophage. A single strand template of the p10 plasmid is prepared with the fl origin of the KS+ plasmid and hybridized with the mutagenesis oligonucleotide. The second strand is synthesized and repaired by the Klenow polymerase and the DNA ligase of the T4 bacteriophage. After transformation of bacteria, the mutated clones are screened with the aid of the oligonucleotide marked with ³²P. The mutagenesis was verified by digesting with Apa LI and also by sequencing.

12th stage:

09 NOV 1996

Replacement of the initiation sequence in the p9 plasmid

A fragment containing the modified sequence of initiation of translation of the gene of resistance to neomycin is isolated from the p11 plasmid through the enzymes Hind III-Eag I and cloned in the p9 plasmid through the same sites.

13th stage:

Insertion of the multiple cloning site in the p12 plasmid

Two complementary oligonucleotides are synthesized (Gene Assembler, Pharmacia) then phosphorylated. After pairing, the MCS is cloned in the sites Apa I-Sac II of the p12 plasmid through its sticky ends.

XmaI Asp 718 XmnI
 ApaI SmaI KpnI XbaI NsiI SacII
 5' CCCC GGGGGTACCTCTAGAAATGCATTCCGC 3'
 3' CCGGGGGGCCCCCATGGAGATCTTACGTAAGG 5'

The multiple cloning site was also verified by sequencing.

II - Addition of flanking sequences at the extremities of the linearized pGN plasmid upstream from LacZ' by an M.C.S. site

The flanking sequences used were chosen in relation to the desired site of insertion (for example, HOXC8 (Hox-3.1), see Fig. 2a and 2b pGMA and pGMD).

In the construction of the mutagenesis plasmid pGMD, two fragments of DNA homologous to the locus HOXC8 (Hox-3.1) were cloned at the sites Apa I-Nsi I and Nsi I-Sac II of the pGN vector. The 5' fragment begins at the site Sac II (CCGCGG) at nucleotide 219 of the DNAC c21 of HOXC8 (Hox-3.1). This fragment extends over 6.8 kb in 5' up to the first BamHI site. The 3'

09 NOV 1996

fragment begins at the site Apa 1 (GGGCCC) at nucleotide 885 of the cDNA c21. This fragment extends over 1.5 kb in 3' up to the first PstI site. A NsiI linker was inserted in the BamHI site of the 5' fragment and in the PstI site of the 3' fragment. The 5' and 3' fragments were cloned in the pGN vector in the sites Nsi I-Sac II and Apa I-Nsi I, respectively. The sequence of the cDNA of HOXC8 (Hox-3.1) c21 has been published (Ref. 1).

0 The mutagenesis plasmid is linearized by digesting with Nsi I before electroporation of E.S. cells. Its extremities are formed by the two genomic fragments cloned at the sites Apa I-Nsi I and Nsi I-Sac II of the pGN vector.

5 The pGMD plasmid does not present any signal of polyadenylation after the gene of resistance but, on the other hand, presents a sequence rich in AU responsible for selective degradation of mRNA, inserted in the sequence of the intron of HOXC8 (Hox-3.1) of the plasmid.

1 Another mutagenesis plasmid, pGMA, presents the same structure as pGMD but contains the signals of polyadenylation and of termination of transcription of SV40 and does not present the AU sequence of degradation of mRNA downstream from the gene Neo^R. The aim of these modifications was to reduce the level of transcripts of Neo^R in clones resulting from random integration. On the other hand, clones resulting from homologous recombination events between pGMD and a locus HOXC8 (Hox-3.1) should have an unaltered growth during the selection at G418, the AT sequence of degradation of mRNA being eliminated by the process of recombination itself, or spliced with the intron HOXC8 (Hox-3.1).

In the experimental stages which follow, the protocol described by Thompson et al. 1989, was followed for the production of chimeric animals.

III - Transfection of embryonic mouse cells

5 The method described by Thompson et al. 1989, was used to transfect embryonic mouse cells. Utilization of the technique of electroporation ensures the introduction of a single copy of the foreign gene (LacZ) per cell. After transfection, several clones
0 expressing β -galactosidase were isolated.

The mutagenesis plasmids pGMD and pGMA were linearized and introduced by electroporation into E.S. cells in order to favor the insertion of only one copy in the genome (Ref. 11).

5 The initial transfections were carried out to compare the targeting efficiency of the HOXC8 (Hox-3.1) of the plasmids pGMA and pGMD (see table I).

Homologous recombination in the gene HOXC8

0 (Hox-3.1)

Exp.	Plasmid of muta- genesis	No. of group analysed	No. clones forming the group	No. of positive PCR results
I	pGMA	3	600	0(2)
II	pGMD	5	250	3(5)
III	pGMD	84	2-3	5(5)

Table I

The E.S. cell line "C.C.E." (Ref. 16) was maintained in a continuous manner on fibroblast feeding layers (Ref. 17). For experiments I and II, 1.5×10^7 E.S. cells in 1.5 ml HeBS were electroporated (Ref. 11) at 200 V, with 40mg of linearized plasmid, and then

09 NOV 1996

spread over four culture trays (diameter 100 mm). For experiment III, the shock was carried out under the same conditions but a quarter of the cells were spread over four 24-well plates. The next day, $250 \mu\text{g} \cdot \text{ml}^{-1}$ of G418 were added. Each transfection produced about 2400 clones with pGMA and about 1000 clones with pGND.

The average number of clones of E.S. cells resistant to G418 in each group is indicated in table I, as well as the number of groups giving a positive result with the P.C.R. technique. A positive result means that a band of 1.6 kb could be observed on an agarose gel coloured with ethidium bromide (see Fig. 7). The number of groups giving a positive signal after a Southern analysis of the P.C.R. mixture and hybridization of a specific probe which does not contain the primer sequences is indicated between brackets (Fig. 8).

Detection of homologous recombination with P.C.R.

P.C.R. was carried out on 10^5 cells of a group of 250 clones of transfection II (see lane D of Figure 7). In the other lanes, four groups of transfection III were analysed together by mixing about 4×5000 cells. The primers 07 and 08 used in the P.C.R. surround the sequence 3'HOXC8 (Hox-3.1) of the mutagenesis plasmid (Fig. 2). The fragment of 1.6 Kb covering this 3' sequence cannot be amplified except in the case of homologous recombination. Lanes 2, 3 and D illustrate these positive results.

The DNA of E.S. clones was prepared at the moment of replica on filter by utilizing the method "boiling-proteinase K digestion boiling" (Ref. 18). 40 amplification cycles (40 seconds at 94°C , 1 minute at 60°C , 7 minutes at 72°C) were carried out in a reaction mixture of $100 \mu\text{l}$, containing 67 mM Tris-HCl (pH 8.6), 16.7 mM $(\text{NH}_4)_2\text{SO}_4$, 6.7 mM MgCl_2 , 10 mM 2-mercaptoethanol, 0.01% (p/v) gelatine, 200 μM dATP,

09 NOV 1996

dTTP and dCTP, 100 μ M dGTP, 100 μ M 7-deaza dGT, 600 ng of each primer (07: AACTTCCCTCTCTGCTATTC and 08: CAGCAGAAACATACAAGCTG) and 3U polymerase Taq (Perkin Elmer Cetus), covered by 100 μ l paraffin. Half of the reaction mixture was applied on an agarose gel 0.7% coloured with ethidium bromide. The main marker is an Eco RI + Hind III digest of lambda DNA.

Southern analyses

Three independent clones of E.S. cells containing the mutated HOXC8 (Hox-3.1) (identified by P.C.R.) were isolated from positive groups using pipettes. Their DNA was examined by Southern analysis after digestion with the restriction enzymes indicated in Figure 8, so as to confirm the specific targeting and to make the distinction between recombined and wild-type loci. Two different probes were used in the analysis of the 3' end of loci HOXC8 (Hox-3.1) in the mutated clones and in the non-mutated E.S. used as controls (Figure 8c). The first probe "a" was contained in the HOXC8 (Hox-3.1) sequences of the mutagenesis plasmid and showed the number of vector integrations and their physical links. One of the three recombined clones also contained a copy of the plasmid integrated at random (Fig. 8a, clone F2). The second "b" probe which was not in the mutagenesis vector made the distinction between the recombined and wild-type alleles HOXC8 (Hox-3.1) (Fig. 8b). The recombined locus HOXC8 (Hox-3.1) presented, with the two probes, the image of hybridization expected from the restriction maps of the mutagenesis vector and of the intact locus. In addition, the existence of two fields of recombination in the 3' fragment of the vector was confirmed by the presence or the absence of the AT sequence in the recombined locus HOXC8 (Hox-3.1), (for example Fig. 8, clone L5). The 5' end of the locus HOXC8 (Hox-3.1) was also analyzed for a homologous recombination event.

09 NOV 1996

Restriction enzymes not presenting sites in the sequence HOXC8 (Hoc-3.1) 5' of 6.8 kb of the mutagenesis vector were utilized in the digestion of the DNAs of recombined clones. These DNAs were then submitted to electrophoresis in a pulsed field to differentiate the fragments of high molecular weight. A Southern analysis of this gel also indicated the correctly recombined alleles and the wild-type HOXC8 (Hoc-3.1) alleles, using a sensor presenting a sequence upstream from the mutagenesis plasmid.

Southern analyses have shown that an allele of the gene HOXC8 (Hoc-3.1) was recombined as predicted. The homologous recombination was equivalent to a double "crossing-over" between the genomic arms of the mutagenesis plasmid and the homologous chromosomal sequences (Fig. 2).

In the recombinant clones, the gene LacZ was placed under the control of promoter and regulatory sequences of HOXC8 (Hox-3.1) upstream from the AUG codon, but the 3' signals of maturation of mRNA came from the SV 40. In these recombined clones, the expression of LacZ was not detectable by colouring with β -Gal, which is coherent with the absence of transcription of HOXC8 (Hoc-3.1) in E.S. cells determined by analysis of RNase protection. The activity of β -Gal could be induced in certain cells after 3 or 4 days of culture in the presence of $5 \cdot 10^{-7}M$ retinoic acid, conditions known for inducing the transcription of HOXC8 (Hoc-3.1) (Ref. 19).

By using the mutagenesis vector pGMA, which presents a total homology of 8.3 Kb DNA with the locus HOXC8 (Hoc-3.1), a fragment of 120 pb was replaced by an insertion of 7.2 Kb. The frequency of this targeted replacement (1/900) is comparable to that obtained recently (1/1000) with HPRT (Ref. 13) or with En-2

09 NOV 1996

(1/260) (Ref. 20), the heterologous fragment inserted being nonetheless in these latter cases of a much less significant size (1.1 and 1.5 Kb respectively). In a surprising manner, it was noted that a very high homologous recombination frequency (1/40) could be obtained with the pGMD vector. The elimination of 3' signals of maturation of mRNA and the addition of the sequence of degradation of mRNA to the gene of resistance to neomycin had the effect of reducing the total number of clones resistant to G418 by 2.4 (table I). The specific targeting result was almost ten times higher (900/40). The mechanism of homologous recombination was probably affected in the experiments with pGMD. A possible explanation for these results could be that an AT sequence of 51 pb could provide, in vivo, an open loop in the mutagenesis plasmid because of its lower fusion temperature. If the HOXC8 (Hoc-3.1) sequences close to the pGMD can be influenced by this opening, on each side of the AT region, they could react in a more efficient manner, in the single-strand state, with the chromosomal locus HOXC8 (Hoc-3.1). The model of mitotic recombination in yeast suggests that it could be initiated by such an exchange of strands, even though the mechanism of homologous recombination remains unknown in the more complex eucaryotes.

Figure 8 shows the results of Southern analysis carried out on positive individual clones (L5 and F2) and E.S. cells (C.C.E.).

The probes used only hybridize with HOXC8 (Hoc-3.1) sequences included in the vector (a) or excluded from the mutagenesis vector (b). The image of hybridization of the recombined locus HOXC8 (Hoc-3.1) (open triangles) is clearly distinguished from the wild-type locus (black triangles). The stars indicate the hybridization bands of a copy of the plasmid which

09 NOV 1996

was integrated at random. The size marker is a digest Eco RI + Hind III of Dna lambda.

Figure 8(c) shows the restriction maps of recombined (rec.) and wild-type (wt) HOXC8 (Hoc-3.1) alleles. The parts of the mutagenesis vector and of the locus HOXC8 (Hoc-3.1) are indicated by the same symbols as those used in Figure 2. In this case, the sequence AT was integrated by homologous recombination. The vertical arrow indicates the 3' end of the mutagenesis plasmid. The localization of probes "a" and "b" used in the Southern analysis is also indicated. The abbreviations used in Figure 8 are the following: B, Bam HI; D, Dra I; E, Eco RI, H, Hind III; S, Sal I; X, Xho I.

IV - Production of chimeric embryos

A micro-injection was made in blastocysts with two recombinant E.S. clones containing an intact allele HOXC8 (Hoc-3.1) and a recombined allele, these clones did not contain any other copy of the mutagenesis plasmid. The karyotypes of these cells were normal.

Ten to fifteen mutated cells were micro-injected per blastocyst. After reimplantation in surrogate mothers, the embryos were taken at 9.5, 10.5 and 12.5 days p.c. and analyzed for LacZ expression. The transcription template of HOXC8 (Hoc-3.1) at these stages had been determined beforehand by hybridization analysis *in situ* (Ref. 1). The HOXC8 (Hoc-3.1) transcripts are detectable for the first time at the stage of late gastrulation and are spread throughout all the tissues of the posterior part of the animal. Later, the distribution becomes progressively limited in space and specific to the tissue. At the stage of 12.5 days p.c. the transcription is localized in the cervical region of the neural tube, at the level of the heart. During embryogenesis, the repartition of the transcription of HOXC8 (Hoc-3.1) thus undergoes

modifications. The stage 10.5 days p.c. seems to be a period of transition, the transcription taking place both in the two rear regions and in the cervical neural tube.

5 In chimeric embryos of 9.5 and 10.5 days p.c., the caudal part to the posterior bud presented intense β -Gal activity, while the marker was never detected in the anterior thoracic region or the head (Fig. 9a). In the posterior region, the cells coloured by β -Gal were
 0 observed in all the tissues and all the embryonic layers. Between the two buds which form the members, coloured cells were distributed in the restricted zones, in the superficial ectoderm (Fig. 9b), as in the rear regions (Fig. 9c) and, in straight lines or rays,
 5 in the neural tube (Fig. 9b). These rays showed an irregular and asymmetric distribution on the wall of the neural tube. The transcription of HOXC8 (Hoc-3.1) was not detected in the thin layer of cells near the closure of the neural tube. Perhaps these cells did not
 10 resist the treatments applied during the hybridization in situ. It was observed that the cells of the neural ectoderm very soon become part of the different parts of the nervous system and migrate in a radial direction, following narrow lateral movements (Ref.
 5 21). These results are thus in line with this observation.

The expression of LacZ thus illustrated correctly the first part of the transcription of the homeogene HOXC8 (Hoc-3.1), that is to say in all the tissues of
 0 the caudal regions of embryos of 9.5 and 10.5 days p.c., and provided new information concerning the mode of transcription of HOXC8 (Hoc-3.1).

On the other hand, the expression of LacZ was not observed in the cervical regions of the neural tube of
 5 chimeric embryos of 12.5 days, nor in the anterior

09 NOV 1996

region of embryos of 10.5 days; this was not the result expected from studies of hybridization in situ. The ulterior phase of transcription of HOXC8 (Hoc-3.1) observed from day 10.5 in the very localized zones of the neural tube were not made evident by the activity of β -Gal. A possible explanation for this result could be that, even though the expression of LacZ was under the control of the promoter HOXC8 (Hoc-3.1), the 3' sequences of HOXC8 (Hoc-3.1) are absent in the reporter gene. It is possible that the 3' sequences of the AUG initiation codon of HOXC8 (Hoc-3.1) have an influence on the late expression of HOXC8 (Hoc-3.1) in the anterior domain. An effect of "gene dosing" could also explain this result. Auto-activation of several homeogenes of *Drosophila* was demonstrated genetically or suggested by the formation of complexes between the DNA and the proteins of the homeoboxes.

If the late component of transcription of HOXC8 (Hoc-3.1) in the neural tube is maintained by a similar mechanism, the inactivation of an allele could have a dominating effect in the cells of the neural ectoderm. Since one allele only would produce the protein HOXC8 (Hoc-3.1), the signal of activation would be diluted on the two promoters. The reduction of auto-inactivation in the two loci could thus lead to a total interruption of the transcription initiation. This could explain why no expression of LacZ was detected in the cervical region of the neural tube of embryos of 10.5 and 12.5 days.

V - Passage of the modification into the germ cell line: production of transgenic animals

The effects in F₁ and in F₂ of the modification brought by targeted insertion were observed after reproduction of the chimeric animals. Passage of the modification into the germ cell line was observed.

09 NOV 1996

BIBLIOGRAPHY

1. Le Mouellic, H., Condamine, H. et Brûlet, P. (1988), *Genes & Development*. 2, 125-135.
2. Cavenir, D.R. (1987). *Nucleic Acid Res.* 15, 1353-1361.
3. Sanes, J.R. Rubenstein, J.L.R. and Nicolas, J.F. (1986) *EMBO J.* 5, 3133-3142.
4. Nolan, G.P., Fiering, S., Nicolas, J.F. and Hertenberg, L.A. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2603-2607.
5. Gorman, C., Padmanabhan, R. and Howard, B.H. (1983). *Science*. 221, 661-553.
6. Gormann, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6777-6781.
7. Southern, P.J. and Berg, P. (1982) *J. Mol. Appl. Genet.* 1, 327-341.
8. Herbomel, P., Bourachot, B. and Yaniv, M. (1984). *Cell*. 39, 653-662.
9. Robertson, E.J. (1987). *Teratocarcinomas and embryonic stem cells: A practical approach*. IRL Press, Oxford.
10. Khan, A. (1988). *Medecine/Sciences* 4, 515-518.
11. G. Chu, Hayakawa, H., Berg, P. (1987). *Nucleic acid Research* 15, No. 3, 1311-1326.
12. Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., Melton, D.W., (1989) *Cell*, 56, 313-321.
13. Thomas, K.R., Capecchi, M.R., (1987) *Cell*, 51 503-512.
14. Evans, M.J., Kaufmann, M.H. (1981) *Nature*, 292, 154-155.
15. Robertson, E.J., (1986) *Trends in Genetics*, 9-13.

09 NOV 1996

16. Robertson, E., Bradley, A., Kuehn, M. & Evans, M., Nature 323, 445-448 (1986).

17. Robertson, E.J. in Teratocarcinomas and Embryonic Stem Cells (ed. Robertson, E.J.) *71-112
5 (IRL, Oxford, 1987).

18. Kim, H.S. & Smithies, O. Nucleic Acids. Res. 16, 8887-8903 (1988).

19. Breier, G., Bucan, M., Francke, U., Colberg-Poley, A.M. & Gruss, P. EMBO J.5, 2209-2215 (1986).

) 20. Joyner, A.L., Skarnes, W.C. & Rossant, J. Nature 338, 153-156 (1989).

21. McKay, R. D. G. Cell 58, 815-821 (1989).



09 NOV 1996

THIS SHEET HAS BEEN
REPLACEDCLAIMS

1. Transgenic or chimeric animal comprising cells whose genome comprises:

- a heterologous DNA, called insertion DNA, and being constituted by the totality of a gene or by a part of a gene;

- the insertion DNA being positioned in the genome in relationship to an endogenous gene, called the receiver gene, comprising a complementary DNA, in such a way that the insertion DNA or the complementary DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

2. Animal according to claim 1, characterized in that the insertion DNA contains either a coding sequence, whose expression takes place under the control of regulatory sequences of the receiver gene, or a regulatory sequence permitting the expression of a coding sequence of the receiver gene.

3. Animal according to claim 2, characterized in that the insertion DNA is constituted by a coding sequence other than the coding sequence of a gene coding for a selection agent.

4. Animal according to claim 1, characterized in that the insertion DNA comprises the totality of a gene.

5. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is a gene which normally is not transcribed in the cell.

6. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is inactivated following the insertion of the insertion DNA, the expression of the insertion DNA substituting that of the receiver gene.





THIS SHEET HAS BEEN
REPLACED

REPLACEMENT SHEET

09 NOV 1996

36

7. Animal as claimed in any one of the preceding claims characterized in that the insertion DNA replaces a part of the receiver gene.

8. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is, apart from the presence of the insertion DNA, in its natural genetic environment.

9. Animal as claimed in any one of the preceding claims characterized in that the genome contains, adjacent to the insertion DNA, a so-called "intercalated" sequence comprising for example a sequence coding for a selection agent with a promoter authorizing the expression of the agent of selection, or a marker.

10. Transgenic animal in which a single copy of a gene which is present in the genome in at least two copies, has been inactivated by the insertion of a gene which is different from the inactivated gene, the insertion gene being inserted in a position which permits the expression of this gene under the control of regulatory sequences of the inactivated endogenous gene.

11. Animal as claimed in any one of the preceding claims utilized for a genetic marking.

12. Process for screening pharmaceutical products presumed to have an activity with regard to expression products of a pathological gene linked to a disease, characterized by the administration of the product to a transgenic animal according to any one of the claims 1 to 11, the insertion DNA being constituted by the pathological gene or a fragment of it.

13. Nucleic acid obtained from the homologous recombination between an insertion DNA and a receiver gene, said nucleic acid comprising:

This Sheet Has Been Replaced





09 NOV 1996

THIS SHEET HAS BEEN
REPLACED

- an insertion DNA, heterologous with respect to the receiver gene, the insertion DNA being constituted by the totality of a gene or by a part of a gene;

- a receiver gene;

- the insertion DNA being positioned in relationship to the receiver gene comprising a complementary DNA in such a way that the insertion DNA or the complementary DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

14. Cell of a transgenic animal as claimed in any one of claims 1 to 10.

15. Cell line issuing from a transgenic animal as claimed in any one of claims 1 to 10.

This Sheet Has Been Replaced



REPLACEMENT SHEET

09 NOV 1996

THIS SHEET HAS BEEN
REPLACEDABSTRACT OF THE DISCLOSURETRANSGENIC OR CHIMERIC ANIMAL AND PROCESS FOR
SCREENING PHARMACEUTICAL PRODUCTS

The invention relates to a transgenic or chimeric animal carrying cells whose genome comprises:

- a heterologous DNA, called insertion DNA, and being constituted by the totality of a gene or by a part of a gene;

- the insertion DNA being positioned in the genome in relationship to an endogenous gene, called the receiver gene, comprising a complementary DNA, in such a way that the insertion DNA or the complementary DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA.

Figure 1

7-2283288

09 NOV 1996

1/12

pGN

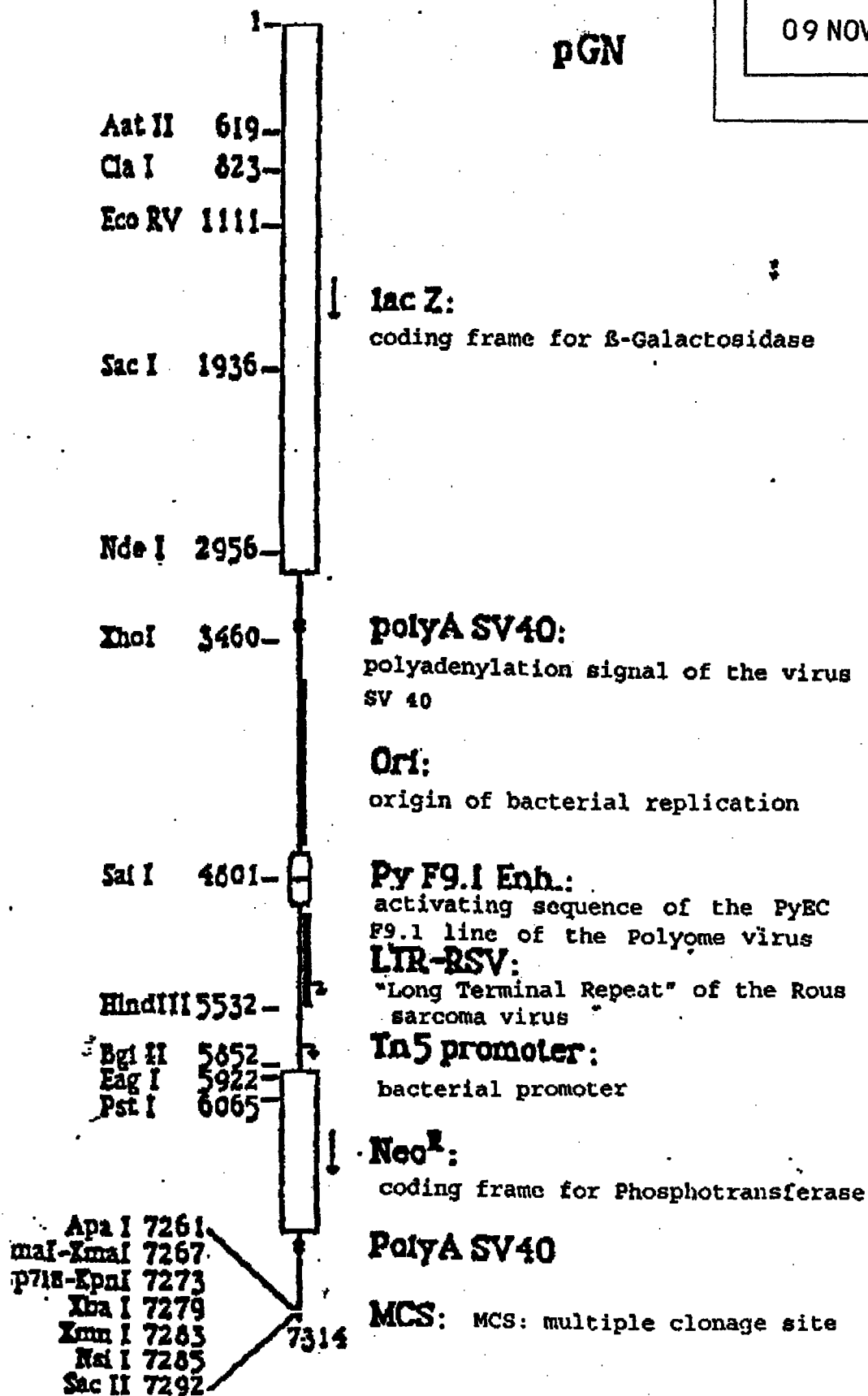


Figure 1: linear plasmid

2/12

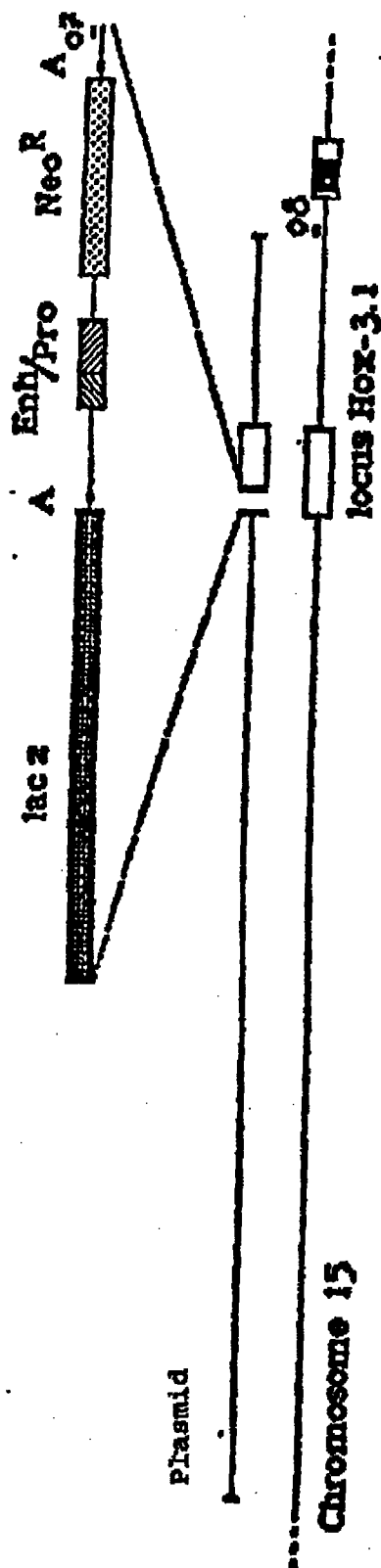


FIGURE 2a

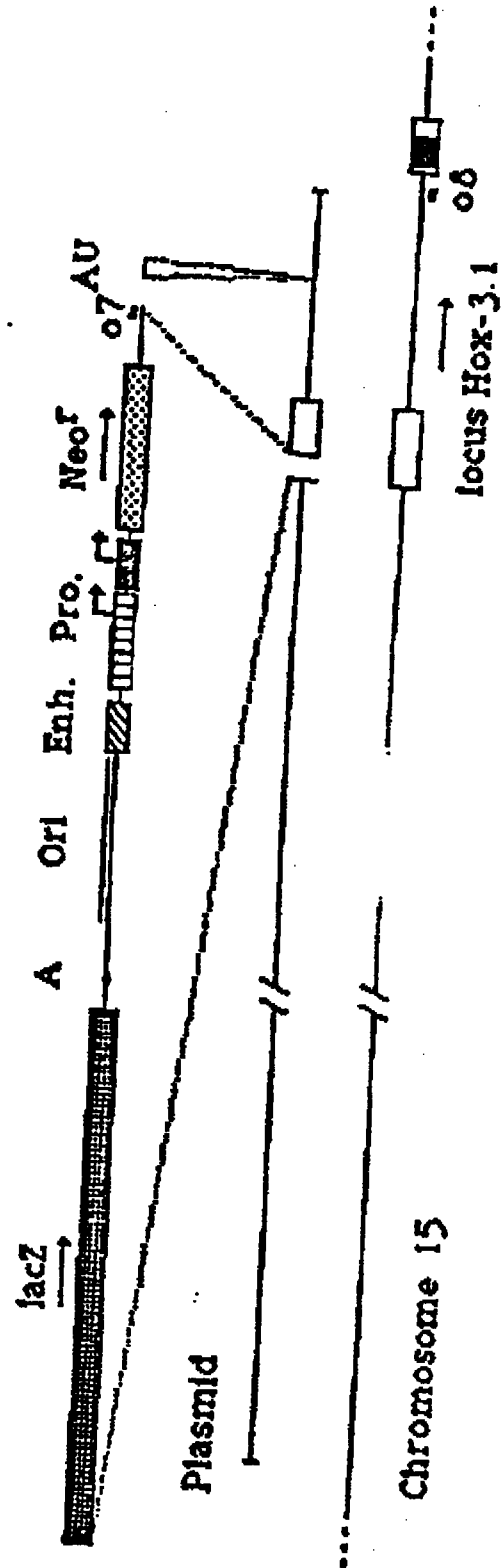


FIGURE 2b

4/12

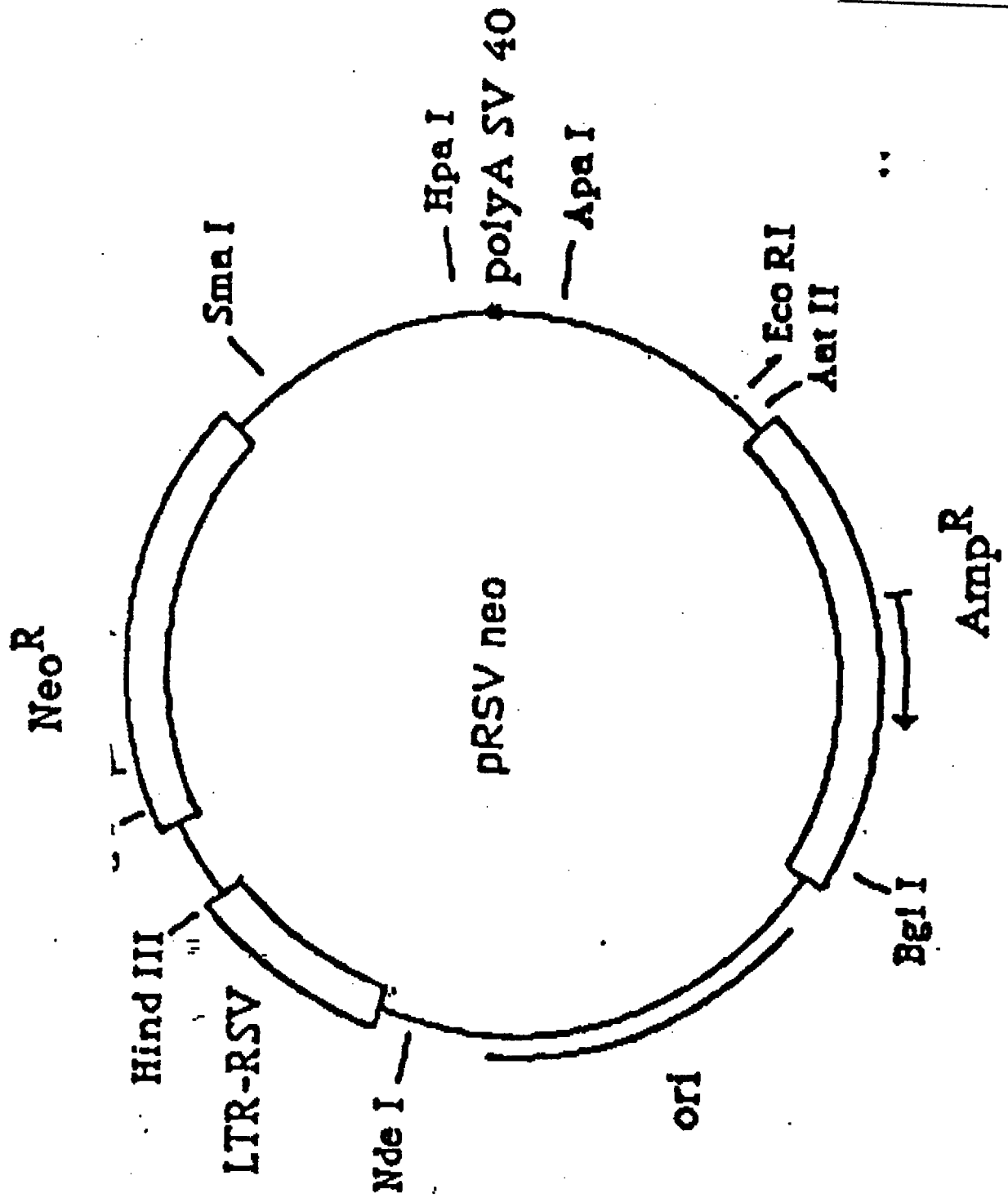
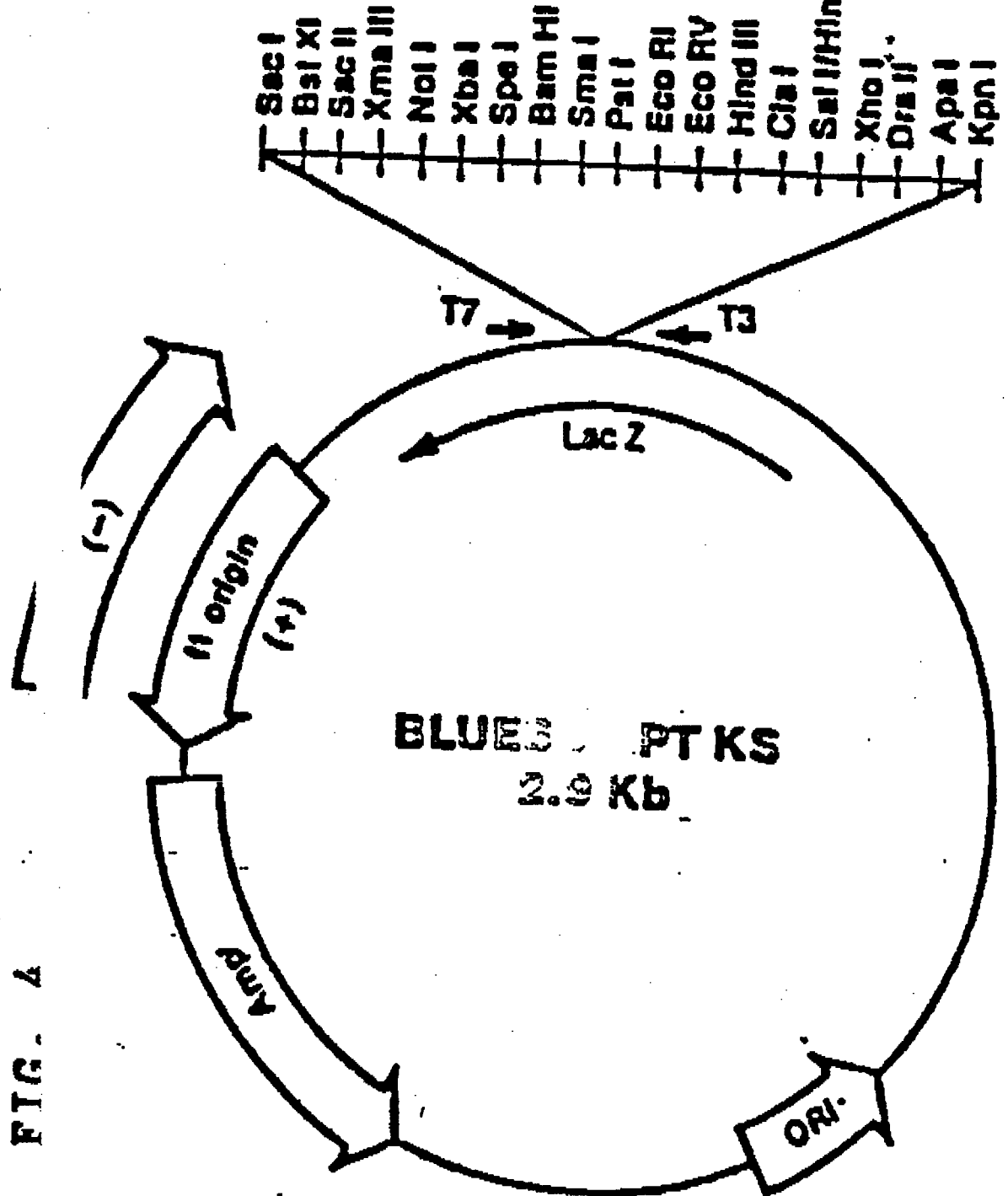


FIG. 3

5/12



09 NOV 1996

6/12

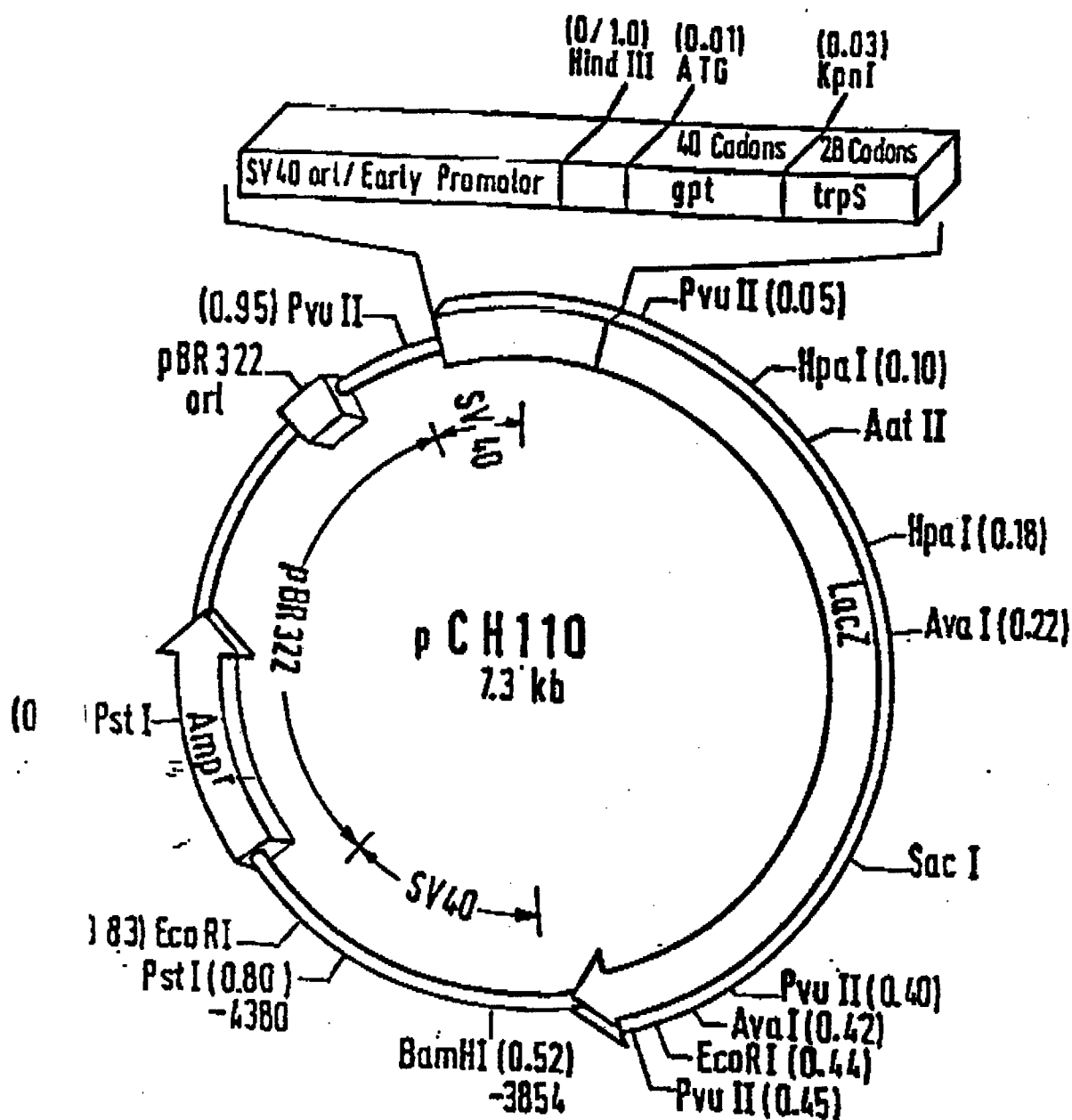
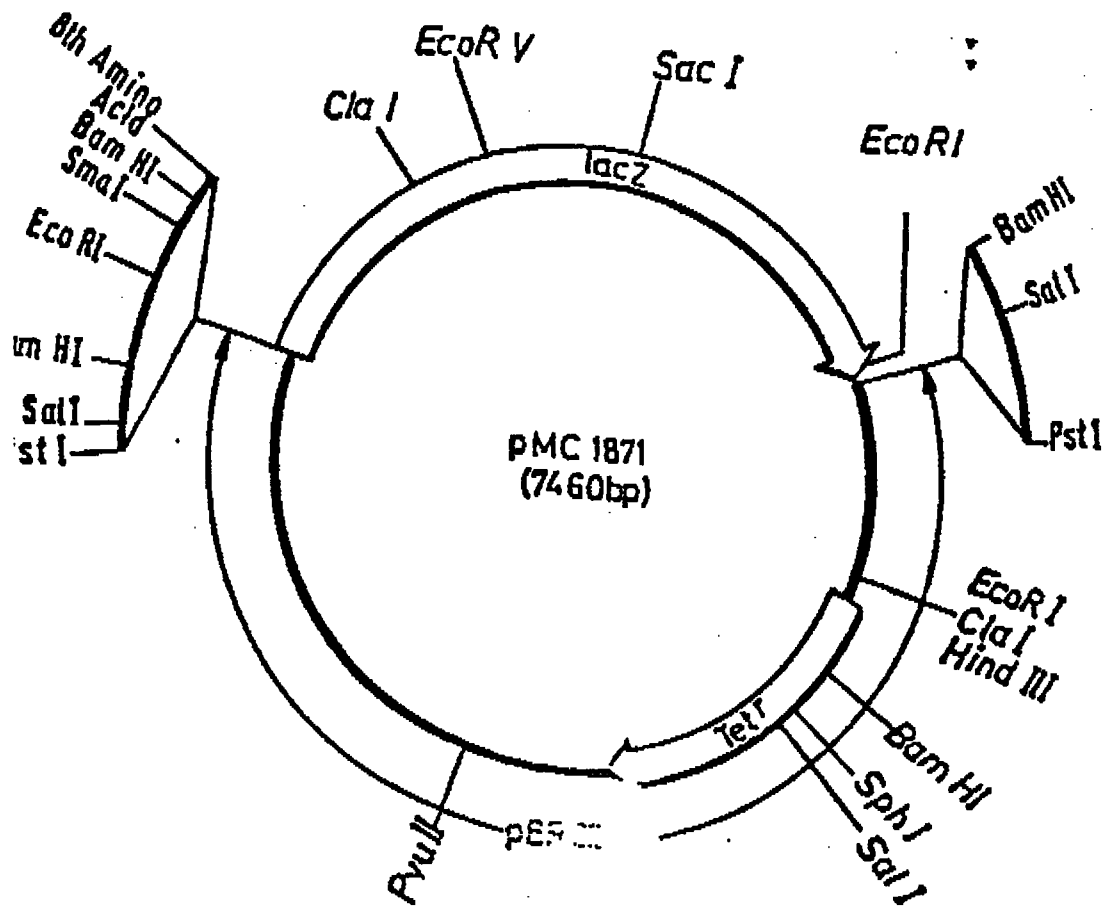


FIG. 5

7/12



	Pst I	Sal I	Bam HI	Eco RI	Sma I
18322	CT	GCA GGT CGACGG	ATC CGG GGAATT	CCC	
	8	9	1021		
	Bam HI	pro val	lys stop stop stop		
G	GAT CCC GTC	lac Z	AAA TAA TAA TAA CCG GGC		
3	GGG GAT CCG TCG ACC TGC AG	pBR322			
	Bam HI	Sal I	Pst I		

FIG.6

09 NOV 1996

8/12

1 2 3 4 5 6 7 8 9 10 M D :

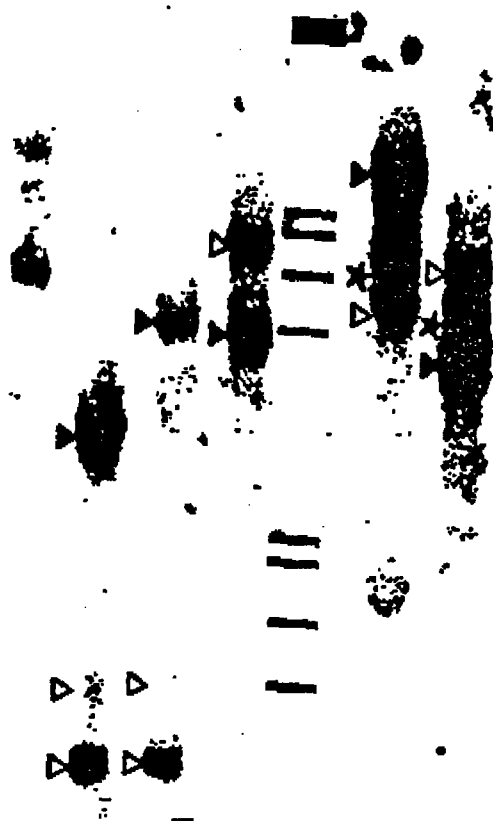


FIG. 7

9/12

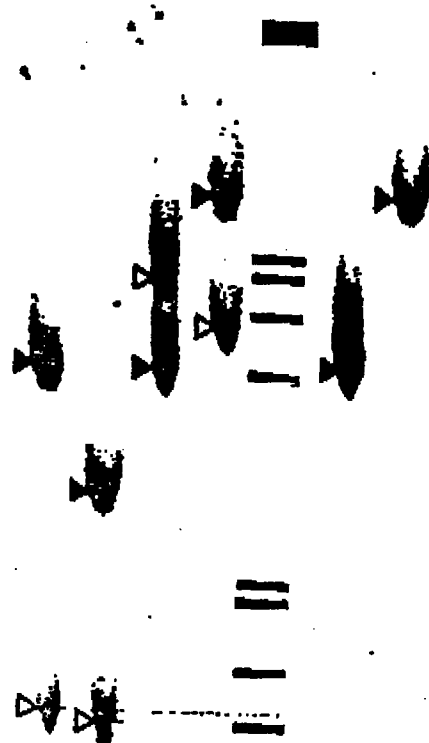
a

$\overline{\text{L5}}$ λ $\overline{\text{F2}}$
 3+H D+E D E B+H E



b

$\overline{\text{L5}}$ λ $\overline{\text{ES}}$
 D D+E E B+H E B+H



c

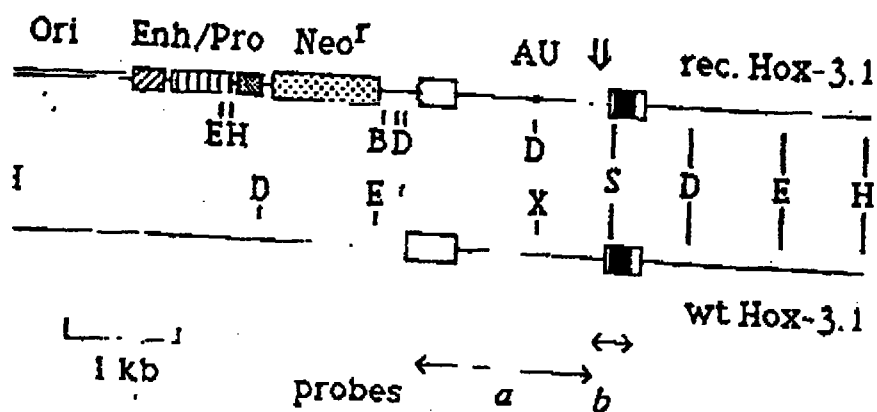


FIG. 8

09 NOV 1996

10/12



Fig. 9A

09 NOV 1996

11/12



Fig. 98



Fig. 9C

1000X

7-1868600



16 JUN 1998

CLAIMS

1. Transgenic or chimeric animal comprising cells whose genome comprises :

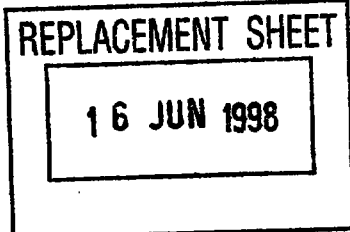
- a heterologous DNA, called insertion DNA, and being constituted by the totality of a gene or by a part of a gene;

- the insertion DNA being positioned in the genome in relationship to an endogenous gene, called the receiver gene, comprising a complementary DNA, in such a way that the insertion DNA or the complementary DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA;

- provided that, when the insertion DNA is constituted by a part of a gene, it is other than a part of a gene conferring a selectable phenotype when associated with the complementary DNA.

2. Animal according to claim 1, characterized in that the insertion DNA contains either a coding sequence, whose expression takes place under the control of regulatory sequences of the receiver gene, or a regulatory sequence permitting the expression of a coding sequence of the receiver gene.

3. Animal according to claim 2, characterized in that the insertion DNA is constituted by a coding sequence other than the coding sequence of a gene coding for a selection agent.



4. Animal according to claim 1, characterized in that the insertion DNA comprises the totality of a gene.

5. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is a gene which normally is not transcribed in the cell.

6. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is inactivated following the insertion of the insertion DNA, the expression of the insertion DNA substituting that of the receiver gene.

7. Animal as claimed in any one of the preceding claims characterized in that the insertion DNA replaces a part of the receiver gene.

8. Animal as claimed in any one of the preceding claims characterized in that the receiver gene is, apart from the presence of the insertion DNA, in its natural genetic environment.

9. Animal as claimed in any one of the preceding claims characterized in that the genome contains, adjacent to the insertion DNA, a so-called « intercalated » sequence comprising for example a sequence coding for a selection agent with a promoter authorizing the expression of the agent of selection, or a marker.

10. Transgenic animal in which a single copy of a gene which is present in the genome in at least two copies, has been inactivated by the insertion of a gene which is different from the inactivated gene, the insertion gene being inserted in a position which



16 JUN 1998

permits the expression of this gene under the control of regulatory sequences of the inactivated endogenous gene.

11. Animal as claimed in any one of the preceding claims utilized for a genetic marking.

12. Process for screening pharmaceutical products presumed to have an activity with regard to expression products of a pathological gene linked to a disease, characterized by the administration of the product to a transgenic animal according to any one of the claims 1 to 11, the insertion DNA being constituted by the pathological gene or a fragment of it.

13. Nucleic acid obtained from the homologous recombination between an insertion DNA and a receiver gene, said nucleic acid comprising :

- an insertion DNA, heterologous with respect to the receiver gene, the insertion DNA being constituted by the totality of a gene or by a part of a gene;

- a receiver gene;

- the insertion DNA being positioned in relationship to the receiver gene comprising a complementary DNA in such a way that the insertion DNA or the complementary DNA is rendered functional, or its functioning is rendered more efficient due to the association of the insertion DNA and the complementary DNA;

- provided that, when the insertion DNA is constituted by a part of a gene, it is other than a part of a gene conferring a selectable phenotype when associated with the complementary DNA.



REPLACEMENT SHEET

16 JUN 1998

38

14. Cell of a transgenic animal as claimed in any one of claims 1 to 10.

15. Cell line issuing from a transgenic animal as claimed in any one of claims 1 to 10.
