



(86) **Date de dépôt PCT/PCT Filing Date:** 2003/10/08
(87) **Date publication PCT/PCT Publication Date:** 2004/04/22
(45) **Date de délivrance/Issue Date:** 2016/01/26
(85) **Entrée phase nationale/National Entry:** 2005/04/06
(86) **N° demande PCT/PCT Application No.:** FR 2003/002962
(87) **N° publication PCT/PCT Publication No.:** 2004/033689
(30) **Priorités/Priorities:** 2002/10/08 (FR0212489);
2003/02/27 (FR0302439); 2003/08/07 (US60/493,490)

(51) **Cl.Int./Int.Cl. C12N 15/52** (2006.01),
C07K 14/36 (2006.01), **C12N 1/21** (2006.01),
C12N 15/63 (2006.01), **C12N 9/00** (2006.01),
C12P 19/62 (2006.01)
(72) **Inventeurs/Inventors:**
BLONDELET-ROUAULT, MARIE-HELENE, FR;
DOMINGUEZ, HELENE, FR;
DARBON-RONGERE, EMMANUELLE, FR;
GERBAUD, CLAUDE, FR;
GONDRAN, ANNE, FR;
...
(73) **Propriétaires/Owners:**
AVENTIS PHARMA S.A., FR;

(54) **Titre : POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES SPIRAMYCINES, SEQUENCES NUCLEOTIDIQUES
CODANT CES POLYPEPTIDES ET LEURS APPLICATIONS**
(54) **Title: POLYPEPTIDES INVOLVED IN SPIRAMYCIN BIOSYNTHESIS, NUCLEOTIDE SEQUENCES ENCODING SAID
POLYPEPTIDES AND USES THEREOF**

(57) **Abrégé/Abstract:**

La présente invention concerne l'isolement et l'identification de nouveaux gènes de la voie de biosynthèse des spiramycines et de nouveaux polypeptides impliqués dans cette biosynthèse. L'invention est également relative à un procédé de production de ces polypeptides. Elle est également relative à l'utilisation de ces gènes dans le but d'augmenter les taux de production et la pureté de la spiramycine produite. L'invention concerne notamment un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III et l'utilisation d'un tel microorganisme. L'invention a également trait à l'utilisation des gènes de la voie de biosynthèse des spiramycines pour la construction de mutants pouvant conduire à la synthèse de nouveaux antibiotiques ou à des formes dérivées de spiramycines. L'invention concerne encore les molécules produites grâce à l'expression de ces gènes et des compositions pharmacologiquement actives d'une molécule produite grâce à l'expression de tels gènes _____

(72) **Inventeurs(suite)/Inventors(continued):** KARRAY, FATMA, FR; LACROIX, PATRICIA, FR;
OESTREICHER-MERMET- BOUVIER, NATHALIE, FR; PERNODET, JEAN-LUC, FR; TUPHILE, KARINE, FR

(73) **Propriétaires(suite)/Owners(continued):**C.N.R.S., FR

(74) **Agent:** BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international(43) Date de la publication internationale
22 avril 2004 (22.04.2004)

PCT

(10) Numéro de publication internationale
WO 2004/033689 A3(51) Classification internationale des brevets⁷ :**C12N 15/52**, 9/00, 15/63, 1/21, C07K 14/36,
C12P 19/62 // (C12P 19/62, C12R 1:465)

(21) Numéro de la demande internationale :

PCT/FR2003/002962

(22) Date de dépôt international : 8 octobre 2003 (08.10.2003)

(25) Langue de dépôt :

français

(26) Langue de publication :

français

(30) Données relatives à la priorité :

0212489	8 octobre 2002 (08.10.2002)	FR
0302439	27 février 2003 (27.02.2003)	FR
60/493,490	7 août 2003 (07.08.2003)	US

(71) **Déposants** (*pour tous les États désignés sauf US*) : **AVENTIS PHARMA S.A.** [FR/FR]; 20 avenue Raymond Aron, F-92160 ANTONY (FR). **CNRS** [FR/FR]; 3 rue Michel Ange, F-75794 Cedex 16 PARIS (FR).(72) **Inventeurs; et**(75) **Inventeurs/Déposants** (*pour US seulement*) : **BLONDELET-ROUAULT, Marie-Hélène** [FR/FR]; 26 avenue de l'Esterel, F-91400 ORSAY (FR). **DOMINGUEZ, Hélène** [FR/FR]; 74 rue Guillaume Puy, F-84000 AVIGNON (FR). **DARBON-RONGERE, Emmanuelle** [FR/FR]; 6 rue du Grand Parc, F-78960 VOISINS LEBRETONNEUX (FR). **GERBAUD, Claude** [FR/FR]; 3 allée Dunand, F-93220 GAGNY (FR). **GONDRAN, Anne** [FR/FR]; 45 boulevard Saint Jacques, F-75014 PARIS (FR). **KARRAY, Fatma** [TN/FR]; Résidence Universitaire les Rives de l'Yvette, Bâtiment 23, ch. 2008, F-91440 BURES SUR YVETTE (FR). **LACROIX, Patricia** [FR/FR]; 28 allée de Bellevue, F-94170 LE PERREUX (FR). **OESTREICHER-MERMET-BOUVIER, Nathalie** [FR/FR]; 9 rue Louise Filliaux-Tiger, F-92140 CLAMART (FR). **PERNODET, Jean-Luc** [FR/FR]; 21 rue des Jardins, F-94230 CACHAN (FR). **TUPHILE, Karine** [FR/FR]; 18A rue de l'Oratoire, F-69300 CALUIRE (FR).(74) **Mandataire** : **BOUVET, Philippe**; AVENTIS PHARMA S.A., Direction brevet, 20 avenue Raymond Aron, F-92160 ANTONY (FR).(81) **États désignés (national)** : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.(84) **États désignés (régional)** : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK,

[Suite sur la page suivante]

(54) **Title:** POLYPEPTIDES INVOLVED IN SPIRAMYCIN BIOSYNTHESIS, NUCLEOTIDE SEQUENCES ENCODING SAID POLYPEPTIDES AND USES THEREOF(54) **Titre** : POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES SPIRAMYCINES, SEQUENCES NUCLEOTIDIQUES CODANT CES POLYPEPTIDES ET LEURS APPLICATIONS.(57) **Abstract:** The invention concerns isolation and identification of novel genes of spiramycin biosynthesis process and novel polypeptides involved in said biosynthesis. The invention also concerns a method for producing sand polypeptides. The invention further concerns the use of said genes for increasing the rate of production and purity in the resulting spiramycin. The invention concerns in particular a micro-organism producing spiramycin I but not producing spiramycin II and III and the use of such a micro-organism. The invention also concerns the use of genes in spiramycin biosynthesis process for constructing mutants capable of leading to synthesis of novel antibiotics or derivative forms of spiramycin. Finally the invention concerns molecules produced through expression of said genes and phamacologically active compositions of one molecule produced through expression of such genes.(57) **Abrégé** : La présente invention concerne l'isolement et l'identification de nouveaux gènes de la voie de biosynthèse des spiramycines et de nouveaux polypeptides impliqués dans cette biosynthèse. L'invention est également relative à un procédé de production de ces polypeptides. Elle est également relative à l'utilisation de ces gènes dans le but d'augmenter les taux de production et la pureté de la spiramycine produite. L'invention concerne notamment un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III et l'utilisation d'un tel micoorganisme. L'invention a également trait à l'utilisation des gènes de la voie de biosynthèse des spiramycines pour la construction de mutants pouvant conduire à la synthèse de nouveaux antibiotiques ou à des formes dérivées de spiramycines. L'invention concerne encore les molécules produites grâce à l'expression de ces gènes et des compositions pharmacologiquement actives d'une molécule produite grâce à l'expression de tels gènes_____

WO 2004/033689 A3

WO 2004/033689 A3



TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(88) Date de publication du rapport de recherche internationale:

26 août 2004

Publiée :

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

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

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __3__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __3__

NOTE: For additional volumes please contact the Canadian Patent Office.

POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES SPIRAMYCINES, SEQUENCES
NUCLEOTIDIQUES CODANT CES POLYPEPTIDES ET LEURS APPLICATIONS

5 La présente invention concerne l'isolement et l'identification de nouveaux gènes de la voie de biosynthèse des spiramycines et de nouveaux polypeptides impliqués dans cette biosynthèse. Elle est également relative à l'utilisation de ces gènes dans le but d'augmenter les taux de production et la pureté de la spiramycine produite.

10 L'invention a également trait à l'utilisation de ces gènes pour la construction de mutants pouvant conduire à la synthèse de nouveaux antibiotiques ou à des formes dérivées de spiramycines. L'invention concerne encore les molécules produites grâce à l'expression de ces gènes et enfin des compositions pharmacologiquement actives d'une molécule produite grâce à l'expression de tels gènes.

15 Les *Streptomyces* sont des bactéries filamenteuses Gram positif du sol. Elles jouent un rôle important dans la décomposition et la minéralisation des matières organiques grâce à la grande diversité des enzymes dégradatives qu'elles sécrètent. Elles présentent des phénomènes de différenciations morphologiques uniques chez les procaryotes s'accompagnant d'une différenciation métabolique caractérisée par la production de métabolites secondaires présentant une extraordinaire diversité de
20 structures chimiques et d'activités biologiques. Parmi ces métabolites, on retrouve les spiramycines produites à l'état naturel par la bactérie *Streptomyces ambofaciens*.

25 La spiramycine est un antibiotique de la famille des macrolides utile aussi bien en médecine vétérinaire qu'en médecine humaine. Les macrolides sont caractérisés par la présence d'un cycle lactone sur lequel sont greffés un ou plusieurs sucres. *Streptomyces ambofaciens* produit à l'état naturel la spiramycine I, II et III (cf. Figure 1), cependant l'activité antibiotique de la spiramycine I est nettement supérieure à celle des

spiramycines II et III (Liu *et al.*, 1999). La molécule de spiramycine I est constituée d'un macro-cycle lactonique, appelé platénolide et de trois sucres : la forosamine, le mycaminose et le mycarose (cf. Figure 1). L'activité antibiotique des spiramycines est due à une inhibition de la synthèse protéique chez les procaryotes par un mécanisme
5 impliquant la liaison de l'antibiotique au ribosome bactérien.

Certains composés membres de la famille des macrolides et possédant également un cycle lactone ont donné lieu à des utilisations hors du domaine des antibiotiques. Ainsi le produit FK506 a des effets immunosuppresseurs et offre des perspectives d'application thérapeutique dans le domaine de la transplantation d'organe, de l'arthrite
10 rhumatoïde et plus généralement dans des pathologies liées à des réactions autoimmunes. D'autres macrolides comme l'ivermectine ont des activités insecticides et anti-helminthiques.

De nombreuses voies de biosynthèse, concernant des antibiotiques appartenant à des classes variées ainsi que d'autres métabolites secondaires (pour une revue, Chater K.
15 1990), ont à ce jour déjà été étudiées chez les *Streptomycètes*. Cependant, la connaissance des voies de biosynthèse des spiramycines n'est que très partielle à ce jour.

La biosynthèse des spiramycines est un processus complexe comportant de nombreuses étapes et impliquant de nombreuses enzymes (Omura *et al.*, 1979a, Omura
20 *et al.*, 1979b). Les spiramycines appartiennent à la large classe des polyketides qui regroupe des molécules complexes particulièrement abondantes dans les microorganismes du sol. Ces molécules sont regroupées non pas par analogie de structure mais par une certaine similarité au niveau d'étapes de leur voie de biosynthèse. En effet, les polyketides sont produits par une série complexe de réactions mais possèdent en commun d'avoir dans leur voie de biosynthèse une série de réactions
25 catalysées par une ou des enzymes appelées « polyketides synthases » (PKS). Chez *Streptomyces ambofaciens*, la biosynthèse du macro-cycle lactonique des spiramycines (platénolide) est réalisée par une série de huit modules codés par cinq gènes PKS différents (Kuhstoss S., 1996, brevet US 5,945,320). Les spiramycines sont obtenues à

partir de ce cycle lactone. Cependant, les diverses étapes et enzymes impliquées dans la synthèse des sucres, ainsi que leur liaison au cycle lactone et la modification de ce cycle pour l'obtention des spiramycines restent encore inconnues à ce jour.

Le brevet US 5,514,544 décrit le clonage d'une séquence appelée *srmR* chez
5 *Streptomyces ambofaciens*. Il est fait l'hypothèse dans ce brevet que le gène *srmR* code une protéine régulatrice de la transcription des gènes impliqués dans la biosynthèse des macrolides.

En 1987, Richardson et collaborateurs (Richardson et al., 1987) ont montré que la
résistance à la spiramycine de *S. ambofaciens* est conférée par au moins trois gènes, ces
10 derniers ont été appelés *srmA*, *srmB* et *srmC*. Le brevet US 4,886,757 décrit plus particulièrement la caractérisation d'un fragment d'ADN de *S. ambofaciens* contenant le gène *srmC*. Cependant la séquence de ce gène n'a pas été divulguée. En 1990, Richardson et collaborateurs (Richardson et al., 1990) ont posé l'hypothèse de l'existence de trois gènes de biosynthèse de la spiramycine proche de *srmB*. Le brevet
15 US 5,098,837 rapporte le clonage de cinq gènes potentiellement impliqués dans la biosynthèse de la spiramycine. Ces gènes ont été baptisés *srmD*, *srmE*, *srmF*, *srmG* et *srmH*.

Une des difficultés majeures dans la production de composés tels que les spiramycines réside dans le fait que de très grands volumes de fermentation sont
20 nécessaires à la production d'une quantité relativement faible de produit. Il est donc souhaitable de pouvoir augmenter l'efficacité de production de telles molécules afin d'en diminuer les coûts de production.

La voie de biosynthèse des spiramycines est un processus complexe et il serait souhaitable d'identifier et de supprimer les réactions parasites qui pourraient exister au
25 cours de ce processus. Le but d'une telle manipulation est d'obtenir un antibiotique plus pur et/ou une amélioration de la productivité. A ce sujet, *Streptomyces ambofaciens* produit à l'état naturel la spiramycine I, II et III (cf. Figure 1), cependant l'activité antibiotique de la spiramycine I est nettement supérieure à celle des spiramycines II et III

(Liu *et al.*, 1999). Il serait donc souhaitable de pouvoir disposer de souches ne produisant que de la spiramycine I.

En raison de l'intérêt commercial des antibiotiques macrolides, l'obtention de nouveaux dérivés, notamment d'analogues de spiramycines ayant des propriétés
5 avantageuses est intensivement recherchée. Il serait souhaitable de pouvoir générer en quantité suffisante les intermédiaires de biosynthèse de la voie de biosynthèse des spiramycines ou des dérivés des spiramycines notamment dans le but de fabriquer des antibiotiques hybrides dérivés des spiramycines.

10 Description générale de l'invention

La présente invention résulte du clonage de gènes dont le produit intervient dans la biosynthèse des spiramycines. L'invention concerne tout d'abord de nouveaux gènes de la voie de biosynthèse des spiramycines et de nouveaux polypeptides impliqués dans cette biosynthèse.

15 Les gènes de la voie de biosynthèse et les séquences codantes associées ont été clonés et la séquence d'ADN de chacun a été déterminée. Les séquences codantes clonées seront désignées par la suite *orf1*c*, *orf2*c*, *orf3*c*, *orf4*c*, *orf5**, *orf6**, *orf7*c*, *orf8**, *orf9**, *orf10**, *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9c*, *orf10*, *orf11c*, *orf12*, *orf13c*, *orf14*, *orf15c*, *orf16*, *orf17*, *orf18*, *orf19*, *orf20*, *orf21c*, *orf22c*, *orf23c*,
20 *orf24c*, *orf25c*, *orf26*, *orf27*, *orf28c*, *orf29*, *orf30c*, *orf31*, *orf32c*, *orf33* et *orf34c*. La fonction des protéines codées par ces séquences dans la voie de biosynthèse des spiramycines est développée dans la discussion ci-après qui est illustrée par les figures 4, 5, 6 et 8.

1) Un premier objet de l'invention concerne un polynucléotide codant un
25 polypeptide impliqué dans la biosynthèse des spiramycines caractérisé en ce que la séquence dudit polynucléotide est :

(a) l'une des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149,

(b) l'une des séquences constituées par les variants des séquences (a),

5 (c) l'une des séquences dérivées des séquences (a) et (b) en raison de la dégénérescence du code génétique.

2) La présente invention a également pour objet un polynucléotide s'hybridant dans des conditions d'hybridation de forte stringence à au moins l'un des polynucléotides selon le paragraphe 1) ci-dessus.

10 3) L'invention concerne également un polynucléotide présentant au moins 70%, 80%, 85%, 90%, 95% ou 98% d'identité en nucléotides avec un polynucléotide comprenant au moins 10, 12, 15, 18, 20 à 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750,
15 1800, 1850 ou 1900 nucléotides consécutifs d'un polynucléotide selon le paragraphe 1) ci dessus.

4) L'invention concerne également un polynucléotide selon le prapagraphe 2) ou 3) ci-dessus caractérisé en ce qu'il est isolé à partir d'une bactérie du genre *Streptomyces*.

5) L'invention concerne également un polynucléotide selon le prapagraphe 2), 3) ou
20 4) ci-dessus caractérisé en ce qu'il code une protéine impliquée dans la biosynthèse d'un macrolide.

6) L'invention concerne également un polynucléotide selon le prapagraphe 2), 3) ou 4) ci-dessus caractérisé en ce qu'il code une protéine ayant une activité similaire à la protéine codée par le polynucléotide avec lequel il s'hybride ou il présente l'identité.

25 7) L'invention concerne également un polypeptide résultant de l'expression d'un polynucléotide selon le prapagraphe 1), 2), 3), 4), 5) ou 6) ci-dessus.

8) Un autre aspect de l'invention concerne un polypeptide impliqué dans la biosynthèse des spiramycines caractérisée en ce que la séquence dudit polypeptide est :

(a) l'une des séquences SEQ ID N° 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 27, 29, 31, 32, 33, 35, 37, 38, 39, 41, 42, 44, 46, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 108, 110, 112, 114, 116, 117, 119, 121, 142, 144, 146, 148 et 150,

(b) l'une des séquences telles que définies en (a) excepté que tout au long de ladite séquence un ou plusieurs acides aminés ont été substitués, insérés ou délévés sans en affecter les propriétés fonctionnelles,

10 (c) l'une des séquences constituées par les variants des séquences (a) et (b).

9) Un autre objet de l'invention concerne un polypeptide présentant au moins 70%, 80%, 85%, 90%, 95% ou 98% d'identité en acides aminés avec un polypeptide comprenant au moins 10, 15, 20, 30 à 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 420, 440, 460, 480, 500, 520, 15 540, 560, 580, 600, 620 ou 640 acides aminés consécutifs d'un polypeptide selon le paragraphe 8) ci-dessus.

10) Un autre aspect de l'invention concerne un polypeptide selon le paragraphe 9) ci-dessus caractérisé en ce qu'il est isolé à partir d'une bactérie du genre *Streptomyces*.

11) Un autre aspect de l'invention concerne un polypeptide selon le paragraphe 9) 20 ou 10) ci-dessus caractérisé en ce qu'il code une protéine impliquée dans la biosynthèse d'un macrolide.

12) Un autre aspect de l'invention concerne un polypeptide selon le paragraphe 9), 10) ou 11) ci-dessus caractérisé en ce qu'il a une activité similaire à celle du polypeptide avec lequel il partage l'identité.

25 13) Un autre aspect de l'invention concerne un ADN recombinant caractérisé en ce qu'il comprend au moins un polynucléotide selon l'un des paragraphes 1), 2), 3), 4), 5) ou 6) ci-dessus.

14) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 13) ci-dessus caractérisé en ce que le dit ADN recombinant est compris dans un vecteur.

5 15) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 14) ci-dessus caractérisé en ce que ledit vecteur est choisi parmi les bactériophages, les plasmides, les phagemides, les vecteurs intégratifs, les fosmides, les cosmides, les vecteurs navettes, les BAC ou les PAC.

16) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 15) ci-dessus caractérisé en ce qu'il est choisi parmi pOS49.1, pOS49.11, 10 pOSC49.12, pOS49.14, pOS49.16, pOS49.28, pOS44.1, pOS44.2, pOS44.4, pSPM5, pSPM7, pOS49.67, pOS49.88, pOS49.106, pOS49.120, pOS49.107, pOS49.32, pOS49.43, pOS49.44, pOS49.50, pOS49.99, pSPM17, pSPM21, pSPM502, pSPM504, pSPM507, pSPM508, pSPM509, pSPM1, pBXL1111, pBXL1112, pBXL1113, pSPM520, pSPM521, pSPM522, pSPM523, pSPM524, pSPM525, pSPM527, 15 pSPM528, pSPM34, pSPM35, pSPM36, pSPM37, pSPM38, pSPM39, pSPM40, pSPM41, pSPM42, pSPM43, pSPM44, pSPM45, pSPM47, pSPM48, pSPM50, pSPM51, pSPM52, pSPM53, pSPM55, pSPM56, pSPM58, pSPM72, pSPM73, pSPM515, pSPM519, pSPM74, pSPM75, pSPM79, pSPM83, pSPM107, pSPM543 et pSPM106.

20 17) Un autre aspect de l'invention concerne un vecteur d'expression caractérisé en ce qu'il comprend au moins une séquence d'acide nucléique codant un polypeptide selon le paragraphe 7), 8), 9), 10), 11) ou 12) ci-dessus.

18) L'invention est également relative à un système d'expression comprenant un vecteur d'expression approprié et une cellule hôte permettant l'expression d'un ou 25 plusieurs polypeptides selon le paragraphe 7), 8), 9), 10), 11) ou 12) ci-dessus.

19) L'invention est également relative à un système d'expression selon le paragraphe 18 ci-dessus caractérisé en ce qu'il est choisi parmi les systèmes d'expression procaryotes ou les systèmes d'expression eucaryotes.

20) L'invention est également relative à un système d'expression selon le paragraphe 19) ci-dessus caractérisé en ce qu'il est choisi parmi les systèmes d'expression chez la bactérie *E. coli*, les systèmes d'expression baculovirus permettant une expression dans des cellules d'insectes, les systèmes d'expression permettant une
5 expression dans des cellules de levures, les systèmes d'expression permettant une expression dans des cellules de mammifères.

21) L'invention est également relative à une cellule hôte dans laquelle a été introduit au moins un polynucléotide et/ou au moins un ADN recombinant et/ou au moins un vecteur d'expression selon l'un des paragraphes 1), 2), 3), 4), 5), 6), 13), 14),
10 15), 16) ou 17) ci-dessus.

22) L'invention est également relative à un procédé de production d'un polypeptide selon le paragraphe 7), 8), 9), 10), 11) ou 12) ci-dessus caractérisé en ce que ledit procédé comprend les étapes suivantes:

- a) insérer au moins un acide nucléique codant ledit polypeptide dans un vecteur
15 approprié ;
- b) cultiver, dans un milieu de culture approprié, une cellule hôte préalablement transformée ou transfectée avec le vecteur de l'étape a) ;
- c) récupérer le milieu de culture conditionné ou un extrait cellulaire;
- d) séparer et purifier à partir dudit milieu de culture ou encore à partir de
20 l'extrait cellulaire obtenu à l'étape c), ledit polypeptide ;
- e) le cas échéant, caractériser le polypeptide recombinant produit.

23) Un autre aspect de l'invention concerne un microorganisme bloqué dans une étape de la voie de biosynthèse d'au moins un macrolide.

24) Un autre aspect de l'invention concerne un microorganisme selon le
25 paragraphe 23) ci-dessus caractérisé en ce qu'il est obtenu en inactivant la fonction d'au moins une protéine impliquée dans la biosynthèse de ce ou de ces macrolides dans un microorganisme producteur de ce ou de ces macrolide.

25) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 24) ci-dessus caractérisé en ce que l'inactivation de cette ou de ces protéines est effectuée par mutagenèse dans le ou les gènes codant ladite ou lesdites protéines ou par expression d'un ou de plusieurs ARN anti-sens complémentaires de l'ARN ou des
5 ARN messagers codant ladite ou lesdites protéines.

26) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 25) ci-dessus caractérisé en ce que l'inactivation de cette ou de ces protéines est effectuée par mutagenèse par irradiation, par action d'agent chimique mutagène, par mutagenèse dirigée ou par remplacement de gène.

10 27) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 25) ou 26) ci-dessus caractérisé en ce que la ou les mutagenèses sont effectuées *in vitro* ou *in situ*, par suppression, substitution, délétion et/ou addition d'une ou plusieurs bases dans le ou les gènes considérés ou par inactivation génique.

15 28) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 23), 24), 25), 26) ou 27) ci-dessus caractérisé en ce que ledit microorganisme est une bactérie du genre *Streptomyces*.

29) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 23), 24), 25), 26), 27) ou 28) ci-dessus caractérisé en ce que le macrolide est la spiramycine.

20 30) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 23), 24), 25), 26), 27), 28) ou 29) ci-dessus caractérisé en ce que ledit microorganisme est une souche de *S. ambofaciens*.

25 31) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 23), 24), 25), 26), 27), 28), 29) ou 30) ci-dessus caractérisé en ce que la mutagenèse est effectuée dans au moins un gène comprenant une séquence selon l'un des paragraphes 1), 2), 3), 4), 5) ou 6) ci-dessus.

32) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 25), 26), 27), 28), 29), 30) ou 31) ci-dessus caractérisé en ce que la mutagenèse est effectuée dans un ou plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149 ;

33) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 25), 26), 27), 28), 29), 30), 31) ou 32) ci-dessus caractérisé en ce que la mutagenèse consiste en l'inactivation génique d'un gène comprenant une séquence correspondant à la séquence SEQ ID N° 13.

34) Un autre aspect de l'invention concerne une souche de *Streptomyces ambofaciens* caractérisée en ce qu'il s'agit d'une souche choisie parmi l'une des souches déposées auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) le 10 juillet 2002 sous le numéro d'enregistrement I-2909, I-2911, I-2912, I-2913, I-2914, I-2915, I-2916 ou I-2917.

35) Un autre aspect de l'invention concerne un procédé de préparation d'un intermédiaire de biosynthèse de macrolide caractérisé en ce qu'il comprend les étapes suivantes :

- a) cultiver, dans un milieu de culture approprié, un microorganisme selon l'un des paragraphes 23), 24), 25), 26), 27), 28), 29), 30), 31), 32), 33) ou 34) ci dessus,
- b) récupérer le milieu de culture conditionné ou un extrait cellulaire,
- c) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), ledit intermédiaire de biosynthèse.

36) Un autre aspect de l'invention concerne un procédé de préparation d'une molécule dérivée d'un macrolide caractérisé en ce que l'on prépare un intermédiaire de biosynthèse selon le procédé du paragraphe 35) ci-dessus et que l'on modifie l'intermédiaire ainsi produit.

37) Un autre aspect de l'invention concerne un procédé de préparation selon le paragraphe 36) ci-dessus caractérisé en ce que l'on modifie ledit intermédiaire par voie chimique, biochimique, enzymatique et/ou microbiologique.

5 38) Un autre aspect de l'invention concerne un procédé de préparation selon le paragraphe 36) ou 37) ci-dessus caractérisé en ce que l'on introduit dans ledit microorganisme un ou plusieurs gènes codant des protéines susceptibles de modifier l'intermédiaire en s'en servant comme substrat.

39) Un autre aspect de l'invention concerne un procédé de préparation selon le paragraphe 36), 37) ou 38) ci-dessus caractérisé en ce que le macrolide est la
10 spiramycine.

40) Un autre aspect de l'invention concerne un procédé de préparation selon le paragraphe 36), 37), 38) ou 39) ci-dessus caractérisé en ce que le microorganisme utilisé est une souche de *S. ambofaciens*.

41) Un autre aspect de l'invention concerne un microorganisme produisant de la
15 spiramycine I mais ne produisant pas de spiramycine II et III

42) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41) ci-dessus caractérisé en ce qu'il comprend l'ensemble des gènes nécessaires à la biosynthèse de la spiramycine I mais que le gène comprenant la séquence SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de
20 celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide de séquence SEQ ID N° 14 ou l'un de ses variants n'est pas exprimé ou a été rendu inactif.

43) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 42) ci-dessus caractérisé en ce que la dite inactivation est réalisée par mutagenèse dans le gène codant ladite protéine ou par l'expression d'un ARN anti-sens
25 complémentaire de l'ARN messenger codant ladite protéine.

44) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 43) ci-dessus caractérisé en ce que la dite mutagénèse est effectuée dans le promoteur de ce gène, dans la séquence codante ou dans une séquence non codante de manière à rendre inactive la protéine codée ou à en empêcher son expression ou sa
5 traduction.

45) Un autre aspect de l'invention concerne un microorganisme selon la revendication 43) ou 44) ci-dessus caractérisé en ce que la mutagénèse est effectuée par irradiation, par action d'agent chimique mutagène, par mutagénèse dirigée ou par remplacement de gène.

10 46) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 43), 44) ou 45) ci-dessus caractérisé en ce que la mutagénèse est effectuée *in vitro* ou *in situ*, par suppression, substitution, délétion et/ou addition d'une ou plusieurs bases dans le gène considéré ou par inactivation génique.

15 47) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41) ou 42) ci-dessus caractérisé en ce que ledit microorganisme est obtenu en exprimant les gènes de la voie de biosynthèse de la spiramycine sans que ceux-ci ne comprennent le gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide de séquence SEQ ID N° 14 ou l'un de ses
20 variants.

48) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41), 42), 43), 44), 45), 46) ou 47) ci-dessus caractérisé en ce que ledit microorganisme est une bactérie du genre *Streptomyces*.

25 49) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41), 42), 43), 44), 45), 46), 47) ou 48) ci-dessus caractérisé en ce que ledit microorganisme est obtenu à partir d'une souche de départ produisant les spiramycines I, II et III.

50) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41), 42), 43), 44), 45), 46), 47), 48) ou 49) ci-dessus caractérisé en ce qu'il est obtenu par mutagenèse dans un gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide de séquence SEQ ID N° 14 ou l'un de ses variants ayant la même fonction.

51) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 41), 42), 43), 44), 45), 46), 47), 48), 49) ou 50) ci-dessus caractérisé en ce que ledit microorganisme est obtenu à partir d'une souche de *S. ambofaciens* produisant les spiramycines I, II et III, dans laquelle est effectuée une inactivation génique du gène comprenant la séquences correspondant à SEQ ID N° 13 ou l'une des séquences dérivées de celle-ci en raison de la dégénérescence du code génétique.

52) Un autre aspect de l'invention concerne une souche de *S. ambofaciens* caractérisée en ce qu'il s'agit de la souche déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) le 10 juillet 2002 sous le numéro d'enregistrement I-2910.

53) Un autre aspect de l'invention concerne un procédé de production de spiramycine I, caractérisé en ce qu'il comprend les étapes suivantes :

- (a) cultiver, dans un milieu de culture approprié, un microorganisme selon l'un des paragraphes 41), 42), 43), 44), 45), 46), 47), 48), 49), 50), 51) ou 52) ci-dessus,
- (b) récupérer le milieu de culture conditionné ou un extrait cellulaire,
- (c) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), la spiramycine I.

54) Un autre aspect de l'invention concerne l'utilisation d'un polynucléotide selon l'un des paragraphes 1), 2), 3), 4), 5) ou 6) ci-dessus pour améliorer la production en macrolides d'un microorganisme.

55) Un autre aspect de l'invention concerne un microorganisme mutant producteur de macrolide caractérisé en ce qu'il possède une modification génétique dans au moins un gène comprenant une séquence selon l'un des paragraphes 1), 2), 3), 4), 5) ou 6) ci-dessus et/ou qu'il surexprime au moins un gène comprenant une séquence selon l'un des paragraphes 1), 2), 3), 4), 5) ou 6) ci-dessus.

56) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55) ci-dessus caractérisé en ce que la modification génétique consiste en une suppression, une substitution, une délétion et/ou une addition d'une ou plusieurs bases dans le ou les gènes considérés dans le but d'exprimer une ou des protéines ayant une activité supérieure ou d'exprimer un niveau plus élevé de cette ou de ces protéines.

57) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55) ci-dessus caractérisé en ce que la surexpression du gène considéré est obtenue en augmentant le nombre de copie de ce gène et/ou en plaçant un promoteur plus actif que le promoteur sauvage.

58) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55) ou 57) ci-dessus caractérisé en ce que la surexpression du gène considéré est obtenue en transformant un microorganisme producteur de macrolide par une construction d'ADN recombinant selon le paragraphe 13, 14 ou 17 ci-dessus, permettant la surexpression de ce gène.

59) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55), 56), 57) ou 58) ci-dessus caractérisé en ce que la modification génétique est effectuée dans un ou plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique.

60) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55), 56), 57), 58) ou 59) ci-dessus caractérisé le microorganisme surexprime un ou plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique.

61) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55), 56), 57), 58), 59) ou 60) ci-dessus caractérisé en ce que ledit microorganisme est une bactéries du genre *Streptomyces*.

62) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55), 56), 57), 58), 59), 60) ou 61) ci-dessus caractérisé en ce que le macrolide est la spiramycine.

63) Un autre aspect de l'invention concerne un microorganisme mutant selon le paragraphe 55), 56), 57), 58), 59), 60), 61) ou 62) ci-dessus caractérisé en ce que ledit microorganisme est une souches de *S. ambofaciens*.

64) Un autre aspect de l'invention concerne un procédé de production de macrolides, caractérisé en ce qu'il comprend les étapes suivantes :

(a) cultiver, dans un milieu de culture approprié, un microorganisme selon l'un des paragraphes 55), 56), 57), 58), 59), 60), 61), 62), 63) ou 64) ci-dessus,

(b) récupérer le milieu de culture conditionné ou un extrait cellulaire,

(c) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), le ou lesdits macrolides produits.

65) Un autre aspect de l'invention concerne l'utilisation d'une séquence et/ou d'un ADN recombinant et/ou d'un vecteur selon l'un des paragraphes 1), 2), 3), 4), 5), 6), 7), 8), 9), 10), 11), 12), 13), 14), 15), 16) ou 17) ci-dessus pour la préparation d'antibiotiques hybrides.

66) Un autre aspect de l'invention concerne l'utilisation d'au moins un polynucléotide et/ou au moins un ADN recombinant et/ou au moins un vecteur d'expression et/ou au moins un polypeptide et/ou au moins une cellule hôte selon l'un des paragraphes 1), 2), 3), 4), 5), 6), 7), 8), 9), 10), 11), 12), 13), 14), 15), 16), 17) ou
5 21) ci-dessus pour réaliser une ou plusieurs bioconversions.

67) Un autre aspect de l'invention concerne un polynucléotide caractérisé en ce qu'il s'agit d'un polynucléotide complémentaire de l'un des polynucléotides selon le paragraphe 1), 2), 3), 4), 5), ou 6) ci-dessus.

68) Un autre aspect de l'invention concerne un microorganisme producteur
10 d'au moins une spiramycine caractérisé en ce qu'il surexprime :

- un gène susceptible d'être obtenu par amplification en chaîne par polymérase (PCR) en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme
15 matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*,
- ou un gène dérivé de celui-ci en raison de la dégénérescence du code génétique.

69) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 68 ou 90 caractérisé en ce qu'il s'agit d'une bactérie du genre
20 *Streptomyces*.

70) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 68, 69 ou 90 caractérisé en ce qu'il s'agit d'une bactérie de l'espèce *Streptomyces ambofaciens*.

71) Un autre aspect de l'invention concerne un microorganisme selon le
25 paragraphe 68, 69, 70 ou 90 caractérisé en ce que la surexpression dudit gène est obtenue par transformation dudit microorganisme par un vecteur d'expression.

72) Un autre aspect de l'invention concerne une souche de *Streptomyces ambofaciens* caractérisée en ce qu'il s'agit de la souche OSC2/pSPM75(1) ou de la souche OSC2/pSPM75(2) déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 6 octobre 2003 sous le numéro d'enregistrement I-3101.

73) Un autre aspect de l'invention concerne un ADN recombinant caractérisé en ce qu'il comprend :

- un polynucléotide susceptible d'être obtenu par amplification en chaîne par polymérase en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*,

- ou un fragment d'au moins 10, 12, 15, 18, 20 à 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1460, 1470, 1480, 1490 ou 1500 nucléotides consécutifs de ce polynucléotide.

74) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 73 ou 91 caractérisé en ce qu'il s'agit d'un vecteur.

75) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 73, 74 ou 91 caractérisé en ce qu'il s'agit d'un vecteur d'expression.

76) Un autre aspect de l'invention concerne une cellule hôte dans laquelle a été introduit au moins un ADN recombinant selon l'un des paragraphes 73, 74, 75 ou 91.

77) Un autre aspect de l'invention concerne un procédé de production d'un polypeptide caractérisé en ce que ledit procédé comprend les étapes suivantes:

a) transformer une cellule hôte par au moins un vecteur d'expression selon le paragraphe 75;

- b) cultiver, dans un milieu de culture approprié, ladite cellule hôte;
- c) récupérer le milieu de culture conditionné ou un extrait cellulaire;
- d) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape c), ledit polypeptide ;
- 5 e) le cas échéant, caractériser le polypeptide recombinant produit.

78) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 51 caractérisé en ce que l'inactivation génique est effectuée par délétion en phase du gène ou d'une partie du gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'une des séquences dérivées de celle-ci en raison de la
10 dégénérescence du code génétique.

79) Un autre aspect de l'invention concerne un microorganisme selon l'un des paragraphes 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51 ou 78 caractérisé en ce en ce qu'il surexprime en outre :

- un gène susceptible d'être obtenu par amplification en chaîne par polymérase en
15 utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*,
- ou un gène dérivé de ceux-ci en raison de la dégénérescence du code génétique.

20 80) Un autre aspect de l'invention concerne un vecteur d'expression caractérisé en ce que le polynucléotide de séquence SEQ ID N° 47 ou un polynucléotide dérivé de celui-ci en raison de la dégénérescence du code génétique est placé sous le contrôle d'un promoteur permettant l'expression de la protéine codée par le dit polynucléotide chez *Streptomyces ambofaciens*.

25 81) Un autre aspect de l'invention concerne un vecteur d'expression selon le paragraphe 80 caractérisé en ce qu'il s'agit du plasmide pSPM524 ou pSPM525.

82) Un autre aspect de l'invention concerne une souche de *Streptomyces ambofaciens* transformée par un vecteur selon le paragraphe 80 ou 81.

83) Un autre aspect de l'invention concerne un microorganisme selon l'un des paragraphes 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 78, 79 ou 92 caractérisé en ce qu'il surexprime en outre le gène de séquence codante SEQ ID N° 47 ou d'une séquence codante dérivée de celle-ci en raison de la dégénérescence du code génétique.

84) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 83 caractérisé en ce qu'il s'agit de la souche SPM502 pSPM525 déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 26 février 2003 sous le numéro d'enregistrement I-2977.

85) Un autre aspect de l'invention concerne un procédé de production de spiramycine(s), caractérisé en ce qu'il comprend les étapes suivantes :

- 15 (a) cultiver, dans un milieu de culture approprié, un microorganisme selon l'un des paragraphes 68, 69, 70, 71, 72, 78, 79, 82, 83, 84, 90 ou 92,
- (b) récupérer le milieu de culture conditionné ou un extrait cellulaire,
- (c) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), les spiramycines.

20 86) Un autre aspect de l'invention concerne un polypeptide caractérisé en ce que sa séquence comprend la séquence SEQ ID N° 112 ou la séquence SEQ ID N° 142.

87) Un autre aspect de l'invention concerne un polypeptide caractérisé en ce que sa séquence correspond à la séquence traduite de la séquence codante :

- d'un gène susceptible d'être obtenu par amplification en chaîne par polymérase (PCR) en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5'
- 25

GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*,

- ou d'un gène dérivé de celui-ci en raison de la dégénérescence du code génétique.

5 88) Un autre aspect de l'invention concerne un vecteur d'expression permettant l'expression d'un polypeptide selon le paragraphe 86, 87 ou 93 chez *Streptomyces ambofaciens*.

89) Un autre aspect de l'invention concerne un vecteur d'expression selon le paragraphe 88 caractérisé en ce qu'il s'agit du plasmide pSPM75.

10 90) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 68 caractérisé en ce que le gène susceptible d'être obtenu par amplification en chaîne par polymérase est le gène de séquence codante SEQ ID N° 141 ou un gène dérivé de celui-ci en raison de la dégénérescence du code génétique.

15 91) Un autre aspect de l'invention concerne un ADN recombinant selon le paragraphe 73 caractérisé en ce que le polynucléotide susceptible d'être obtenu par amplification en chaîne par polymérase est un polynucléotide de séquence SEQ ID N° 141.

20 92) Un autre aspect de l'invention concerne un microorganisme selon le paragraphe 79 caractérisé en ce que le gène susceptible d'être obtenu par amplification en chaîne par polymérase est le gène de séquence codante SEQ ID N° 141 ou un gène dérivé de celui-ci en raison de la dégénérescence du code génétique.

93) Un autre aspect de l'invention concerne un polypeptide caractérisé en ce que sa séquence est SEQ ID N° 142.

Définitions générales

Le terme “ isolé ” au sens de la présente invention désigne un matériel biologique (acide nucléique ou protéine) qui a été soustrait à son environnement originel
5 (l'environnement dans lequel il est localisé naturellement).

Par exemple un polynucléotide présent à l'état naturel dans une plante ou un animal n'est pas isolé. Le même polynucléotide séparé des acides nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de la plante ou l'animal est considéré comme “ isolé ”.

10 Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeurer néanmoins à l'état isolé du fait que le vecteur ou la composition ne constitue pas son environnement naturel.

Le terme “ purifié ” ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusive de la présence d'autres composés. Il s'agit plutôt d'une
15 définition relative.

Un polynucléotide est à l'état “ purifié ” après purification du matériel de départ ou du matériel naturel d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

Aux fins de la présente invention le terme ORF (« Open Reading Frame », c'est à
20 dire cadre ouvert de lecture) a été employé pour désigner en particulier la séquence codante d'un gène.

Aux fins de la présente description, l'expression “ séquence nucléotidique ” peut être employée pour désigner indifféremment un polynucléotide ou un acide nucléique. L'expression “ séquence nucléotidique ” englobe le matériel génétique lui-même et n'est
25 donc pas restreinte à l'information concernant sa séquence.

Les termes “ acide nucléique ”, “ polynucléotide ”, “ oligonucléotide ” ou encore “ séquence nucléotidique ” englobent des séquences d'ARN, d'ADN, d'ADNc ou encore des séquences hybrides ARN/ADN de plus d'un nucléotide, indifféremment sous la forme simple chaîne ou sous la forme de duplex.

5 Le terme “ nucléotide ” désigne à la fois les nucléotides naturels (A, T, G, C) ainsi que des nucléotides modifiés qui comprennent au moins une modification telle que (1) un analogue d'une purine, (2) un analogue d'une pyrimidine, ou (3) un sucre analogue, des exemples de tels nucléotides modifiés étant décrits par exemple dans la demande PCT N°WO 95/04 064.

10 Aux fins de la présente invention, un premier polynucléotide est considéré comme étant “ complémentaire ” d'un second polynucléotide lorsque chaque base du premier polynucléotide est appariée à la base complémentaire du second polynucléotide dont l'orientation est inversée. Les bases complémentaires sont A et T (ou A et U), ou C et G.

15 Le terme « gènes de la voie de biosynthèse des spiramycines » comprend également les gènes régulateurs et les gènes conférant la résistance aux microorganismes producteurs.

On entendra par “ fragment ” d'un acide nucléique de référence selon l'invention, une séquence nucléotidique de longueur réduite par rapport à l'acide nucléique de référence et comprenant, sur la partie commune, une séquence en nucléotides identique à
20 l'acide nucléique de référence.

Un tel “ fragment ” d'acide nucléique selon l'invention peut être le cas échéant, compris dans un polynucléotide plus grand duquel il est constitutif.

De tels fragments comprennent ou alternativement consistent en, des polynucléotides de longueur allant de 8, 10, 12, 15, 18, 20 à 25, 30, 40, 50, 60, 70, 80,
25 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600,

1650, 1700, 1750, 1800, 1850 ou 1900 nucléotides consécutifs d'un acide nucléique selon l'invention.

Par " fragment " d'un polypeptide selon l'invention, on entendra un polypeptide dont la séquence d'acides aminés est plus courte que celle du polypeptide de référence et
5 qui comprend sur toute la partie commune avec ces polypeptides de référence, une séquence en acides aminés identique.

De tels fragments peuvent, le cas échéant, être compris au sein d'un polypeptide plus grand duquel ils font partie.

De tels fragments d'un polypeptide selon l'invention peuvent avoir une longueur
10 de 10, 15, 20, 30 à 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 420, 440, 460, 480, 500, 520, 540, 560, 580, 600, 620 ou 640 acides aminés.

Par " conditions d'hybridation de forte stringence " au sens de la présente invention, on entendra des conditions d'hybridation défavorisant l'hybridation de brins
15 d'acides nucléiques non homologues. Des conditions d'hybridations de forte stringence peuvent être par exemple décrites comme des conditions d'hybridation dans le tampon décrit par Church & Gilbert (Church & Gilbert, 1984) à une température comprise entre 55°C et 65°C, de manière préférée la température d'hyridation est de 55°C, de manière encore plus préférée la température d'hybridation est de 60°C et de manière tout à fait
20 préférée la température d'hybridation est de 65°C, suivi d'un ou plusieurs lavages effectué en tampon 2X SSC (le tampon 1X SSC correspond à une solution aqueuse 0,15M NaCl, 15 mM de citrate de sodium) à une température comprise entre 55°C et 65°C, de manière préférée cette température est de 55°C, de manière encore plus préférée cette température est de 60°C et de manière tout à fait préférée cette température
25 est de 65°C, suivi d'un ou plusieurs lavages en tampon 0.5X SSC à une température comprise entre 55°C et 65°C, de manière préférée cette température est de 55°C, de manière encore plus préférée cette température est de 60°C et de manière tout à fait préférée cette température est de 65°C.

Il va sans dire que les conditions d'hybridation ci-dessus décrites peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon des techniques connues de l'homme du métier. Les conditions convenables d'hybridation peuvent par exemple être adaptées
5 selon l'ouvrage de F. Ausubel *et al* (2002).

Par " variant " d'un acide nucléique selon l'invention, on entendra un acide nucléique qui diffère d'une ou plusieurs bases par rapport au polynucléotide de référence. Un acide nucléique variant peut être d'origine naturelle, tel qu'un variant allélique retrouvé naturellement, ou peut être aussi un variant non naturel obtenu par
10 exemple par des techniques de mutagenèse.

En général, les différences entre l'acide nucléique de référence et l'acide nucléique variant sont réduites de telle sorte que les séquences nucléotidiques de l'acide nucléique de référence et de l'acide nucléique variant sont très proches et, dans de nombreuses régions, identiques. Les modifications de nucléotides présentes dans un acide nucléique
15 variant peuvent être silencieuses, ce qui signifie qu'elles n'altèrent pas les séquences d'acides aminés codées par ledit acide nucléique variant.

Cependant, les changements de nucléotides dans un acide nucléique variant peuvent aussi résulter de substitutions, additions, délétions dans le polypeptide codé par l'acide nucléique variant par rapport aux peptides codés par l'acide nucléique de
20 référence. En outre, des modifications de nucléotides dans les régions codantes peuvent produire des substitutions, conservatives ou non conservatives dans la séquence d'acides aminés.

De préférence, les acides nucléiques variants selon l'invention codent des polypeptides qui conservent sensiblement la même fonction ou activité biologique que le
25 polypeptide de l'acide nucléique de référence ou encore la capacité à être reconnus par des anticorps dirigés contre les polypeptides codés par l'acide nucléique initial.

Certains acides nucléiques variants coderont ainsi des formes mutées des polypeptides dont l'étude systématique permettra de déduire des relations structure activité des protéines en question.

Par " variant " d'un polypeptide selon l'invention, on entendra principalement un polypeptide dont la séquence d'acides aminés contient une ou plusieurs substitutions, additions ou délétions d'au moins un résidu d'acide aminé, par rapport à la séquence d'acides aminés du polypeptide de référence, étant entendu que les substitutions d'acides aminés peuvent être indifféremment conservatives ou non conservatives.

De préférence, les polypeptides variants selon l'invention conservent sensiblement la même fonction ou activité biologique que le polypeptide de référence ou encore la capacité à être reconnus par des anticorps dirigés contre les polypeptides initiaux.

Par polypeptide ayant " une activité similaire " à un polypeptide de référence au sens de l'invention, on entend un polypeptide ayant une activité biologique proche, mais pas nécessairement identique, de celle du polypeptide de référence tel que mesuré dans un essai biologique convenable à la mesure de l'activité biologique du polypeptide de référence.

Par " antibiotique hybride " au sens de l'invention on entend un composé, généré par la construction d'une voie de biosynthèse artificielle utilisant la technologie de l'ADN recombinant.

Description détaillée de l'invention

La présente invention a plus particulièrement pour objet, de nouveaux gènes de la voie de biosynthèse des spiramycines et de nouveaux polypeptides impliqués dans cette biosynthèse tels que présentés dans la description détaillée ci-dessous.

5 Les gènes de la voie de biosynthèse ont été clonés et la séquence d'ADN de ces gènes a été déterminée. Les séquences obtenues ont été analysées grâce au programme FramePlot (Ishikawa J & Hotta K. 1999). On a identifié, parmi les phases ouvertes de lecture, les phases ouvertes de lecture présentant un usage des codons typiques de *Streptomyces*. Cette analyse a montré que cette région comporte 44 ORFs, localisées de
10 part et d'autre de cinq gènes codant l'enzyme « polyketide synthase » (PKS), et présentant un usage des codons typique de *Streptomyces*. De part et d'autre de ces cinq gènes codant la PKS, respectivement 10 et 34 ORFs ont été identifiées en aval et en amont (l'aval et l'amont étant définis par l'orientation des 5 gènes de PKS tous orientés dans le même sens) (cf. figure 3 et 37). Ainsi, les 34 phases ouvertes de lecture de ce
15 type, occupant une région d'environ 41,7 kb (cf. SEQ ID N° 1 présentant une première région de 31 kb contenant 25 ORFs et SEQ ID N° 140 présentant une région d'environ 12,1 kb dont 1,4 kb chevauchent la séquence précédente (SEQ ID N° 1) et environ 10,7 kb correspondent à la suite de la séquence, cette dernière partie d'environ 10,7 kb contenant 9 ORFs supplémentaires (dont un ORF de séquence partielle), cf. également
20 figure 3 et 37 et ci-dessous), ont été identifiées en amont des 5 gènes codant la PKS et 10 occupant une région d'environ 11,1 kb (SEQ ID N° 2 et figure 3), ont été identifiées en aval des gènes de la PKS. Les 10 gènes situés en aval des 5 gènes PKS ont ainsi été nommés *orf1*c*, *orf2*c*, *orf3*c*, *orf4*c*, *orf5**, *orf6**, *orf7*c*, *orf8**, *orf9**, *orf10** (SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19 et 21). Le « c » ajouté dans le nom du gène signifiant
25 pour l'ORF en question que la séquence codante est dans l'orientation inverse (le brin codant est donc le brin complémentaire de la séquence donnée en SEQ ID N° 2 pour ces gènes). En utilisant la même nomenclature, les 34 ORFs en amont des gènes de la PKS ont été nommés *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9c*, *orf10*, *orf11c*, *orf12*, *orf13c*, *orf14*, *orf15c*, *orf16*, *orf17*, *orf18*, *orf19*, *orf20*, *orf21c*, *orf22c*, *orf23c*, *orf24c*,

orf25c, *orf26*, *orf27*, *orf28c*, *orf29*, *orf30c*, *orf31*, *orf32c*, *orf33* et *orf34c* (SEQ ID N° 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149) (cf. figure 3 et 37).

Les séquences protéiques déduites de ces phases ouvertes de lecture ont été comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et al.*, 1997), CD-search, COGs (Cluster of Orthologous Groups) (ces trois programmes sont accessibles notamment auprès du National Center for Biotechnology Information (NCBI) (Bethesda, Maryland, USA)), FASTA ((Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990)), BEAUTY (Worley K. C. *et al.*, 1995)), (ces deux programmes sont accessibles notamment auprès du centre de ressources INFOBIOGEN, Evry, France). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse des spiramycines.

15 Gènes situés en aval des gènes codant les PKS

Une représentation schématique de l'organisation de la région est présentée en figure 3. Ainsi qu'il sera démontré ci-dessous, sur les 10 gènes identifiés en aval des gènes codant les PKS 9 semblent impliqués dans la biosynthèse ou la résistance aux spiramycines. Il s'agit des 9 gènes suivants : *orf1*c*, *orf2*c*, *orf3*c*, *orf4*c*, *orf5**, *orf6**, *orf7*c*, *orf8** et *orf9**.

Dans le tableau 1 suivant sont présentées les références à la séquence en ADN et en acides aminés des 10 gènes identifiés en aval des 5 gènes PKS.

Tableau 1

Gène	Position dans la séquence SEQ ID N 2	Séquence ADN	Séquences polypeptidiques
<i>orf1*c</i>	10882 à 10172	SEQ ID N 3	SEQ ID N 4
<i>orf2*c</i>	10052 à 8781	SEQ ID N 5	SEQ ID N 6

<i>orf3*c</i>	8741 à 7476	SEQ ID N 7	SEQ ID N 8
<i>orf4*c</i>	7459 à 6100	SEQ ID N 9	SEQ ID N 10
<i>orf5*</i>	5302 à 5976	SEQ ID N 11	SEQ ID N 12
<i>orf6*</i>	4061 à 5305	SEQ ID N 13	SEQ ID N 14
<i>orf7*c</i>	3665 à 2817	SEQ ID N 15	SEQ ID N 16
<i>orf8*</i>	1925 à 2755	SEQ ID N 17	SEQ ID N 18
<i>orf9*</i>	1007 à 1888	SEQ ID N 19	SEQ ID N 20
<i>orf10*</i>	710 à 937	SEQ ID N 21	SEQ ID N 22

Le « c » ajouté dans le nom du gène indique que la séquence codante est dans l'orientation inverse (le brin codant est donc le brin complémentaire de la séquence donnée en SEQ ID N° 2 pour ces gènes).

Dans le but de déterminer la fonction des polypeptides identifiés, trois types d'expériences ont été menés : la comparaison des séquences identifiées avec des séquences de fonctions connues, des expériences d'inactivation de gènes, conduisant à la construction de souches mutantes et des analyses de la production en spiramycines et en intermédiaires de biosynthèse des spiramycines par ces souches mutantes.

Les séquences protéiques déduites de ces phases ouvertes de lecture ont tout d'abord été comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et al.*, 1997), CD-search, COGs (Cluster of Orthologous Groups), FASTA ((Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990), BEAUTY (Worley K. C.. *et al.*, 1995)). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycine. Le tableau 2 montre les protéines présentant une forte similitude avec les produits 10 gènes situés en aval des 5 gènes PKS.

Tableau 2

Produit de gène	Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
<i>orf1*c</i>	TylMI(<i>orf3*</i>) (<i>S. fradiae</i>)	CAA57473	287	N-méthyltransférase
<i>orf2*c</i>	dnrQ gene product (<i>Streptomyces peucetius</i>)	AAD15266	153	Inconnue
<i>orf3*c</i>	TylMII(<i>orf2*</i>) (<i>S. fradiae</i>)	CAA57472	448	Glycosyltransférase
<i>orf4*c</i>	Crotonyl-CoA réductase (<i>S. coelicolor</i>)	NP_630556	772	Crotonyl-CoA réductase
<i>orf5*</i>	MdmC (<i>S. mycarofaciens</i>)	B42719	355	O-méthyltransférase
<i>orf6*</i>	3-O-acyltransférase (<i>S. mycarofaciens</i>)	Q00718	494	Acyltransférase
<i>orf7*c</i>	MdmA (<i>S. mycarofaciens</i>)	A60725	380	Protéine impliquée dans la résistance à la midécamycine
<i>orf8*</i>	ABC-transporteur (<i>S. griseus</i>)	CAC22119	191	ABC-transporteur
<i>orf9*</i>	ABC-transporteur (<i>S. griseus</i>)	CAC22118	269	ABC-transporteur
<i>orf10*</i>	Putative small conserved hypothetical protein (<i>S. coelicolor</i>)	NP_627432	109	Inconnue

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

- Des expériences d'inactivation de gènes ont été réalisées pour confirmer ces résultats.
- 5 Les méthodes utilisées consistent à effectuer un remplacement de gène. Le gène cible à interrompre est remplacé par une copie de ce gène interrompue par une cassette conférant la résistance à un antibiotique (par exemple l'apramycine, la généticine ou

l'hygromycine). Les cassettes utilisées sont bordées de part et d'autre par des codons de terminaison de la traduction dans toutes les phases de lecture et par des terminateurs de transcription actifs chez *Streptomyces*. L'insertion de la cassette dans le gène cible peut s'accompagner ou non d'une délétion dans ce gène cible. La taille des régions flanquant la cassette peut aller de quelques centaines à plusieurs milliers de paires de bases. Un deuxième type de cassettes peut être utilisé pour l'inactivation de gènes : des cassettes dites « cassettes excisables ». Ces cassettes présentent l'avantage de pouvoir être excisées chez *Streptomyces* par un événement de recombinaison de site spécifique après avoir été introduites dans le génome de *S. ambofaciens*. Le but est d'inactiver certains gènes dans des souches de *Streptomyces* sans laisser dans la souche finale de marqueurs de sélection ou de grandes séquences d'ADN n'appartenant pas à la souche. Après excision il subsiste uniquement une courte séquence d'une trentaine de paires de bases (appelé site « cicatriciel ») dans le génome de la souche (cf. figure 10). La mise en oeuvre de ce système consiste, dans un premier temps, au remplacement de la copie sauvage du gène cible (grâce à deux événements de recombinaison homologue, cf. figure 9) par une construction dans laquelle une cassette excisable a été insérée dans ce gène cible. L'insertion de cette cassette est accompagnée d'une délétion dans le gène cible (cf. figure 9). Dans un deuxième temps, l'excision de la cassette excisable du génome de la souche est provoquée. La cassette excisable fonctionne grâce à un système de recombinaison site spécifique et a pour avantage de permettre l'obtention de mutants de *Streptomyces* ne portant finalement pas de gène de résistance. On s'affranchit également d'éventuels effets polaires sur l'expression des gènes situés en aval du ou des gènes inactivés (cf. figure 10). Les souches ainsi construites ont été testées pour leur production en spiramycines.

L'inactivation des gènes *orf1*c*, *orf2*c*, *orf3*c* et *orf4*c* n'a pas été effectuée car les expériences de comparaison de séquence ont permis de déterminer que ces gènes avaient une similitude relativement importante avec des gènes impliqués dans la biosynthèse d'un antibiotique relativement proche. Ainsi, le gène *orf1*c* code une protéine présentant une identité de 66% (déterminée grâce au programme BLAST) avec la protéine codée par gène *tylMI* qui code une N-méthyltransférase impliquée dans la

biosynthèse de tylosine et catalysant la 3 N-méthylation durant la production du mycaminose chez *Streptomyces fradiae* (Gandecha, A.R. *et al.*, 1997 ; Numéro d'accès GenBank : CAA57473 ; Score BLAST : 287). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique relativement proche et plus particulièrement dans la biosynthèse du mycaminose suggère que le gène *orf1*c* code une N-méthyltransférase responsable d'une N-méthylation lors de la biosynthèse de la forosamine ou du mycaminose (cf. figure 5 et 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf1*c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 3).

10

Tableau 3

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
méthyltransférase (<i>S. antibioticus</i>)	CAA05643	277	méthyltransférase
N,N-diméthyltransférase (<i>S. venezuelae</i>)	AAC68678	268	N,N-diméthyltransférase
probable N-méthylase snogX (<i>S. nogalater</i>)	T46679	243	N-méthyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf2*c* code une protéine présentant une relative forte similitude (35% d'identité) avec une protéine codée par le gène *tylMIII* codant une NDP hexose 3,4 isomérase impliquée dans la biosynthèse de la tylosine chez *Streptomyces fradiae* (Gandecha, A.R., *et al.*, 1997 ; Numéro d'accès GenBank : CAA57471 ; Score BMAST : 130). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche et plus particulièrement dans la biosynthèse du mycaminose suggère fortement que le gène *orf2*c* code une NDP hexose 3,4 isomérase responsable

d'isomérisation lors de la biosynthèse d'un des sucres de la spiramycine, éventuellement le mycaminose (cf. figure 5 et 6).

Le gène *orf3*c* code une protéine présentant une relative forte similitude (59% d'identité) avec une protéine codée par le gène *tylMII* codant une glycosyltransférase impliquée dans la biosynthèse de la tylosine chez *Streptomyces fradiae* (Gandecha, A.R. *et al.*, 1997 ; Numéro d'accès GenBank : CAA57472 ; Score BLAST : 448). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf3*c* code une glycosyltransférase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf3*c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 4).

Tableau 4

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Glycosyl transférase (<i>S. venezuelae</i>)	AAC68677	426	Glycosyltransférase
Glycosyltransférase (<i>S. antibioticus</i>)	CAA05642	425	Glycosyltransférase
Glycosyltransférase (<i>Saccharopolyspora erythraea</i>)	CAA74710	395	Glycosyltransférase
Glycosyltransférase (<i>S. antibioticus</i>)	CAA05641	394	Glycosyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf4*c* code une protéine présentant une relative forte similitude avec plusieurs crotonyl-CoA réductases. Notamment, la protéine codée par *orf4*c* possède une similitude importante avec une crotonyl CoA réductase de chez *Streptomyces coelicolor* (Redenbach, M *et al.*, 1996 ; Numéro d'accès GenBank : NP_630556 ; Score

BLAST : 772). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf4*c* code également une crotonyl-CoA réductase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf4*c* présente une forte similitude avec d'autres protéines de fonction similaire
 5 chez d'autres organismes (cf. tableau 5).

Tableau 5

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
trans-2-énoyl-CoA réductase (EC1.3.1.38) (<i>S. collinus</i>)	S72400	764	trans-2-énoyl-CoA réductase
Crotonyl CoA réductase (<i>S. fradiae</i>)	CAA57474	757	Crotonyl CoA réductase
Crotonyl-CoA réductase (<i>S. cinnamomensis</i>)	AAD53915	747	Crotonyl-CoA réductase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf6** présente une certaine similitude avec le gène *mdmB* présent chez
 10 *Streptomyces mycarofaciens* (Hara et Hutchinsnson, 1992 ; numéro d'accès GenBank : A42719 ; Score BLAST : 489) producteur d'antibiotique macrolide. Chez ce producteur, le gène est impliqué dans l'acylation du cycle lactone. Le gène *orf6** coderait donc une acyltransférase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf6** présente une forte similitude avec d'autres protéines de fonction similaire chez
 15 d'autres organismes (cf. tableau 6).

Tableau 6

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
<i>AcyA (Streptomyces thermotolerans)</i>	J4001	450	macrolide 3-O-acyltransférase
Midécamycine 4"-O-propionyl transférase (<i>S. mycarofaciens</i>)	BAA09815	234	Midécamycine 4"-O-propionyl transférase
Mycarose O-acyltransférase (<i>Micromonospora megalomicea subsp. Nigra</i>)	AAG13909	189	Mycarose O-acyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'inactivation du gène *orf6** a été réalisée par une délétion/insertion en phase et il a été montré que la souche résultante ne produit plus de spiramycine II et III mais uniquement de la spiramycine I (cf figure 1). Ceci confirme que le gène *orf6** est bien impliqué dans la synthèse de spiramycine II et III. L'enzyme codée par ce gène est responsable de la formation de spiramycine II et III par fixation d'un groupement acétyl ou butyryl sur le carbone 3. Les souches n'exprimant plus la protéine codée par le gène *orf6** sont particulièrement intéressantes puisqu'elles ne produisent plus de spiramycine II et III mais seulement de la spiramycine I. Comme il a été précisé ci-dessus, l'activité antibiotique de la spiramycine I est nettement supérieure à celle des spiramycines II et III (Liu *et al.*, 1999).

Le gène *orf5** code une protéine présentant une relative forte similitude avec plusieurs O-méthyltransférase. Notamment, la protéine codée par *orf5** possède une similitude importante avec une O-méthyltransférase (EC 2.1.1.-) MdmC de

Streptomyces mycarofaciens (Hara & Hutchinson, 1992 ; Numéro d'accès GenBank :B42719; Score BLAST : 355). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique suggère fortement que le gène *orf5** code également une O-méthyltransférase. Le gène *orf5** serait impliqué dans la formation de précurseurs incorporés dans le cycle lactone. En effet, d'après les comparaisons de séquences, le produit du gène *orf5** est également relativement proche de FkbG qui est responsable de la méthylation de l'hydroxymalonyl-ACP d'après (Wu *et al.*, 2000 ; Hoffmeister *et al.*, 2000 ; Numéro d'accès GenBank : AAF86386 ; Score BLAST : 247) (cf. figure 8). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf5** présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 7).

Tableau 7

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Probable O-méthyltransférase (EC 2.1.1.-) safC (<i>Myxococcus xanthus</i>)	T18553	223	O-méthyltransférase
4-O-méthyltransférase (EC 2.1.1.-) – (<i>Streptomyces sp.</i>)	JC4004	222	O-méthyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Grâce à l'effet polaire de l'insertion d'une cassette non excisée dans le gène *orf6**, il a pu être déterminé que le gène *orf5** est essentiel à la voie de biosynthèse des spiramycines. En effet, l'insertion de la cassette excisable dans la partie codante du gène *orf6** entraîne un arrêt total de la production en spiramycines. Cependant, une fois que la cassette insérée a été excisée (et donc lorsque seul le gène *orf6** est inactivé, cf exemples 14 et 15), on retrouve une production en spiramycine I. Ceci montre que le

gène *orf 5** est essentiel à la voie de biosynthèse des spiramycines puisque son inactivation entraîne un arrêt total de la production en spiramycines.

Le gène *orf5** code une protéine présentant une relative forte similitude avec plusieurs O-méthyltransférase. Le gène *orf5** serait une O-méthyl transférase impliquée
5 soit directement dans la synthèse du platénolide, soit dans la synthèse d'un précurseur méthylé (méthoxymalonyl, voir figure 8) incorporé dans le platénolide par la PKS. Pour vérifier cette hypothèse, des expériences d'analyse CL/SM et de RMN ont été conduites sur une souche de *S. ambofaciens* de génotype : *orf6*::att1Ωhyg+*. Dans cette souche le gène *orf5** n'est pas exprimé à cause de l'effet polaire de l'insertion, dans le gène *orf6**,
10 de la cassette qui contient des terminateurs de transcription (cf. exemple 27). Il a été montré que cette souche produit une molécule dont le spectre UV a une allure similaire à celui de la spiramycine I mais le spectre de masse montre un ion moléculaire à 829. La différence de masse de 14 par rapport à la masse de la spiramycine peut s'expliquer par l'absence de méthyl sur l'oxygène portée par le carbone n°4 du cycle lactone (la
15 structure de ce composé est présenté en figure 39). Les résultats obtenus par RMN sont compatibles avec cette hypothèse. La présence d'un composé à 829 permet de valider l'hypothèse du rôle de *orf5** dans la biosynthèse des spiramycines. En outre, le produit correspondant à la spiramycine sans le groupement méthyle présente une très faible activité microbologique (plus faible d'un facteur 10) par rapport à la spiramycine non
20 modifiée, lorsque testée sur le microorganisme *Micrococcus luteus*.

Le gène *orf7*c* code une protéine présentant une relative forte similitude avec une protéine codée par le gène *mdmA* de *Streptomyces mycarofaciens*, ce dernier codant une protéine impliquée dans la résistance à la midécamycine chez cet organisme producteur (Hara *et al.*, 1990 ; Numéro d'accès GenBank : A60725 ; Score BLAST : 380). Cette
25 similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique suggère fortement que le gène *orf7*c* code également une protéine impliquée dans la résistance à la spiramycine. Plus particulièrement, l'enzyme codée par le gène *orf7*c* possède une activité méthyltransférase et est impliquée dans la résistance à la spiramycine chez *Streptomyces ambofaciens*. Il a été démontré que ce gène confère

une résistance de type MLS I, résistance dont on sait qu'elle est due à la mono-méthylation en position A2058 de l'ARN ribosomal 23S (Pernodet *et al.*, 1996). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf7*c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes

5 (cf. tableau 8).

Tableau 8

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
macrolide-lincosamide-streptogramin B résistance déterminant (<i>S. fradiae</i>)	JC5319	238	23S rRNA méthyltransférase
23S ribosomal RNA methyltransferase ErmML (<i>Micrococcus luteus</i>)	AAL68827	119	23S rRNA méthyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf8** code une protéine présentant une relative forte similitude avec une

10 protéine de type transporteur ABC chez *Streptomyces griseus* (Campelo, 2002, Numéro d'accès GenBank : CAC22119 ; Score BLAST : 191). Cette similitude avec une protéine de type transporteur ABC suggère fortement que le gène *orf8** code également une protéine de type transporteur ABC pouvant être impliquée dans la résistance à la spiramycine. Cette hypothèse est étayée par le fait que la protéine codée par le gène

15 *orf8** présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 9).

Tableau 9

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
AcrW (<i>Streptomyces galilaeus</i>)	BAB72060	94	Transporteur ABC
Daunorubicin resistance transmembrane protéine (<i>Streptomyces peucetius</i>)	P32011	89	Résistance à la daunorubicine
Probable ABC-transporter, transmembrane component. (<i>Streptomyces coelicolor</i>)	NP_626506	89	Transporteur ABC

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

5 Le gène *orf9** code une protéine présentant une relative forte similitude avec une protéine de type transporteur ABC chez *Streptomyces griseus* (Campelo, 2002, Numéro d'accès GenBank : CAC22118 ; Score BLAST : 269). Cette similitude avec une protéine de type transporteur ABC suggère fortement que le gène *orf9** code également une protéine de type transporteur ABC pouvant être impliquée dans la résistance à la spiramycine. Cette hypothèse est étayée par le fait que la protéine codée par le gène
10 *orf8** présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 10).

Tableau 10

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Probable ABC-type transport protein, ATP-binding component (<i>S. coelicolor</i>)	NP_626505	231	Transporteur ABC

Putative ABC transporter ATP-binding component (<i>Streptomyces Coelicolor</i>)	NP_627624	228	Transporteur ABC
---	-----------	-----	------------------

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf10** code une protéine présentant une relative forte similitude avec une protéine de fonction inconnue. Cependant des gènes similaires à *orf10** sont retrouvés au milieu de plusieurs groupes de gènes impliqués dans la biosynthèse d'antibiotiques. Ainsi un gène proche de *orf10** est retrouvé chez *S. coelicolor* (Redenbach *et al.*, 1996, numéro d'accès GenBank : NP_627432, score BLAST : 109). Un gène proche (*CouY*) est également retrouvé chez *S. rishiriensis* (Wang *et al.*, 2000, numéro d'accès GenBank : AAG29779, score BLAST 97).

10

Gènes situés en amont des gènes codant les PKS

Dans la séquence d'ADN située en amont des gènes codant les PKS (l'aval et l'amont étant défini par l'orientation des 5 gènes de PKS tous orientés dans le même sens) (cf. figure 3), 34 ORFs ont été identifiées (cf. ci-dessus). Ainsi, les 34 phases ouvertes de lecture de ce type, occupant une région d'environ 41,7 kb (cf. SEQ ID N° 1 présentant une première région de 31 kb contenant 25 ORFs et SEQ ID N° 140 présentant une région d'environ 12,1 kb dont 1,4 kb chevauchent la séquence précédente (SEQ ID N° 1) et environ 10,7 kb correspondent à la suite de la séquence), cette dernière partie d'environ 10,7 kb contenant 9 ORFs supplémentaires (dont un ORF de séquence partielle), cf. également figure 3 et 37 et ci-dessous). Une représentation schématique de l'organisation de la région est présentée en figure 3 et 37. Les 34 gènes identifiés ont été nommés : *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9c*, *orf10*, *orf11c*, *orf12*, *orf13c*, *orf14*, *orf15c*, *orf16*, *orf17*, *orf18*, *orf19*, *orf20*, *orf21c*, *orf22c*, *orf23c*, *orf24c*, *orf25c*, *orf26*, *orf27*, *orf28c*, *orf29*, *orf30c*, *orf31*, *orf32c*, *orf33* et *orf34c*.

20

Dans le tableau 11 suivant sont présentées les références à la séquence en ADN et en acides aminés des 34 gènes identifiés en amont des 5 gènes PKS.

Tableau 11

Gène ¹	Position dans la séquence SEQ ID N 1	Séquence ADN	Séquence(s) polypeptidique(s) ²
<i>orf1</i>	658 à 1869	SEQ ID N 23	SEQ ID N 24
<i>orf2</i>	1866 à 2405	SEQ ID N 25	SEQ ID N 26 et 27
<i>orf3</i>	2402 à 3568	SEQ ID N 28	SEQ ID N 29
<i>orf4</i>	3565 à 4473	SEQ ID N 30	SEQ ID N 31, 32 et 33
<i>orf5</i>	4457 à 5494	SEQ ID N 34	SEQ ID N 35
<i>orf6</i>	5491 à 6294	SEQ ID N 36	SEQ ID N 37, 38 et 39
<i>orf7</i>	6296 à 7705	SEQ ID N 40	SEQ ID N 41 et 42
<i>orf8</i>	8011 à 9258	SEQ ID N 43	SEQ ID N 44
<i>orf9c</i>	10081 à 9362	SEQ ID N 45	SEQ ID N 46
<i>orf10</i>	10656 à 12623	SEQ ID N 47	SEQ ID N 48
<i>orf11c</i>	14482 à 12734	SEQ ID N 49	SEQ ID N 50, 51 et 52
<i>orf12</i>	14601 à 16031	SEQ ID N 53	SEQ ID N 54, 55, 56, 57, 58 et 59
<i>orf13c</i>	17489 à 16092	SEQ ID N 60	SEQ ID N 61
<i>orf14</i>	17809 à 18852	SEQ ID N 62	SEQ ID N 63
<i>orf15c</i>	20001 à 18961	SEQ ID N 64	SEQ ID N 65
<i>orf16</i>	20314 à 21552	SEQ ID N 66	SEQ ID N 67
<i>orf17</i>	21609 à 22879	SEQ ID N 68	SEQ ID N 69
<i>orf18</i>	22997 à 24175	SEQ ID N 70	SEQ ID N 71

<i>orf19</i>	24177 à 25169	SEQ ID N 72	SEQ ID N 73
<i>orf20</i>	25166 à 26173	SEQ ID N 74	SEQ ID N 75
<i>orf21c</i>	27448 à 26216	SEQ ID N 76	SEQ ID N 77
<i>orf22c</i>	28560 à 27445	SEQ ID N 78	SEQ ID N 79
<i>orf23c</i>	29770 à 28649	SEQ ID N 80	SEQ ID N 81
<i>orf24c</i>	30074 à 29763	SEQ ID N 82	SEQ ID N 83
<i>orf25c</i>	30937 à 30071	SEQ ID N 84	SEQ ID N 85
Gène¹	Position dans la séquence SEQ ID NO : 140	Séquence ADN	Séquence(s) polypeptidique(s)²
<i>orf26</i>	1647 à 2864	SEQ ID N 107	SEQ ID N 108
<i>orf27</i>	2914 à 3534	SEQ ID N 109	SEQ ID N 110
<i>orf28c</i>	4967 à 3804	SEQ ID N° 141	SEQ ID N° 142
<i>orf29</i>	5656 à 6663	SEQ ID N 113	SEQ ID N 114
<i>orf30c</i>	7723 à 6686 7534 à 6686	SEQ ID N 115 SEQ ID N° 143	SEQ ID N 116 et 117 SEQ ID N° 144
<i>orf31</i>	7754 à 8728	SEQ ID N 118	SEQ ID N 119
<i>orf32c</i>	10488 à 8977	SEQ ID N 145	SEQ ID N 146
<i>orf33</i>	10562 à 10837	SEQ ID N 147	SEQ ID N 148
<i>orf34c</i>	12134 à 10899	SEQ ID N 149	SEQ ID N 150

¹ Le « c » ajouté dans le nom du gène indique que la séquence codante est dans l'orientation inverse (le brin codant est donc le brin complémentaire de la séquence donnée en SEQ ID N° 1 ou SEQ ID N° 140 pour ces gènes).

² Lorsque plusieurs séquences protéiques sont indiquées pour une seule orf, les protéines correspondantes sont issues de plusieurs sites possibles de démarrage de la traduction.

Dans le but de déterminer la fonction des polypeptides identifiés dans le tableau 11 ci-dessus, trois types d'expériences ont été menées : la comparaison de l'identité des séquences identifiées avec des séquences de fonctions connues, des expériences d'inactivation de gènes et des analyses de la production en spiramycines de ces souches mutantes.

Les séquences protéiques déduites de ces phases ouvertes de lecture ont tout d'abord été comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et al.*, 1997), CD-search, COGs (Cluster of Orthologous Groups), FASTA ((Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990), BEAUTY (Worley K. C. *et al.*, 1995)), (cf. ci-dessus). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycine. Le tableau 12 montre les protéines présentant une forte similitude avec les 34 gènes situés en amont des 5 gènes PKS.

Tableau 12

Gène	Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
<i>orf1</i>	Cytochrome P450 tyII (<i>Streptomyces fradiae</i>)	S49051	530	Cytochrome P45
<i>orf2</i>	ORF15x4 (<i>Listonella anguillarum</i>)	AAB81630	113	Inconnue
<i>orf3</i>	aminotransferase-like protein (<i>Streptomyces antibioticus</i>)	AAF59939	431	Amino-transférase
<i>orf4</i>	alpha-D-glucose-1-phosphate thymidyltransferase (<i>Streptomyces venezuelae</i>)	AAC68682	404	alpha-D-glucose-1-phosphate thymidyltransferase
<i>orf5</i>	AprE (<i>Streptomyces tenebrarius</i>)	AAG18457	476	dTDP-glucose 4,6-déshydratase

<i>orf6</i>	Thioesterase (<i>Streptomyces avermitilis</i>)	BAB69315	234	Thioestérase
<i>orf7</i>	TylCVI (<i>Streptomyces fradiae</i>)	AAF29379	461	dNTP hexose 2,3-déshydratase
<i>orf8</i>	probable aminotransferase (<i>Saccharopolyspora spinosa</i>)	AAG23279	465	Aminotransférase
<i>orf9c</i>	SrmX (<i>Streptomyces ambofaciens</i>)	S25204	445	Méthyltransférase
<i>orf10</i>	SrmR (<i>Streptomyces ambofaciens</i>)	S25203	1074	Protéine de régulation
<i>orf11c</i>	SrmB (<i>Streptomyces ambofaciens</i>)	S25202	955	Résistance à la spiramycine
<i>orf12</i>	UrdQ (<i>Streptomyces fradiae</i>)	AAF72550	634	NDP-hexose 3,4 déshydratase
<i>orf13c</i>	SC4H2.17 (<i>Streptomyces coelicolor</i>)	T35116	619	Inconnue
<i>orf14</i>	putative reductase (<i>Streptomyces coelicolor</i>)	CAB90862	147	Réductase
<i>orf15c</i>	Probable 3-ketoreductase (<i>Streptomyces antibioticus</i>)	T51102	285	3-cétoréductase
<i>orf16</i>	Hypothetical NDP hexose 3,4 isomerase (<i>Streptomyces fradiae</i>)	CAA57471	209	NDP hexose 3,4 isomérase
<i>orf17</i>	Glycosyl transférase (<i>Streptomyces venezuelae</i>)	AAC68677	400	Glycosyl transférase
<i>orf18</i>	Glycosyl transférase (<i>Streptomyces rishiriensis</i>)	AAG29785	185	Glycosyl transférase
<i>orf19</i>	NDP-hexose 4-cétoreductase TylCIV (<i>Streptomyces fradiae</i>)	AAD41822	266	NDP-hexose 4- cétoréductase
<i>orf20</i>	EryBII (<i>Saccharopolyspora erythraea</i>)	AAB84068	491	aldocéto réductase
<i>orf21c</i>	TylCIII (<i>Streptomyces fradiae</i>)	AAD41823	669	NDP-hexose 3-C méthyltransférase

<i>orf22c</i>	FkbH (<i>Streptomyces hygroscopicus</i>)	AAF86387	463	Impliqué dans la biosynthèse du méthoxymalonyl
<i>orf23c</i>	FkbI (<i>Streptomyces hygroscopicus</i>)	AAF86388	387	Acyl CoA déshydrogénase
<i>orf24c</i>	FkbJ (<i>Streptomyces hygroscopicus</i>)	AAF86389	87	Impliqué dans la biosynthèse du méthoxymalonyl
<i>orf25c</i>	FkbK (<i>Streptomyces hygroscopicus</i>)	AAF86390	268	Acyl CoA déshydrogénase
<i>orf26</i>	TylCV (<i>Streptomyces fradiae</i>)	AAD41824	471	Mycarosyl transférase
<i>orf27</i>	TylCVII (<i>Streptomyces fradiae</i>)	AAD41825	243	NDP-hexose 3,5- (ou 5-) épimérase
<i>orf28c</i>	AcyB2 (<i>Streptomyces thermotolerans</i>)	JC2032	451	protéine régulatrice
<i>orf29</i>	Béta-manannase (<i>Sorangium cellulosum</i>)	AAK19890	139	Glycosyl hydrolase
<i>orf30c</i>	Epimérase de sucre-nucléoside-diphosphate (<i>Corynebacterium glutamicum</i>)	NP_600590	89	Epimérase de sucre-nucléoside-diphosphate
<i>orf31</i>	Oxidoréductase (<i>Streptomyces coelicolor</i>)	NP_631148	261	Oxidoréductase
<i>orf32c</i>	Protéine régulatrice de la famille GntR (<i>Streptomyces avermitilis</i>)	NP_824604	282	Protéine régulatrice
<i>orf33</i>	Protéine hypothétique (<i>Xanthomonas campestris</i>)	NP_635564	54	Inconnue
<i>orf34c</i>	Arabino-furanosidase (<i>Streptomyces coelicolor</i>)	NP_630049	654	Arabino-furanosidase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Des expériences d'inactivation de gènes ont été réalisées pour confirmer ces résultats. Les méthodes utilisées consistent à effectuer un remplacement de gène. Le gène cible à interrompre est remplacé par une copie de ce gène interrompue par une cassette conférant la résistance à un antibiotique (par exemple l'apramycine ou l'hygromycine). Les cassettes utilisées sont bordées de part et d'autre par des codons de terminaison de la traduction dans toutes les phases de lecture et par des terminateurs de transcription actifs chez *Streptomyces*. L'insertion de la cassette dans le gène cible peut s'accompagner ou non d'une délétion dans ce gène cible. La taille des régions flanquant la cassette peut aller de quelques centaines à plusieurs milliers de paires de bases. Un deuxième type de cassettes peut être utilisé pour l'inactivation de gènes : des cassettes dites « cassettes excisables » (cf. ci dessus). Les souches ainsi construites ont été testées pour leur production en spiramycines.

Le gène *orf1* code une protéine présentant une relative forte similitude avec plusieurs cytochrome P450. Notamment, la protéine codée par *orf1* possède une similitude importante avec la protéine codée par le gène *tyl1* impliquée dans la biosynthèse de tylosine chez *Streptomyces fradiae* (Merson-Davies, L.A. *et al.*, 1994 ; Numéro d'accès GenBank : S49051 ; Score BLAST : 530). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf1* code également une cytochrome P450. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf1* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 13).

Tableau 13

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Putative cytochrome P450 YJIB (<i>Bacillus subtilis</i>)	O34374	248	Cytochrome P450

Cytochrome P450 113A1 (<i>Saccharopolyspora erythraea</i>)	P48635	237	Cytochrome P450
Cytochrome P-450 hydroxylase homolog (<i>Streptomyces caelestis</i>)	AAC46023	208	Cytochrome P-450 hydroxylase
Cytochrome P450 monooxygénase (<i>Streptomyces venezuelae</i>)	AAC64105	206	Cytochrome P450 monooxygénase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf2* code une protéine présentant une relative forte similitude avec une
5 dTDP-6-deoxy-3,4-keto-hexulose isomerase de *Aneurinibacillus thermoaerophilus*
(Pfoestl, A. *et al.*, 2003, Numéro d'accès GenBank : AAO06351 ; Score BLAST : 118).
Cette similitude suggère fortement que le gène *orf2* code une isomérase responsable de
la réaction d'isomérisation nécessaire à la biosynthèse d'un des sucres présents dans la
molécule de spiramycine, ce sucre pouvant être le mycarose (cf. figure 5). L'inactivation
10 du gène *orf2* a été réalisée. Il a pu être montré que la souche résultante ne produit plus de
spiramycines. Ceci confirme que le gène *orf2* est bien impliqué dans la biosynthèse des
spiramycines.

Le gène *orf3* code une protéine présentant une relative forte similitude avec
plusieurs aminotransférases. Notamment, la protéine codée par *orf3* possède une
15 similitude importante avec une aminotransférase de *Streptomyces antibioticus* impliquée
dans la biosynthèse de l'oléandomycine (Draeger, G., *et al.*, 1999 ; Numéro d'accès
GenBank : AAF59939 ; Score BLAST : 431). Cette similitude avec une protéine
impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement
que le gène *orf3* code une 3 amino-transférase responsable de la réaction de
20 transamination nécessaire à la biosynthèse d'un des sucres aminés des spiramycines (cf.
figure 5). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf3*

présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 15).

Tableau 15

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Aminotransférase (<i>Streptomyces antibioticus</i>)	T51111	429	Aminotransférase
Transaminase (<i>Streptomyces venezuelae</i>)	AAC68680	419	Transaminase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'inactivation du gène *orf3* a été réalisée. Il a ainsi pu être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf3* est bien impliqué dans la biosynthèse des spiramycines. L'enzyme codée par ce gène est donc bien responsable d'une étape de bioconversion essentielle à la biosynthèse des spiramycines. La production de spiramycines peut être complétée par l'expression de la protéine TylB de *S. fradiae* (cf exemple 23). Ceci démontre que le gène *orf3* code une 3 amino-transférase responsable de la réaction de transamination nécessaire à la biosynthèse du mycaminoose (cf. figure 5). Comme le mycaminoose est le premier sucre à être fixé sur le platénolide, il est attendu que la souche interrompue dans *orf3* (OS49.67) accumule du platénolide.

Les intermédiaires de biosynthèse de la souche interrompue dans le gène *orf3* ont été étudiés (cf. exemple 20). Ces expériences ont permis de démontrer que cette souche produit deux formes de platénolide : le platénolide A et le platénolide B, la structure déduite de ces deux molécules est présentée en figure 36. Cette souche produit également du platénolide A + mycarose et du platénolide B + mycarose (cf. exemple 20

et figure 40). Ces composés comportent un sucre mais ne comportent pas de mycaminose. De plus si on les compare à de la spiramycine (cf. figure 1), ces composés comportent du mycarose à la place du mycaminose. Ces résultats sont en accord avec l'implication du produit du gène *orf3* dans la biosynthèse du mycaminose et son rôle en tant que 3 amino-transférase responsable de la réaction de transamination nécessaire à la biosynthèse du mycaminose (cf. figure 5). On peut remarquer que la spécificité de glycosylation ne semble pas absolue puisque l'on trouve des molécules avec le mycarose fixé à la position normalement occupée par le mycaminose (cf. figure 40).

Le gène *orf4* code une protéine présentant une relative forte similitude avec plusieurs NDP-glucose synthétases. Notamment, la protéine codée par *orf4* possède une similitude importante avec une alpha-D-glucose-1-phosphate thymidyltransférase de *Streptomyces venezuelae* (Xue Y *et al.*, 1998 ; Numéro d'accès GenBank : AAC68682 ; Score BLAST 404). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf4* code une NDP-glucose synthétase responsable de la synthèse du NDP-glucose nécessaire à la biosynthèse des trois sucres atypiques incorporés dans la molécule de spiramycine (cf. figure 4, 5 et 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf4* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 16).

20

Tableau 16

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Glucose-1-phosphate thymidyltransférase (<i>Streptomyces avermitilis</i>)	BAA84594	402	Glucose-1-phosphate thymidyltransférase

AclY (<i>Streptomyces galilaeus</i>)	BAB72036	400	dTDP-1-glucose synthétase
Putative glucose-1-phosphate thymidyltransférase (<i>Saccharopolyspora spinosa</i>)	AAK83289	399	glucose-1-phosphate thymidyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf5* code une protéine présentant une relative forte similitude avec plusieurs glucose déshydratases. Notamment, la protéine codée par *orf5* possède une similitude importante avec une dTDP-glucose 4,6-déshydratase de *Streptomyces tenebrarius* (Li,T.B. *et al.*,2001 ; Numéro d'accès GenBank : AAG18457, Score BLAST : 476). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf5* code une NDP glucose déshydratase nécessaire à la biosynthèse des trois sucres atypiques incorporés dans la molécule de spiramycine (cf. figure 4, 5 et 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf5* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 17).

Tableau 17

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
DTDP-glucose 4,6-déshydratase (<i>Saccharopolyspora spinosa</i>)	AAK83290	464	DTDP-glucose 4,6-déshydratase
thymidine diphosphoglucose 4,6-déshydratase (<i>Saccharopolyspora erythraea</i>)	AAA68211	445	thymidine diphosphoglucose 4,6-déshydratase

dTDPglucose 4,6-déshydratase (EC 4.2.1.46) – (<i>Streptomyces fradiae</i>)	S49054	443	dTDPglucose 4,6-déshydratase
TDP-glucose-4,6-déshydratase (<i>Streptomyces venezuelae</i>)	AAC68681	421	TDP-glucose-4,6-déshydratase
SgcA (<i>Streptomyces globisporus</i>)	AAF13998	418	dNDP-glucose-4,6-déshydratase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf6* code une protéine présentant une relative forte similitude avec plusieurs thioestérases. Notamment, la protéine codée par *orf6* possède une similitude importante avec une thioestérase de *Streptomyces avermitilis* (Omura, S. et al., 2001 ; Numéro d'accès GenBank : BAB69315 ; Score BLAST : 234). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf6* code également une thioestérase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf6* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 18).

Tableau 18

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
RifR (<i>Amycolatopsis mediterranei</i>)	AAG52991	216	Thioestérase
Thioestérase – (<i>Streptomyces fradiae</i>)	S49055	215	Thioestérase
Thioestérase (<i>Streptomyces avermitilis</i>)	BAB69188	213	Thioestérase

Thioestérase II (EC 3.1.2.-) – (<i>Streptomyces venezuelae</i>)	T17413	203	Thioestérase
PimI protein (<i>Streptomyces natalensis</i>)	CAC20922	201	Thioestérase
Thioestérase (<i>Streptomyces griseus</i>)	CAC22116	200	Thioestérase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf7* code une protéine présentant une relative forte similitude avec plusieurs hexose déshydratases. Notamment, la protéine codée par *orf7* possède une similitude importante avec une dNTP hexose 2,3-déshydratase (codée par le gène TylCVI) de *Streptomyces fradiae* impliquée dans la biosynthèse de la tylosine (Merson-Davies, L.A. *et al.*, 1994 ; Numéro d'accès GenBank : AAF29379 ; Score BLAST : 461). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf7* code également une hexose 2-3 déshydratase nécessaire à la biosynthèse de deux sucres atypiques incorporés dans la molécule de spiramycine (cf. figure 4 et 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf7* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 19).

Tableau 19

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Rif18 (<i>Amycolatopsis mediterranei</i>)	AAG52988	459	Hexose déshydratase
SimB3 (<i>Streptomyces antibioticus</i>)	AAK06810	444	dNDP-4-keto-6-déoxy-glucose-2,3-déshydratase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf8* code une protéine présentant une relative forte similitude avec plusieurs aminotransférases. Notamment, la protéine codée par *orf8* possède une similitude importante avec une aminotransférase probablement impliquée dans la biosynthèse de la forosamine chez *Saccharopolyspora spinosa* (Waldron, C. *et al.*, 2001 ;

5 Numéro d'accès GenBank : AAG23279 ; score BLAST : 465). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche, suggère fortement que le gène *orf8* code une 4 amino-transférase responsable de la réaction de transamination nécessaire à la biosynthèse de la forosamine (cf. figure 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf8* présente une forte

10 similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 20).

Tableau 20

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Putative amino-sugar biosynthesis protein (<i>Bordetella bronchiseptica</i>)	CAA07666	213	Proteine impliquée dans la biosynthèse d'amino-sucre

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

15 L'inactivation du gène *orf8* a été réalisée. Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf8* est bien impliqué dans la biosynthèse des spiramycines. L'enzyme codée par ce gène est donc bien responsable d'une étape de bioconversion essentielle à la biosynthèse des spiramycines. La validation de l'hypothèse du rôle joué par le produit du gène *orf8* dans

20 la biosynthèse de la forosamine est apportée par le fait qu'un mutant inactivé pour le gène *orf8* produit de la forocidine, ce mutant est donc bloqué au stade de la forocidine et ne produit pas de néo-spiramycine (cf. figure 7 et exemple 25). Ces résultats sont en

accord avec une implication du produit du gène *orf8* dans la biosynthèse de forosamine (cf. figure 6).

Le gène *orf9c* a déjà été identifié chez *Streptomyces ambofaciens* et a été désigné *srmX* par Geistlich *et al.* (Geistlich, M., *et al.*, 1992). La similitude de la protéine codée par ce gène avec plusieurs méthyltransférases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que le gène *orf9c* code une méthyltransférase responsable de la réaction de méthylation nécessaire à la biosynthèse du mycaminose ou de la forosamine (cf. figure 5 et 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf9c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 21).

Tableau 21

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
N,N-diméthyltransférase (<i>Streptomyces venezuelae</i>)	AAC68678	240	N,N-diméthyltransférase
Méthyltransférase (<i>Streptomyces antibioticus</i>)	CAA05643	232	Méthyltransférase
Putative amino methylase (<i>Streptomyces nogalater</i>)	AAF01819	219	Aminométhylase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf10* a déjà été identifié chez *Streptomyces ambofaciens* et a été désigné *srmR* par Geistlich *et al.* (Geistlich, M., *et al.*, 1992). La protéine codée par ce gène est impliquée dans la régulation de la voie de biosynthèse des spiramycines chez *Streptomyces ambofaciens*. L'inactivation du gène *orf10* a été réalisée. Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le

gène *orf10* est bien impliqué dans la biosynthèse des spiramycines. La protéine codée par ce gène est donc bien essentielle à la biosynthèse des spiramycines.

D'autre part ; le point de démarrage de la traduction de *orf 10* a été déterminé et il a pu être montré que la surexpression de ce gène entraîne une amélioration de la production en spiramycines. Le site de démarrage de la traduction correspond à un ATG
5 situé en amont de l'ATG proposé par Geistlich *et al.* (Geistlich,M., *et al.*, 1992). Il a en outre été démontré que cette extrémité 5' est essentielle à la fonction de Orf10 puisqu'un messenger tronqué en 5' est inactif (cf. exemple 17). Pour obtenir l'effet recherché sur la production en spiramycines, il est donc essentiel que la surexpression de *orf10* soit
10 réalisée en prenant garde de ne pas exprimer un messenger tronqué en 5' de *orf10*.

Le gène *orf11c* a déjà été identifié chez *Streptomyces ambofaciens* et a été désigné *srmB* par Geistlich *et al.* (Geistlich,M. *et al.*, 1992) et Schoner *et al.* (Schoner B *et al.*, 1992). La protéine codée par ce gène est impliquée dans la résistance à la spiramycine chez *Streptomyces ambofaciens* et est un transporteur de type ABC.

15 Le gène *orf12* code une protéine présentant une relative forte similitude avec plusieurs hexose déshydratases. Notamment, la protéine codée par *orf12* possède une similitude importante avec une NDP-hexose 3,4-déshydratase codée par le gène *UrdQ* de *Streptomyces fradiae* et impliquée dans la biosynthèse de l'urdamycine (Hoffmeister,D. *et al.*, 2000 ; Numéro d'accès GenBank : AAF72550 ; Score BLAST :
20 634). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique proche suggère fortement que le gène *orf12* code une 3,4 déshydratase responsable de la réaction de déshydratation nécessaire à la biosynthèse de la forosamine (cf. figure 6). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf12* présente une forte similitude avec d'autres protéines de fonction similaire chez
25 d'autres organismes (cf. tableau 22).

Tableau 22

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
AknP (<i>Streptomyces galilaeus</i>)	AAF73452	625	3-déshydratase
NDP-hexose 3,4-déshydratase homolog (<i>Streptomyces cyanogenus</i>)	AAD13547	624	NDP-hexose 3,4-déshydratase
RdmI (<i>Streptomyces purpurascens</i>)	AAL24451	608	hexose-C-3-déshydratase
Probable CDP-4-keto-6-deoxyglucose-3-déshydratase (E1) (<i>Streptomyces violaceoruber</i>)	T46528	602	CDP-4-céto-6-déoxyglucose-3-déshydratase
Probable NDP-hexose-3,4-déshydratase (<i>Saccharopolyspora spinosa</i>)	AAG23278	582	NDP-hexose-3,4-déshydratase
dNTP-hexose déshydratase (<i>Amycolatopsis mediterranei</i>)	AAC01730	576	dNTP-hexose déshydratase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'inactivation du gène *orf12* a été réalisée. Il a pu être montré que la souche
 5 résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf12* est bien
 impliqué dans la biosynthèse de la spiramycine. L'enzyme codée par ce gène est donc
 bien responsable d'une étape de bioconversion essentielle à la biosynthèse de la
 spiramycine. La validation de l'hypothèse du rôle joué par Orf12 dans la biosynthèse de
 la forosamine est apportée par le fait qu'un mutant inactivé pour le gène *orf12* ne produit
 10 plus de forosamine. Il produit cependant une faible quantité de forocidine. Ce mutant est
 donc bloqué au stade de la forocidine et ne produit pas de néo-spiramycine (cf. figure 7

et exemple 26). Ce mutant produit en outre un composé de structure présentée en figure 38. Ce dernier composé comporte deux sucres, le mycaminose et le mycarose mais ne comporte pas de forosamine. De plus si on le compare à la structure de la spiramycine (cf. figure 1), ce composé comporte le sucre mycarose à la place attendue de la forosamine. Ces résultats sont en accord avec l'implication du produit du gène *orf12* dans la biosynthèse de la forosamine (cf. figure 6). On peut remarquer que la spécificité de glycosylation n'est pas absolue puisque l'on observe des molécules dans lesquelles le mycarose est fixé à la position normalement occupée par la forosamine (voir figure 38).

Le gène *orf13c* code une protéine présentant une relative forte similitude avec une protéine de fonction inconnue chez *Streptomyces coelicolor*. Cette protéine a été nommée SC4H2.17 (numéro d'accès GeneBank : T35116 ; Score BLAST : 619). La protéine codée par le gène *orf13c* présente également une forte similitude avec d'autres protéines d'autres organismes (cf. tableau 23).

Tableau 23

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
hflX protein (<i>Mycobacterium leprae</i>)	S72938	473	Inconnue
Possible ATP/GTP-binding protein (<i>Mycobacterium leprae</i>)	NP_301739	470	Protéine fixant l'ATP/GTP
GTP-binding protein (<i>Mycobacterium tuberculosis CDC1551</i>)	AAK47114	468	Protéine fixant le GTP
ATP/GTP-binding protein (<i>Streptomyces fradiae</i>)	T44592	388	Protéine fixant l'ATP/GTP

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Aucune fonction précise n'a été assignée aux protéines proches de celle codée par *orf13c*. L'inactivation du gène *orf13c* a été réalisée dans le but d'étudier la fonction de ce gène dans la voie de biosynthèse des spiramycines chez *Streptomyces ambofaciens*. Il a pu être montré que la souche résultante produit des spiramycines. Ceci
5 indique que le gène *orf13c* n'est pas essentiel à la biosynthèse des spiramycines et qu'il n'est pas essentiel à la survie de la bactérie. L'enzyme codée par ce gène n'est donc pas responsable d'une étape de bioconversion essentielle à la biosynthèse des spiramycines.

Le gène *orf14* code une protéine présentant une relative forte similitude avec une réductase putative (Redenbach, M., *et al.*, 1996 ; Bentley *et al.*, 2002 ; numéro d'accès
10 GenBank : CAB90862 ; Score BLAST : 147).

L'inactivation du gène *orf14* a été réalisée. Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf14* est bien impliqué dans la biosynthèse des spiramycines. L'enzyme codée par ce gène est donc bien responsable d'une étape de bioconversion essentielle à la biosynthèse des
15 spiramycines. Les intermédiaires de biosynthèse de la souche interrompue dans le gène *orf14* ont été étudiés (cf. exemple 20). Ces expériences ont permis de démontrer que cette souche produit du platénolide A mais ne produit pas de platénolide B (cf. figure 36).

Le gène *orf15c* code une protéine présentant une relative forte similitude avec
20 plusieurs cétoréductases. Notamment, la protéine codée par *orf15c* possède une similitude importante avec une 3-cétoréductase chez *Streptomyces antibioticus* (Numéro d'accès GenBak : T51102, Score BLAST : 285). Cette similitude suggère fortement que le gène *orf15c* code une 3 céto-réductase responsable de la réaction de réduction nécessaire à la biosynthèse de la forosamine (cf. figure 6). Cette hypothèse est étayée par
25 le fait que la protéine codée par le gène *orf15c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 24).

Tableau 24

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
oxidoreductase homolog (<i>Streptomyces cyanogénus</i>)	AAD13550	272	Oxidoréductase
D-oliose 4-cétoreductase (<i>Streptomyces argillaceus</i>)	CAB96550	265	D-oliose 4-cétoréductase
AknQ (<i>Streptomyces galilaeus</i>)	AAF73453	263	putative 3-cétoréductase
Probable NDP-hexose-3-ketoreductase (<i>Saccharopolyspora spinosa</i>)	AAG23275	253	NDP-hexose-3-cétoréductase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf16* code une protéine présentant une relative forte similitude avec
5 plusieurs isomérases. Notamment, la protéine codée par *orf16* possède une similitude
importante avec une NDP hexose 3,4 isomérase chez *Streptomyces fradiae* (Gandecha et
al., 1997 ; Numéro d'accès GenBak : CAA57471, Score BLAST : 209). Cette similitude
suggère fortement que le gène *orf16* code une protéine impliquée dans la biosynthèse
d'un des sucres de la spiramycine (cf. figure 5 et 6). Cette hypothèse est étayée par le fait
10 que la protéine codée par le gène *orf16* présente une forte similitude avec d'autres
protéines de fonction similaire chez d'autres organismes (cf. tableau 25).

Tableau 25

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Putative tautomerase (<i>Streptomyces venezuelae</i>)	AAC68676	145	Tautomérase
TDP-4-céto-6-déoxyhexose 3,4-isomérase (<i>Micromonospora megalomicea subsp. nigra</i>)	AAG13907	112	TDP-4-céto-6-déoxyhexose 3,4-isomérase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf17* code une protéine présentant une relative forte similitude avec plusieurs glycosyl transférases. Notamment, la protéine codée par *orf17* possède une similitude importante avec une glycosyl transférase de *Streptomyces venezuelae* (Xue, Y. *et al.*, 1998 ; Numéro d'accès GenBank : AAC68677 ; score BLAST : 400). La similitude de la protéine codée par le gène *orf17* avec plusieurs glycosyl transférases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code également une glycosyl transférase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf17* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 26).

Tableau 26

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Glycosyltransférase (<i>Streptomyces fradiae</i>)	CAA57472	399	Glycosyltransférase

Glycosyltransférase (<i>Streptomyces antibioticus</i>)	CAA05642	378	Glycosyltransférase
Glycosyltransférase (<i>Streptomyces antibioticus</i>)	CAA05641	360	Glycosyltransférase
Glycosyltransférase (<i>Saccharopolyspora erythraea</i>)	CAA74710	344	Glycosyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf18* code une protéine présentant une relative forte similitude avec plusieurs glycosyl transférases. Notamment, la protéine codée par *orf18* possède une similitude importante avec une glycosyl transférase de *Streptomyces rishiriensis* (Wang *et al.*, 2000 ; Numéro d'accès GenBank : AAG29785 ; score BLAST : 185). La similitude de la protéine codée par le gène *orf18* avec plusieurs glycosyl transférases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code également une glycosyl transférase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf18* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 27).

Tableau 27

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
NovM (<i>Streptomyces spheroides</i>)	AAF67506	184	Glycosyltransférase
probable glycosyl transferase (<i>Streptomyces violaceoruber</i>)	T46519	169	Glycosyltransférase

Glycosyl transferase homolog (<i>Streptomyces cyanogenus</i>)	AAD13553	167	Glycosyltransférase
Glycosyl transferase homolog (<i>Streptomyces cyanogenus</i>)	AAD13555	163	Glycosyltransférase
dNTP-hexose glycosyl transférase (<i>Amycolatopsis mediterranei</i>)	AAC01731	160	Glycosyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf19* code une protéine présentant une relative forte similitude avec
5 plusieurs cétoréductases. Notamment, la protéine codée par *orf19* possède une similitude
importante avec une NDP-hexose 4-cétoréductase (TylCIV) de *Streptomyces fradiae*
(Bate *et al.*, 2000 ; Numéro d'accès GenBank : AAD41822 ; score BLAST : 266). La
similitude de la protéine codée par le gène *orf19* avec cette cétoréductase impliquée dans
la voie de biosynthèse d'un antibiotique proche suggère fortement que ce gène code
10 également une 4-cétoréductase responsable de la réaction de réduction nécessaire à la
biosynthèse du mycarose (cf. figure 4). Cette hypothèse est étayée par le fait que la
protéine codée par le gène *orf19* présente une forte similitude avec d'autres protéines de
fonction similaire chez d'autres organismes (cf. tableau 28).

Tableau 28

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
NDP-4-céto-6-déoxyhexose 4-cétoreductase (<i>Streptomyces venezuelae</i>)	AAL14256	251	NDP-4-céto-6-déoxyhexose 4-cétoréductase

EryBIV (<i>Saccharopolyspora erythraea</i>)	AAB84071	249	oxidoréductase
TDP-4-céto-6-déoxyhexose 4-cétoreductase (<i>Micromonospora megalomicea</i> subsp. <i>Nigra</i>)	AAG13916	218	TDP-4-céto-6-déoxyhexose 4-cétoreductase
dTDP-4-céto-6-déoxy-L-hexose 4-réductase (<i>Streptomyces avermitilis</i>)	BAA84595	212	dTDP-4-céto-6-déoxy-L-hexose 4-réductase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf20* code une protéine présentant une relative forte similitude avec plusieurs hexose réductases. Notamment, la protéine codée par *orf20* possède une similitude importante avec le gène EryBII de *Saccharopolyspora erythraea* qui code une dTDP-4-céto-L-6-déoxy-hexose 2,3-réductase (Summers, R.G., *et al.*, 1997), Numéro d'accès GenBank : AAB84068 ; Score BLAST : 491). La similitude de la protéine codée par le gène *orf20* avec plusieurs hexose réductases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code une 2-3 réductase responsable de la réaction de réduction nécessaire à la biosynthèse du mycarose (cf. figure 4). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf20* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 29).

Tableau 29

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
TylCII (<i>Streptomyces fradiae</i>)	AAD41821	464	NDP-hexose 2,3-énoyl réductase

TDP-4-céto-6-deoxyhexose 2,3-réductase (<i>Micromonospora megalomicea</i> subsp. <i>Nigra</i>)	AAG13914	446	TDP-4-céto-6-déoxyhexose 2,3-réductase
dTDP-4-céto-6-déoxy-L-hexose 2,3-réductase (<i>Streptomyces avermitilis</i>)	BAA84599	377	dTDP-4-céto-6-déoxy-L-hexose 2,3-réductase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf21c* code une protéine présentant une relative forte similitude avec plusieurs hexose méthyltransférases. Notamment, la protéine codée par *orf21c* possède une similitude importante avec le gène *TylCIII* de *Streptomyces fradiae* qui code une NDP-hexose 3-C-méthyltransférase (Bate,N *et al.*, 2000 ; Numéro d'accès GenBank : AAD41823 ; Score BLAST : 669). La similitude de la protéine codée par le gène *orf21c* avec plusieurs hexose méthyltransférases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code une hexose méthyltransférase responsable de la réaction de méthylation nécessaire à la biosynthèse du mycarose (cf. figure 4). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf21c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 30).

Tableau 30

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
EryH (<i>Saccharopolyspora erythraea</i>)	228448	592	Gène de biosynthèse de l'érythromycine

S-adenosyl-dependent methyl transferase (<i>Coxiella burnetii</i>)	AAK71270	358	Méthyltransferase
NovU (<i>Streptomyces spheroides</i>)	AAF67514	184	C-méthyltransférase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf22c* code une protéine présentant une relative forte similitude avec la protéine codée par le gène *fkfH* de *Streptomyces hygroscopicus* var. *ascomyceticus* qui
5 code une enzyme impliquée dans la biosynthèse du méthoxymalonyl (Wu, K. *et al.*, 2000 ; Numéro d'accès GenBank : AAF86387 ; Score BLAST : 463). La similitude de la protéine codée par le gène *orf22c* avec cette protéine impliquée dans la voie de biosynthèse d'un autre macrolide proche suggère fortement que ce gène code également une enzyme impliquée dans la biosynthèse du méthoxymalonyl chez *Streptomyces*
10 *ambofaciens* (cf. figure 8).

Le gène *orf23c* code une protéine présentant une relative forte similitude avec la protéine codée par le gène *fkfI* de *Streptomyces hygroscopicus* var. *ascomyceticus* qui code une acyl CoA déshydrogénase impliquée dans la biosynthèse du méthoxymalonyl (Wu, K. *et al.*, 2000 ; Numéro d'accès GenBank : AAF86388 ; Score BLAST 387). La
15 similitude de la protéine codée par le gène *orf23c* avec plusieurs acyl CoA déshydrogénases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code une acyl CoA déshydrogénase impliquée dans la biosynthèse du méthoxymalonyl (cf. figure 8). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf23c* présente une forte similitude avec d'autres protéines
20 de fonction similaire chez d'autres organismes (cf. tableau 31).

Tableau 31

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Acyl-CoA déshydrogénase (<i>Polyangium cellulosum</i>)	AAK19892	171	Acyl-CoA déshydrogénase
Probable acyl-CoA dehydrogenase – (<i>Streptomyces coelicolor</i>)	T36802	160	acyl-CoA déshydrogénase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf24c* code une protéine présentant une relative forte similitude avec la protéine codée par le gène *fkfJ* de *Streptomyces hygroscopicus* var. *ascomyceticus* qui coderait la protéine de liaison du groupement acyl (Acyl Carrier Protein (ACP)) impliquée dans la biosynthèse du méthoxymalonyl (Wu, K., *et al.*, 2000 ; Numéro d'accès GenBank : AAF86389 ; Score BLAST : 87). La similitude de la protéine codée par le gène *orf24c* avec cette protéine impliquées dans la voie de biosynthèse d'un autre macrolide proche suggère fortement que ce gène code une protéine impliquée dans la biosynthèse du méthoxymalonyl chez *Streptomyces ambofaciens* (cf. figure 8).

Le gène *orf25c* code une protéine présentant une relative forte similitude avec la protéine codée par le gène *fkfK* de *Streptomyces hygroscopicus* var. *ascomyceticus* qui code une acyl CoA déshydrogénase impliquée dans la biosynthèse du méthoxymalonyl (Wu, K., *et al.*, 2000 ; Numéro d'accès GenBank : AAF86390 ; score BLAST : 268). La similitude de la protéine codée par le gène *orf25c* avec plusieurs acyl CoA déshydrogénases impliquées dans la voie de biosynthèse d'autres antibiotiques proches suggère fortement que ce gène code une acyl CoA déshydrogénase impliquée dans la biosynthèse du méthoxymalonyl (cf. figure 8). Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf25c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 32).

Tableau 32

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Probable 3-Hydroxybutyryl-CoA Dehydrogenase (<i>Bacillus subtilis</i>)	P45856	177	3-Hydroxybutyryl-CoA déshydrogénase
3-hydroxybutyryl-CoA dehydrogenase protein (<i>Bacillus thuringiensis</i> serovar <i>kurstaki</i>)	AAL32270	174	3-hydroxybutyryl-CoA déshydrogénase
3-hydroxybutyryl-CoA dehydrogenase (<i>Deinococcus radiodurans</i>)	NP_294792	167	3-hydroxybutyryl-CoA déshydrogénase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf26* code une protéine présentant une identité de 65% (déterminée grâce au programme BLAST) avec la protéine codée par le gène *tylCV* qui code une mycarosyl transférase impliquée dans la biosynthèse de tylosine chez *Streptomyces fradiae* (Bate, N. *et al.*, 2000 ; Numéro d'accès GenBank : AAD41824, Score BLAST : 471). Plus particulièrement, TylCV est une glycosyl-transférase liant la molécule de mycarose lors de la synthèse de la tylosine. Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un autre antibiotique relativement proche et plus particulièrement dans le transfert du mycarose, suggère que le gène *orf26* code une glycosyl transférase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf26* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 33).

Tableau 33

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Glycosyl transferase (<i>Streptomyces avermitilis</i>)	BAA84592	218	Glycosyl transferase
CalG4 (<i>Micromonospora echinospora</i>)	AAM70365	217	Glycosyl transferase
CalG2 (<i>Micromonospora echinospora</i>)	AAM70348	197	Glycosyl transferase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

Le gène *orf27* code une protéine présentant une identité de 70% (déterminée grâce
5 au programme BLAST) avec la protéine codée par le gène *tylCVII* qui code une NDP-
hexose 3,5- (ou 5-) épimérase impliquée dans la biosynthèse de tylosine chez
Streptomyces fradiae (Bate, N. *et al.*, 2000 ; Numéro d'accès GenBank : AAD41825,
Score BLAST : 243). Plus particulièrement, TylCVII est une hexose 3,5- (ou 5-)
épimérase impliquée dans la biosynthèse du mycarose. Cette similitude avec une
10 protéine impliquée dans la voie de biosynthèse d'un autre antibiotique relativement
proche et plus particulièrement dans la biosynthèse du mycarose, suggère que le gène
orf27 code une épimérase. Cette hypothèse est étayée par le fait que la protéine codée
par le gène *orf27* présente une forte similitude avec d'autres protéines de fonction
similaire chez d'autres organismes (cf. tableau 34). Une analyse des séquences proches
15 obtenues grâce au programme BLAST suggère fortement que le gène *orf27* code une 5
épimérase responsable de la réaction d'épimérisation nécessaire à la biosynthèse du
mycarose (cf. figure 4).

Tableau 34

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
LanZ1 (<i>Streptomyces cyanogenus</i>)	AAD13558	172	NDP-hexose 3,5-epimerase
Epi (<i>Saccharopolyspora spinosa</i>)	AAK83288	169	TDP-4-keto-6-deoxyglucose 3,5-epimerase
dNTP-hexose 3,5 epimerase (<i>Amycolatopsis mediterranei</i>)	AAC01732	166	dNTP-hexose 3,5-epimerase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

La séquence d'*orf28c* a dans un premier temps été déterminée de manière partielle, 5
 puisque la séquence d'une région d'environ 450 paires de bases n'a été déterminée qu'après reséquençage (cette région est symbolisée par des « N » dans la séquence incomplète SEQ ID N° 106). La séquence partielle de cette ORF (SEQ ID N° 111) a néanmoins été utilisée pour l'analyse avec les différents programmes informatiques comme expliqué ci-dessus. Il a ainsi pu être déterminé que le gène *orf28c* code une 10
 protéine présentant une identité de 64% sur la séquence déterminée (SEQ ID N° 112, qui est la séquence partielle de la protéine Orf28c) (déterminée grâce au programme BLAST) avec la protéine codée par le gène *acyB2* qui code une protéine régulatrice impliquée dans la biosynthèse de carbomycine chez *Streptomyces thermotolerans* (Arisawa, A., *et al.*, 1993 ; Numéro d'accès GenBank : JC2032, Score BLAST : 329). 15
 Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'un antibiotique relativement proche suggère que le gène *orf28c* code une protéine régulatrice impliquée dans la biosynthèse des spiramycines. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf28c* présente également une forte

similitude avec la protéine TylR qui est une protéine régulatrice impliquée dans la biosynthèse de tylosine chez *Streptomyces fradiae* (Bate, N. *et al.*, 1999; Numéro d'accès GenBank : AAF29380, Score BLAST : 167).

5 Le gène *orf28c* a pu être amplifié grace à des oligonucléotides situés de part et d'autre de la séquence non déterminée et sous cloné dans un vecteur d'expression. Il a pu ainsi être démontré que la surexpression du gène *orf28c* augmente significativement la production en spiramycines de la souche OSC2 (cf. exemple 24). Ceci démontre que la surexpression de *orf28c* conduit à une augmentation de la production en spiramycines et confirme son rôle en tant que régulateur de la voie de biosynthèse des spiramycines.

10 La séquence partielle d'*orf28c* a été complétée par la suite et la région manquante d'environ 450 paires de bases a été déterminée (cf. SEQ ID N° 140 et SEQ ID N° 141). La séquence complète de cette ORF (SEQ ID N° 141) a été utilisée pour l'analyse avec les différents programmes informatiques comme expliqué ci-dessus. Il a ainsi pu être déterminé que le gène *orf28c* code une protéine présentant une identité de 69% sur la
15 séquence déterminée (SEQ ID N° 142, qui est la séquence complète de la protéine Orf28c) (déterminée grâce au programme BLAST) avec la protéine codée par le gène *acyB2* qui code une protéine régulatrice impliquée dans la biosynthèse de carbomycine chez *Streptomyces thermotolerans* (Arisawa, A., *et al.*, 1993 ; Numéro d'accès GenBank : JC2032, Score BLAST : 451). Cette similitude avec une protéine impliquée
20 dans la régulation de la biosynthèse d'un antibiotique relativement proche suggère que le gène *orf28c* code une protéine régulatrice impliquée dans la biosynthèse des spiramycines. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf28c* présente également une forte similitude avec la protéine TylR qui est une protéine régulatrice impliquée dans la régulation de la biosynthèse de tylosine chez *Streptomyces*
25 *fradiae* (Bate, N. *et al.*, 1999; Numéro d'accès GenBank : AAF29380, Score BLAST : 224). Les résultats de la surexpression de ce gène (cf. exemple 24) confirment son rôle en tant que régulateur de la voie de biosynthèse des spiramycines.

L'inactivation du gène *orf28c* a été réalisée. Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf28c* est bien impliqué dans la biosynthèse des spiramycines et est essentiel à la biosynthèse des spiramycines. Ces résultats associés aux résultats de la surexpression de ce gène (cf. exemple 24) sont en accord avec un rôle d'activateur essentiel à la biosynthèse des spiramycines d'Orf28c.

Le gène *orf29* code une protéine présentant une identité de 31% (déterminée grâce au programme BLAST) avec une probable glycosyl hydrolase localisée dans le groupe de gènes impliqué dans la biosynthèse de soraphen A (un antifongique de la classe des polyketides) chez *Sorangium cellulosum* (Ligon, J., *et al.*, 2002 ; Numéro d'accès GenBank : AAK19890, Score BLAST : 139). Cette similitude avec une protéine impliquée dans la voie de biosynthèse d'une molécule relativement proche suggère que le gène *orf29* code une protéine ayant une activité glycosyl hydrolase. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf29* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 35). Une analyse de la séquence de la protéine codée par *orf29* grâce au programme CD-search (cf. ci-dessus) suggère également que le gène *orf29* code une glycosyl hydrolase.

Tableau 35

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
ManA (<i>Caldicellulosiruptor saccharolyticus</i>)	AAC44232	136	beta-1,4-mannanase
ManA (<i>Dictyoglomus thermophilum</i>)	AAB82454	129	beta-mannanase

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'analyse de la séquence protéique déduite de l'*orf29* par le programme SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) (Nielsen, H., *et al.*, 1997) montre que cette

protéine possède une séquence signal C-terminale avec un site de coupure prédit entre les positions 30 et 31 (QSA/QA). On peut prédire que cette protéine est extra-cellulaire. Elle pourrait, en temps que glycosyl-hydrolase, avoir un rôle dans la réactivation de la spiramycine inactivé par glycosylation par les glycosyl-transférases GimA et/ou GimB
5 (Gourmelen *et al*, 1998).

Le gène *orf30c* code une protéine présentant une identité de 31% (déterminée grâce au programme BLAST) avec une épimérase de sucre-nucléoside-diphosphate de *Corynebacterium glutamicum* (Numéro d'accès GenBank : NP_600590, Score BLAST : 89). Cette similitude suggère que le gène *orf30c* code une épimérase. Cette hypothèse est
10 étayée par le fait qu'une analyse de la séquences grâce au programme CD-search (cf. ci-dessus) suggère également que le gène *orf30c* code une épimérase.

L'*orf30c* présente deux codons d'initiations possibles (cf. SEQ ID N° 115) qui donnent deux protéines possibles de 345 et 282 acides aminés respectivement (SEQ ID N° 116 et 117). Toutefois, l'usage des codons est typique de *Streptomyces* seulement à
15 partir du deuxième ATG de plus, la séquence protéique déduite de la séquence entre le premier ATG et le deuxième ne s'aligne pas avec les séquences proches identifiées, alors que la séquence protéique la plus courte (à partir du 2^{ème} ATG : SEQ ID N° 144) s'aligne bien avec le début de ces protéines. On peut donc en déduire que le deuxième ATG est le bon codon d'initiation et que la séquence de cet orf est donc celui présenté
20 en SEQ ID N° 143 qui une fois traduit correspond à la protéine de séquence SEQ ID N° 144.

Le gène *orf31* code une protéine présentant une identité de 52% (déterminée grâce au programme BLAST) avec une oxidoreductase chez *Streptomyces coelicolor* (Numéro d'accès GenBank : NP_631148, Score BLAST : 261). Cette similitude suggère que le
25 gène *orf31* code une réductase. Cette hypothèse est étayée par le fait qu'une analyse de la séquences grâce au programme CD-search (cf. ci-dessus) suggère également que le gène *orf31* code une réductase. Cette hypothèse est également étayée par le fait que la

protéine codée par le gène *orf31* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 36).

Tableau 36

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Putative oxidoreductase (<i>Streptomyces griseus</i>)	BAB79295	173	Oxidoreductase
MocA (<i>Xanthomonas axonopodis</i>)	NP_640644	171	Oxidoreductase

*une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'inactivation du gène *orf31* a été réalisée. Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines. Ceci confirme que le gène *orf31* est bien impliqué dans la biosynthèse des spiramycines. L'enzyme codée par ce gène est donc bien responsable d'une étape de bioconversion essentielle à la biosynthèse des spiramycines.

La séquence d'*orf32c* a tout d'abord été déterminée de manière partielle (cf. exemple 19), puisque la séquence codante en 5' n'a été déterminée que dans un second temps. La séquence partielle de cette orf (SEQ ID N° 120) a néanmoins été utilisée pour l'analyse avec les différents programmes informatiques comme expliqué ci-dessus. Il a ainsi pu être déterminé que le gène *orf32c* code une protéine présentant une identité de 47% sur la séquence déterminée (SEQ ID N° 121, qui est la séquence partielle de la protéine Orf32c) (déterminée grâce au programme BLAST) avec une protéine régulatrice de la famille GntR chez *Streptomyces coelicolor* (Numéro d'accès GenBank : NP_625576, Score BLAST : 229). Cette similitude suggère que le gène *orf32c* code un régulateur transcriptionnel de la famille GntR. Cette hypothèse est étayée par le fait que la

protéine codée par le gène *orf32c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes.

La séquence partielle d'*orf32c* a été complétée par la suite et la région manquante a été déterminée (cf. SEQ ID N° 140 et SEQ ID N° 145). La séquence complète de cette orf code une protéine présentant une identité de 44% (déterminée grâce au programme BLAST) avec une protéine régulatrice de la famille GntR chez *Streptomyces avermitilis* (Numéro d'accès GenBank : NP_824604, Score BLAST : 282). Cette similitude suggère que le gène *orf32c* code un régulateur transcriptionnel de la famille GntR. Cette hypothèse est étayée par le fait que la protéine codée par le gène *orf32c* présente une forte similitude avec d'autres protéines de fonction similaire chez d'autres organismes (cf. tableau 37).

Tableau 37

Protéine présentant une similitude significative	Numéro d'accès GenBank	Score BLAST*	Fonction rapportée
Protéine régulatrice de la famille GntR (<i>Streptomyces avermitilis</i>)	NP_828241	270	Protéine régulatrice de la famille GntR
Protéine régulatrice de la famille GntR (<i>Streptomyces coelicolor</i>)	NP_625576	266	Protéine régulatrice de la famille GntR
SC5G8.04 (<i>Streptomyces coelicolor</i>)	NP_628991	258	Protéine régulatrice de la famille GntR
transcriptional regulator (<i>Streptomyces venezuelae</i>)	AAF01064	224	Régulateur transcriptionnel
Protéine régulatrice de la famille GntR (<i>Streptomyces avermitilis</i>)	NP_827432	239	Protéine régulatrice de la famille GntR

* une plus grande similitude de séquence est associée à un score BLAST plus élevé (Altschul *et al.*, 1990).

L'inactivation du gène *orf32c* a été réalisée dans le but d'étudier la fonction de ce gène dans la voie de biosynthèse des spiramycines chez *Streptomyces ambofaciens*. Il a pu être montré que la souche résultante produit des spiramycines. Ceci indique que le gène *orf32c* n'est pas essentiel à la biosynthèse des spiramycines et qu'il n'est pas essentiel à la survie de la bactérie.

Le gène *orf33* code une protéine présentant une identité de 49% (déterminée grâce au programme BLAST) avec une protéine hypothétique de *Xanthomonas campestris* (Numéro d'accès GenBank : NP_635564, Score BLAST : 54).

La séquence d'*orf34c* est partielle. En effet, les comparaisons effectuées entre le produit de cette orf et les banques de données suggèrent que la partie C-terminale de cette protéine n'est pas dans le produit déduit de la séquence nucléotidique et donc que cette orf serait plus longue et continuerait en dehors de la région séquencée. La séquence partielle de cette ORF a néanmoins été utilisée pour l'analyse avec les différents programmes informatiques comme expliqué ci-dessus. Il a ainsi pu être déterminé que le gène *orf34c* code une protéine présentant une identité de 91% sur la séquence déterminée (SEQ ID N° 150, qui est la séquence partielle de la protéine Orf34c) (déterminée grâce au programme BLAST) avec une arabinofuranosidase de chez *Streptomyces coelicolor* (Bentley *et al.*, 2002; Numéro d'accès GenBank : NP_630049, Score BLAST : 654). Chez *S. coelicolor* le gène codant cette arabinofuranosidase ne semble pas impliqué dans la biosynthèse de métabolite secondaire. Chez *S. ambofaciens*, ce gène n'est donc probablement pas impliqué dans la biosynthèse de spiramycine.

La présente invention a également pour objet des polynucléotides s'hybridant dans des conditions d'hybridation de forte stringence à au moins l'un des polynucléotides de séquence SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des

séquences dérivées de celles-ci en raison de la dégénérescence du code génétique. Préférentiellement, cesdit polynucléotides sont isolés à partir d'une bactérie du genre *Streptomyces*, plus préférentiellement, ces polynucléotides codent des protéines impliquées dans la biosynthèse d'un macrolide et encore plus préférentiellement, ces

5 polynucléotides codent une protéine ayant une activité similaire à la protéine codée par les polynucléotides avec lesquels ils s'hybrident. Les conditions d'hybridation de forte stringence peuvent être définies comme des conditions d'hybridation défavorisant l'hybridation de brins d'acides nucléiques non homologues. Des conditions d'hybridations de forte stringence peuvent être par exemple décrites comme des

10 conditions d'hybridation dans le tampon décrit par Church & Gilbert (Church & Gilbert, 1984) à une température comprise entre 55°C et 65°C, de manière préférée la température d'hyridation est de 55°C, de manière encore plus préférée la température d'hybridation est de 60°C et de manière tout à fait préférée la température d'hybridation est de 65°C, suivi d'un ou plusieurs lavages effectué en tampon 2X SSC à une

15 température comprise entre 55°C et 65°C, de manière préférée cette température est de 55°C, de manière encore plus préférée cette température est de 60°C et de manière tout à fait préférée cette température est de 65°C, suivi d'un ou plusieurs lavages en tampon 0.5X SSC à une température comprise entre 55°C et 65°C, de manière préférée cette température est de 55°C, de manière encore plus préférée cette température est de 60°C

20 et de manière tout à fait préférée cette température est de 65°C. Les conditions d'hybridation ci-dessus décrites peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon des techniques connues de l'homme du métier. Les conditions convenables d'hybridation peuvent par exemple être adaptées selon l'ouvrage de F.Ausubel *et al.*,

25 2002.

L'invention concerne également un polynucléotide ayant au moins 70%, de façon plus préférée 80%, de façon plus préférée 85%, de façon encore plus préférée 90%, de façon encore plus préférée 95% et de manière tout à fait préférée 98% d'identité en nucléotides avec un polynucléotide comprenant au moins 10, 12, 15, 18, 20 à 25, 30,

30 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700,

750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850 ou 1900 nucléotides consécutifs d'un polynucléotide choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 5 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique, ou un polynucléotide de séquence complémentaire. Préférentiellement, ces dits polynucléotides sont isolés à partir d'une bactérie du genre *Streptomyces*, plus préférentiellement, ces polynucléotides codent des 10 protéines impliquées dans la biosynthèse d'un macrolide et encore plus préférentiellement, ces polynucléotides codent des protéines ayant des activités similaires aux protéines codées par les polynucléotides avec lesquels ils présentent l'identité. De façon tout à fait préférée, un polynucléotide selon l'invention est choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 15 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou un polynucléotide de séquence complémentaire.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus, par exemple ceux du package 20 FASTA (Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990), accessible notamment auprès du centre de ressources INFOBIOGEN, Evry, France. À titre d'illustration, le pourcentage d'identité de séquence pourra être déterminé à l'aide du logiciel LFASTA (Chao K.-M. *et al.*, 1992) ou LALIGN (Huang X. et Miller W., 1991). Les programmes LFASTA et LALIGN font partie du package FASTA. LALIGN 25 retourne les alignements locaux optimaux : ce programme est plus rigoureux, mais aussi plus lent que LFASTA.

Un autre aspect de l'invention concerne un polypeptide résultant de l'expression d'une séquence d'acide nucléique telle que définie ci-dessus. Préférentiellement, les polypeptides selon l'invention présentent au moins 70%, de façon plus préférée 80%, de

façon plus préférée 85%, de façon encore plus préférée 90%, de façon encore plus préférée 95% et de manière tout à fait préférée 98% d'identité en acides aminés avec un polypeptide comprenant au moins 10, 15, 20, 30 à 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 420, 440, 460, 480, 500, 520, 540, 560, 580, 600, 620 ou 640 acides aminés consécutifs d'un polypeptide choisi parmi SEQ ID N° 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 27, 29, 31, 32, 33, 35, 37, 38, 39, 41, 42, 44, 46, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 108, 110, 112, 114, 116, 117, 119, 121, 142, 144, 146, 148 et 150, ou l'une de ces séquences excepté que tout au long de ladite séquence un ou plusieurs acides aminés ont été substitués, insérés ou délétés sans en affecter les propriétés fonctionnelles ou l'un des variant de ces séquences. Préférentiellement les polypeptides selon l'invention sont exprimés à l'état naturel par une bactérie du genre *Streptomyces*, plus préférentiellement, ces polypeptides sont impliqués dans la biosynthèse d'un macrolide et encore plus préférentiellement, ces polypeptides ont une activité similaire à celle du polypeptide avec lequel il partage l'identité. De façon préférée, un polypeptide selon l'invention est choisi dans le groupe constitué des séquences polypeptidiques SEQ ID N° 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 27, 29, 31, 32, 33, 35, 37, 38, 39, 41, 42, 44, 46, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 108, 110, 112, 114, 116, 117, 119, 121, 142, 144, 146, 148 et 150, ou l'une de ces séquences excepté que tout au long de ladite séquence un ou plusieurs acides aminés ont été substitués, insérés ou délétés sans en affecter les propriétés fonctionnelles ou l'un des variant de ces séquences. De façon tout à fait préférée, un polypeptide selon l'invention est choisi dans le groupe constitué des séquences polypeptidiques SEQ ID N° 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 27, 29, 31, 32, 33, 35, 37, 38, 39, 41, 42, 44, 46, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 108, 110, 112, 114, 116, 117, 119, 121, 142, 144, 146, 148 et 150.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus, par exemple ceux du package

FASTA (Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990), accessible notamment auprès du centre de ressources INFOBIOGEN, Evry, France. À titre d'illustration, le pourcentage d'identité de séquence pourra être déterminé à l'aide du logiciel LFASTA (Chao K.-M. *et al.*, 1992) ou LALIGN (Huang X. et Miller W., 1991), en utilisant les paramètres par défaut tel que défini par le centre de ressources INFOBIOGEN, Evry, France. Les programmes LFASTA et LALIGN font partie du package FASTA. LALIGN retourne les alignements locaux optimaux : ce programme est plus rigoureux, mais aussi plus lent que LFASTA.

Un autre aspect de l'invention concerne un ADN recombinant comprenant au moins un polynucléotide tel que décrit ci-dessus. Préférentiellement ; cet ADN recombinant est un vecteur. Encore plus préférentiellement, le vecteur est choisi parmi les bactériophages, les plasmides, les phagemides, les vecteurs intégratifs, les fosmides, les cosmides, les vecteurs navettes, les BAC (Bacterial Artificial Chromosome) ou les PAC (P1-derived Artificial Chromosome). A titre illustratif on peut citer comme bactériophages le phage lambda et le phage M13. Comme plasmides on peut citer les plasmides se répliquant chez *E. coli* par exemple pBR322 et ses dérivés, pUC18 et ses dérivés, pUC19 et ses dérivés, pGB2 et ses dérivés (Churchward G. *et al.*, 1984), pACYC177 (numéro d'accès GenBank : X06402) et ses dérivés, pACYC184 (numéro d'accès GenBank : X06403) et ses dérivés. On peut également citer les plasmides se répliquant chez *Streptomyces* comme par exemple pIJ101 et ses dérivés, pSG5 et ses dérivés, SLP1 et ses dérivés, SCP2* et ses dérivés (Kieser *et al.* 2000). Comme phagemides on peut citer à titre illustratif pBluescript II et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)), pGEM-T et ses dérivés (commercialisé par la société Promega (Madison, Wisconsin, USA)), λ ZAPII et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)). Comme vecteurs intégratifs on peut citer à titre illustratif, les vecteurs intégratifs chez *Streptomyces* comme par exemple ceux dérivés de SLP1 (Kieser *et al.*, 2000), ceux dérivés de pSAM2 (Kieser *et al.*, 2000), les vecteurs utilisant les systèmes d'intégration des phages PhiC31 (Kieser *et al.*, 2000) (par exemple pSET152 (Bierman *et al.*, 1992)) ou de VWB (Van Mellaert L, *et al.* 1998), ainsi que les vecteurs utilisant

le système d'intégration de IS117 (Kieser et al. 2000). Comme fosmides on peut citer à titre illustratif le fosmide pFOS1 (commercialisé par la société New England Biolabs Inc., Beverly, Massachusetts, USA) et ses dérivés. Comme cosmides on peut citer à titre illustratif le cosmide SuperCos et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)), le cosmide pWED15 (Wahl *et al.*, 1987) et ses dérivés. Comme vecteurs navettes on peut citer à titre illustratif les plasmides navettes *E. coli* / *Streptomyces* comme par exemple pIJ903 et ses dérivés, la série des plasmides pUWL, pCAO106, pWHM3, pOJ446 et leurs dérivés (Kieser *et al.* 2000), les BAC navettes *E. coli* / *Streptomyces* comme par exemple ceux décrits dans la demande de brevet WO 01/40497. Comme BAC (Bacterial Artificial Chromosome) on peut citer à titre illustratif le BAC pBeloBAC11 (numéro d'accès GenBank :U51113). Comme PAC (P1-derived Artificial Chromosome) on peut citer à titre illustratif le vecteur pCYPAC6 (numéro d'accès GenBank : AF133437). De façon tout à fait préférée, un vecteur selon l'invention est choisi parmi pOS49.1, pOS49.11, pOSC49.12, pOS49.14, pOS49.16, pOS49.28, pOS44.1, pOS44.2, pOS44.4, pSPM5, pSPM7, pOS49.67, pOS49.88, pOS49.106, pOS49.120, pOS49.107, pOS49.32, pOS49.43, pOS49.44, pOS49.50, pOS49.99, pSPM17, pSPM21, pSPM502, pSPM504, pSPM507, pSPM508, pSPM509, pSPM1, pBXL1111, pBXL1112, pBXL1113, pSPM520, pSPM521, pSPM522, pSPM523, pSPM524, pSPM525, pSPM527, pSPM528, pSPM34, pSPM35, pSPM36, pSPM37, pSPM38, pSPM39, pSPM40, pSPM41, pSPM42, pSPM43, pSPM44, pSPM45, pSPM47, pSPM48, pSPM50, pSPM51, pSPM52, pSPM53, pSPM55, pSPM56, pSPM58, pSPM72, pSPM73, pSPM515, pSPM519, pSPM74, pSPM75, pSPM79, pSPM83, pSPM107, pSPM543 et pSPM106.

Un autre aspect de l'invention concerne un système d'expression comprenant un vecteur d'expression approprié et une cellule hôte permettant l'expression d'un ou plusieurs polypeptides tels que décrits ci-dessus dans un système biologique. Les vecteurs d'expression selon l'invention comprennent une séquence d'acide nucléique codant un ou plusieurs polypeptides tels que décrits ci-dessus et peuvent être destinés à l'expression des différents polypeptides selon l'invention dans diverses cellules hôtes bien connues de l'homme du métier. A titre d'exemple, on peut citer les systèmes

d'expression procaryote comme les systèmes d'expression chez la bactérie *E. coli*, les systèmes d'expression eucaryotes, comme le système d'expression baculovirus permettant une expression dans des cellules d'insectes, les systèmes d'expression permettant une expression dans des cellules de levures, les systèmes d'expression permettant une expression dans des cellules de mammifères, notamment des cellules humaines. Les vecteurs d'expression utilisables dans de tels systèmes sont bien connus de l'homme du métier, en ce qui concerne les cellules procaryotes, on peut citer à titre illustratif les vecteurs d'expression chez *E. coli* par exemple, de la famille pET commercialisés par la société Stratagene (LaJolla, Californie, USA), les vecteurs de la famille GATEWAY commercialisés par la société Invitrogen (Carlsbad, Californie, USA), les vecteurs de la famille pBAD commercialisés par la société Invitrogen (Carlsbad, Californie, USA), les vecteurs de la famille pMAL commercialisés par la société New England Biolabs Inc. (Beverly, Massachusetts, USA), les vecteurs d'expression inductibles par le rhamnose cités dans la publication Wilms B *et al.*, 2001 et leurs dérivés, on peut également citer les vecteurs d'expression chez *Streptomyces* comme par exemple les vecteurs pIJ4123, pIJ6021, pPM927, pANT849, pANT 850, pANT 851, pANT1200, pANT1201, pANT1202 et leurs dérivés (Kieser *et al.*, 2000). En ce qui concerne les cellules de levures, on peut citer à titre illustratif le vecteur pESC commercialisé par la société Stratagene (LaJolla, Californie, USA). En ce qui concerne le système d'expression baculovirus permettant une expression dans des cellules d'insectes on peut citer à titre illustratif le vecteur BacPAK6 commercialisé par la société BD Biosciences Clontech, (Palo Alto, Californie, USA). En ce qui concerne les cellules de mammifères, on peut citer à titre illustratif des vecteurs comprenant le promoteur des gènes précoces du virus CMV (Cytomegalovirus) (par exemple le vecteur pCMV et ses dérivés commercialisés par la société Stratagene (LaJolla, Californie, USA), ou le promoteur des gènes immédiats (SV40 early promoter) du virus simien vacuolisant SV40 (par exemple le vecteur pSG5 commercialisé par la société Stratagene (LaJolla, Californie, USA)).

L'invention est également relative à un procédé de production d'un polypeptide tel que décrit ci-dessus ledit procédé comprenant les étapes suivantes:

a) insérer un acide nucléique codant ledit polypeptide dans un vecteur approprié ;

b) cultiver, dans un milieu de culture approprié, une cellule hôte préalablement transformée ou transfectée avec le vecteur de l'étape a) ;

5 c) récupérer le milieu de culture conditionné ou un extrait cellulaire, par exemple par sonication ou par choc osmotique ;

d) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape c), ledit polypeptide ;

e) le cas échéant, caractériser le polypeptide recombinant produit.

10 Un polypeptide recombinant selon l'invention peut être purifié par passage sur une série appropriée de colonnes de chromatographie, selon les méthodes connues de l'homme de l'art et décrites par exemple dans F.Ausubel et al (2002). On peut citer à titre illustratif la technique du « Tag-Histidine », qui consiste à ajouter une courte
séquence poly-histidine au polypeptide à produire, ce dernier pouvant ensuite être
15 purifié sur colonne de nickel. Un polypeptide selon l'invention peut être également préparé par les techniques de synthèse *in vitro*. A titre illustratif de telles techniques, un polypeptide selon l'invention pourra être préparé grâce au système « rapid translation system (RTS) », commercialisé notamment par la société Roche Diagnostics France S.A, Meylan, France.

20 Un autre aspect de l'invention concerne des cellules hôtes dans laquelle a été introduit au moins un polynucléotide et/ou au moins un ADN recombinant et/ou au moins un vecteur d'expression selon l'invention.

Un autre aspect de l'invention concerne des microorganismes bloqués dans une étape de la voie de biosynthèse d'au moins un macrolide. L'intérêt réside d'une part dans
25 l'étude de la fonction des protéines mutées et d'autre part dans la réalisation de microorganismes produisant des intermédiaires de biosynthèse. Ces intermédiaires

peuvent être modifiés, éventuellement après séparation, soit par ajout de composants particuliers dans les milieux de production, soit par introduction dans les microorganismes ainsi mutés d'autres gènes codant des protéines susceptibles de modifier l'intermédiaire en s'en servant comme substrat. Ces intermédiaires peuvent
5 ainsi être modifiés par voie chimique, biochimique, enzymatique et/ou microbiologique. Les microorganismes bloqués dans une étape de la voie de biosynthèse de macrolides peuvent être obtenus en inactivant la fonction d'une ou plusieurs protéines impliquées dans la biosynthèse de ce ou de ces macrolides dans des microorganismes producteurs de ce ou de ces macrolide. Selon la ou les protéines inactivées, on pourra ainsi obtenir
10 des microorganismes bloqués dans les différentes étapes de la voie de biosynthèse de ce ou de ces macrolides. L'inactivation de cette ou de ces protéines peut être effectuée par toute méthode connue de l'homme du métier, par exemple par mutagenèse dans le ou les gènes codant ladite ou lesdites protéines ou par l'expression d'un ou de plusieurs ARN anti-sens complémentaires de l'ARN ou des ARN messagers codant ladite ou
15 lesdites protéines. La mutagenèse peut, par exemple, être effectuée par irradiation, par action d'agent chimique mutagène, par mutagenèse dirigée, par remplacement de gène ou toute autre méthode connue de l'homme du métier. Les conditions convenables d'une telle mutagenèse peuvent par exemple être adaptées selon l'enseignement contenu dans les ouvrages de Kieser, T *et al* (2000) et Ausubel et al., (2002). La mutagenèse
20 peut être effectuée *in vitro* ou *in situ*, par suppression, substitution, délétion et/ou addition d'une ou plusieurs bases dans le gène considéré ou par inactivation génique. Cette mutagenèse peut être effectuée dans un gène comprenant une séquence telle que décrite ci-dessus. Préférentiellement, les microorganismes bloqués dans une étape de la voie de biosynthèse de macrolides sont des bactéries du genre *Streptomyces*. Plus
25 préférentiellement, l'inactivation de la fonction d'une ou plusieurs protéines impliquées dans la biosynthèse du ou des macrolides en question est effectuée par mutagenèse. Encore plus préférentiellement, le macrolide en question est la spiramycine et les microorganismes dans lesquels sont effectuées la ou les mutagenèses sont des souches de *S. ambofaciens*. Plus préférentiellement, la mutagenèse est effectuée dans un ou
30 plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des

séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149. De façon préférée, la ou les mutagénèses sont effectuées par inactivation génique. De façon tout à fait préférée, la mutagénèse consiste
5 en l'inactivation génique d'un gène comprenant une séquence correspondant à la séquence SEQ ID N° 13.

A titre illustratif, les microorganismes suivants peuvent être cités à titre d'exemple de tels microorganismes : OS49.16 (*orf3*:: Ω hyg, cf. exemple 2), OS49.67 (*orf3*délétion en phase, cf. exemple 6), OS49.107 (*orf8*:: Ω hyg, cf. exemple 7), OS49.50
10 (*orf10*:: Ω hyg, cf. exemple 8), SPM21 (*orf2*::*att3* Ω aac-, cf. exemple 10), SPM22 (*orf2*::*att3* délétion en phase, cf. exemple 10), SPM501 (*orf6**::*att1* Ω hyg+, cf. exemple 14), SPM502 (*orf6**::*att1* délétion en phase, cf. exemple 14), SPM507 (*orf12*::*att3* Ω aac-, cf. exemple 11), SPM508 (*orf13c*::*att3* Ω aac-, cf. exemple 12), SPM509 (*orf14*::*att3* Ω aac-, cf. exemple 13), SPM107 (*orf28c*::*att3a*ac+, cf. exemple
15 29), SPM543 (*orf31*::*att3a*ac+, cf. exemple 30), SPM106 (*orf32c*::*att3a*ac+, cf. exemple 31).

Un autre aspect de l'invention concerne un procédé de préparation d'un intermédiaire de biosynthèse de macrolide utilisant les microorganismes bloqués dans une étape de la voie de biosynthèse de macrolides, tels que décrits ci-dessus. Le procédé
20 consiste à cultiver, dans un milieu de culture approprié, un microorganisme bloqué dans une étape de la voie de biosynthèse de macrolides, tels que décrits ci-dessus, récupérer le milieu de culture conditionné ou un extrait cellulaire, par exemple par sonication ou par choc osmotique, séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape précédente, ledit intermédiaire de biosynthèse.
25 Les conditions de culture de tels microorganismes pourront être déterminées selon des techniques bien connues de l'homme du métier. Le milieu de culture pourra par exemple être le milieu MP5 ou le milieu SL11 pour les *Streptomyces* et notamment pour *Streptomyces ambofaciens* (Pernodet *et al.*, 1993). L'homme du métier pourra notamment se référer à l'ouvrage de Kieser *et al.* (2000) en ce qui concerne la culture

des *Streptomyces*. L'intermédiaire produit peut être récupéré par toutes techniques connues de l'homme du métier. L'homme du métier pourra se référer, par exemple, aux techniques enseignées dans le brevet américain US 3,000,785 et plus particulièrement aux méthodes d'extraction des spiramycines décrites dans ce brevet.

5 Un autre objet de l'invention concerne un procédé de préparation d'une molécule dérivée d'un macrolide utilisant les microorganismes bloqués dans une étape de la voie de biosynthèse de ce macrolide, tels que décrits ci-dessus. Le procédé consiste à obtenir un intermédiaire de biosynthèse selon le procédé ci-dessus et à modifier l'intermédiaire ainsi produit, éventuellement après séparation du milieu de culture. Les conditions de
10 culture de tels microorganismes pourront être déterminées selon des techniques bien connues de l'homme du métier. Le milieu de culture pourra par exemple être le milieu MP5 ou le milieu SL11 pour les *Streptomyces* et notamment pour *Streptomyces ambofaciens* (Pernodet *et al.*, 1993). L'homme du métier pourra notamment se référer à l'ouvrage de Kieser *et al.* 2000 en ce qui concerne la culture des *Streptomyces*. Les
15 intermédiaires produits peuvent être modifiés, éventuellement après séparation, soit par ajout de composants appropriés dans les milieux de production, soit par introduction dans les microorganismes d'autres gènes codant des protéines susceptibles de modifier l'intermédiaire en s'en servant comme substrat. Ces intermédiaires peuvent ainsi être modifiés par voie chimique, biochimique, enzymatique et/ou microbiologique. Plus
20 préférentiellement, le macrolide en question est la spiramycine et les microorganismes dans lesquels sont effectuées la ou les mutagénèses sont des souches de *S. ambofaciens*.

L'invention est également relative à un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III. Ce microorganisme comprend l'ensemble des gènes nécessaires à la biosynthèse de la spiramycine I mais ne
25 produit pas de spiramycine II et III car le gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide de séquence SEQ ID N° 14 ou l'un de ses variants n'est pas exprimé ou a été rendu inactif. L'inactivation de cette protéine peut être effectuée par toute méthode connue de

l'homme du métier, par exemple par mutagénèse dans le gène codant ladite protéine ou par l'expression d'un ARN anti-sens complémentaire de l'ARN messager codant ladite protéine, étant entendu que si l'expression de *orf5** est affectée par cette manipulation, il sera nécessaire de procéder à une autre modification pour que le gène *orf5** soit
5 correctement exprimé. La mutagénèse peut être effectuée dans la séquence codante ou dans une séquence non codante de manière à rendre inactive la protéine codée ou à empêcher son expression ou sa traduction. La mutagénèse peut être effectuée par mutagénèse dirigée, par remplacement de gène ou toute autre méthode connue de l'homme du métier. Les conditions convenables d'une telle mutagénèse peuvent par
10 exemple être adaptées selon l'enseignement contenu dans les ouvrages de Kieser, T *et al* (2000) et Ausubel *et al.*, 2002. La mutagénèse peut être effectuée *in vitro* ou *in situ*, par suppression, substitution, délétion et/ou addition d'une ou plusieurs bases dans le gène considéré ou par inactivation génique. Le microorganisme peut être également obtenu en exprimant les gènes de la voie de biosynthèse de la spiramycine sans que
15 ceux-ci ne comprennent le gène comprenant la séquence SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide de séquence SEQ ID N° 14 ou l'un de ses variants. Préférentiellement, le microorganisme est une bactérie du genre *Streptomyces*. Plus préférentiellement, le microorganisme produisant de la spiramycine I mais ne
20 produisant pas de spiramycine II et III est obtenu à partir d'un microorganisme de départ produisant les spiramycines I, II et III. Encore plus préférentiellement, le microorganisme est obtenu par mutagénèse dans un gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et codant un polypeptide
25 de séquence SEQ ID N° 14 ou l'un de ses variants ayant la même fonction. Encore plus préférentiellement, cette mutagénèse est effectuée par inactivation génique. De manière préférée le microorganisme est obtenu à partir d'une souche de *S. ambofaciens* produisant les spiramycines I, II et III dans laquelle est effectuée une inactivation génique du gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'une des
30 séquences dérivées de celle-ci en raison de la dégénérescence du code génétique. De

manière tout à fait préférée, l'inactivation génique est effectuée par délétion en phase du gène ou d'une partie du gène comprenant la séquence correspondant à SEQ ID N° 13 ou l'une des séquences dérivées de celle-ci en raison de la dégénérescence du code génétique. A titre illustratif, la souche SPM502 (*orf6*::att1*, cf. exemple 14) peut être citée comme un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III.

L'invention est également relative à un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III tel que décrit ci-dessus caractérisé en ce en ce qu'il surexprime en outre :

- un gène susceptible d'être obtenu par amplification en chaîne par polymérase en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, plus préférentiellement il s'agit du gène de séquence codante SEQ ID N° 141,
- ou un gène dérivé de ceux-ci en raison de la dégénérescence du code génétique.

Un exemple de séquence d'un tel gène est donné en SEQ ID N° 111 (ADN), cependant cette séquence est partielle puisqu'elle ne comprend pas la partie 3' de la séquence codante correspondante. La traduction en protéine de cette partie de séquence codante est donnée en SEQ ID N° 112. L'homme du métier pourra aisément la compléter notamment en utilisant l'enseignement présenté dans l'exemple 24. La séquence indéterminée dans SEQ ID N° 111 a été déterminée dans un second temps et la séquence complète de cette orf (*orf28c*) est donnée en SEQ ID N° 141. La traduction en protéine de cette séquence codante est donnée en SEQ ID N° 142. Il est présenté dans l'exemple 24 une méthode pour cloner le gène *orf28c* et pour fabriquer un vecteur d'expression permettant l'expression de *orf28c*. Il est également montré dans cet exemple que la surexpression du gène *orf28c* dans la souche OSC2 conduit à l'amélioration de la production en spiramycines de cette souche. La surexpression du gène *orf28c* peut être obtenue en augmentant le nombre de copies de ce gène et/ou en plaçant un promoteur

plus actif que le promoteur sauvage. De façon préférée, la surexpression dudit gène est obtenue en apportant dans le microorganisme une construction d'ADN recombinant, permettant la surexpression de ce gène. De façon préférée, cette construction d'ADN recombinant augmente le nombre de copie dudit gène et permet d'obtenir une
5 surexpression dudit gène. Dans cette construction d'ADN recombinant, la séquence codante du gène peut être placée sous le contrôle d'un promoteur plus actif que le promoteur sauvage. On peut citer à titre d'illustration le promoteur ptrc actif chez *Streptomyces ambofaciens* (Amann, E. *et al.*, 1988) ainsi que le promoteur ermE*. Ainsi, de façon préférée, la ou les copies du gène *orf28c* apportée sont placées sous le
10 contrôle du promoteur ermE* comme c'est le cas dans la construction pSPM75 (cf. exemple 24).

L'invention est également relative à un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III tel que décrit ci-dessus caractérisé en ce en ce qu'il surexprime en outre le gène de séquence codante SEQ ID
15 N° 47 ou d'une séquence codante dérivée de celle-ci en raison de la dégénérescence du code génétique. Préférentiellement, ce microorganisme est caractérisé en ce qu'il s'agit de la souche SPM502 pSPM525 déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 26 février 2003 sous le numéro d'enregistrement I-2977.

20 L'invention est également relative à un procédé de production de spiramycine I, le procédé consiste à cultiver dans un milieu de culture approprié un microorganisme produisant de la spiramycine I mais ne produisant pas de spiramycine II et III tel que décrit ci-dessus, récupérer le milieu de culture conditionné ou un extrait cellulaire, séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait
25 cellulaire obtenu à l'étape précédente la spiramycine I. Les conditions de culture de tel microorganisme pourront être déterminées selon des techniques bien connues de l'homme du métier. Le milieu de culture pourra par exemple être le milieu MP5 ou le milieu SL11 pour les *Streptomyces* et notamment pour *Streptomyces ambofaciens* (Pernodet *et al.*, 1993). L'homme du métier pourra notamment se référer à l'ouvrage de

Kieser et al. 2000 en ce qui concerne la culture des *Streptomyces*. La spiramycine I produite peut être récupérée par toutes techniques connues de l'homme du métier. L'homme du métier pourra se référer, par exemple, aux techniques enseignées dans le brevet américain US 3,000,785 et plus particulièrement aux méthodes d'extraction des spiramycines décrites dans ce brevet.

Un autre aspect de l'invention concerne l'utilisation d'une séquence nucléotidique selon l'invention pour améliorer la production en macrolides d'un microorganisme. Ainsi, l'invention concerne un microorganisme mutant producteur de macrolide caractérisé en ce qu'il possède une modification génétique dans au moins un gène comprenant une séquence telle que définie ci-dessus et/ou qu'elle surexprime au moins un gène comprenant une séquence telle que définie ci-dessus. La modification génétique peut consister en une suppression, une substitution, une délétion et/ou une addition d'une ou plusieurs bases dans le ou les gènes considérés dans le but d'exprimer une ou des protéines ayant une activité supérieure ou d'exprimer un niveau plus élevé de cette ou de ces protéines. La surexpression du gène considéré peut être obtenue en augmentant le nombre de copies de ce gène et/ou en plaçant un promoteur plus actif que le promoteur sauvage. On peut citer à titre d'illustration le promoteur ptrc actif chez *Streptomyces ambofaciens* (Amann, E. et al., 1988) ainsi que le promoteur ermE* (Bibb et al., 1985), (Bibb et al., 1994). Ainsi la surexpression du gène considéré peut être obtenue en apportant dans un microorganisme producteur du macrolide considéré une construction d'ADN recombinant selon l'invention, permettant la surexpression de ce gène. En effet, certaines étapes de la biosynthèse des macrolides sont limitantes et si l'on exprime une ou plusieurs protéines plus actives ou un niveau d'expression plus important de la ou des protéines sauvages impliquées dans ces étapes limitantes, on peut améliorer la production du ou des macrolides concernés. Une augmentation du taux de production de la tylosine a par exemple été obtenue chez *Streptomyces fradiae* par duplication du gène codant une méthyl transférase limitante convertissant la macrocine en tylosine (Baltz R, 1997). L'obtention de l'expression d'une protéine plus active peut être obtenue notamment par mutagenèse, l'homme du métier pourra par exemple se référer à ce sujet à l'ouvrage de F.Ausubel et al (2002). Préférentiellement, ces microorganismes mutants

améliorés pour leur production en macrolides sont des bactéries du genre *Streptomyces*. Plus préférentiellement, le macrolide en question est la spiramycine et les microorganismes dans lesquels sont effectuées la ou les mutagénèses sont des souches de *S. ambofaciens*. Plus préférentiellement, la modification génétique est effectuée dans

5 un ou plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique. De façon

10 préférée, le microorganisme surexprime un ou plusieurs gènes comprenant une des séquences correspondant à une ou plusieurs des séquences SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 107, 109, 111, 113, 115, 118, 120, 141, 143, 145, 147 et 149, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la

15 dégénérescence du code génétique. De façon préférée, le microorganisme surexprime le gène comprenant une séquence correspondant à SEQ ID N° 111 ou 141, ou l'un de ses variants ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique. La séquence donnée en SEQ ID N° 111 est partielle, cependant l'homme du métier pourra aisément la compléter notamment en utilisant l'enseignement

20 présenté dans l'exemple 24. La séquence indéterminée dans SEQ ID N° 111 a été déterminée dans un second temps et la séquence complète de cette orf (*orf28c*) est donnée en SEQ ID N° 141. La traduction en protéine de cette séquence codante est donnée en SEQ ID N° 142. Il est ainsi présenté l'exemple 24 une méthode pour cloner le gène *orf28c* et pour fabriquer un vecteur d'expression permettant l'expression de *orf28c*.

25 Il est également montré dans cet exemple que la surexpression du gène *orf28c* dans la souche OSC2 conduit à l'amélioration de la production en spiramycines de cette souche. La surexpression du gène *orf28c* peut être obtenue en augmentant le nombre de copies de ce gène et/ou en plaçant un promoteur plus actif que le promoteur sauvage. De façon préférée, la surexpression du gène *orf28c* est obtenue en apportant dans le

30 microorganisme une construction d'ADN recombinant, permettant la surexpression de

ce gène. De façon préférée, cette construction d'ADN recombinant augmente le nombre de copie du gène *orf28c* et permet d'obtenir une surexpression du gène *orf28c*. Dans cette construction d'ADN recombinant, la séquence codante de *orf28c* peut être placée sous le contrôle d'un promoteur plus actif que le promoteur sauvage. On peut citer à titre
5 d'illustration le promoteur ptrc actif chez *Streptomyces ambofaciens* (Amann, E. et al., 1988) ainsi que le promoteur ermE*. Ainsi, de façon préférée, la ou les copies du gène *orf28c* apportée sont placées sous le contrôle du promoteur ermE* comme c'est le cas dans la construction pSPM75 (cf. exemple 24).

Un autre aspect de l'invention concerne un procédé de production de macrolides
10 utilisant les microorganismes décrits au paragraphe précédent. Ce procédé consiste à cultiver dans un milieu de culture approprié un microorganisme défini dans le paragraphe précédent, récupérer le milieu de culture conditionné ou un extrait cellulaire, séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait
15 cellulaire obtenu à l'étape précédente le ou les macrolides produits. Les conditions de culture de tels microorganismes pourront être déterminées selon des techniques bien connues de l'homme du métier. Le milieu de culture pourra par exemple être le milieu MP5 ou le milieu SL11 pour les *Streptomyces* et notamment pour *Streptomyces ambofaciens* (Pernodet et al., 1993). L'homme du métier pourra notamment se référer à l'ouvrage de Kieser et al. 2000 en ce qui concerne la culture des *Streptomyces*. Le ou les
20 macrolides produits peuvent être récupérés par toutes techniques connues de l'homme du métier. L'homme du métier pourra se référer, par exemple, aux techniques enseignées dans le brevet américain US 3,000,785 et plus particulièrement aux méthodes d'extraction des spiramycines décrites dans ce brevet. Préférentiellement, les microorganismes utilisés dans un tel procédé sont des bactéries du genre *Streptomyces*.
25 Plus préférentiellement, le macrolide en question est la spiramycine et les microorganismes mutants améliorés pour leur production en spiramycines sont des souches de *S. ambofaciens*.

Un autre aspect de l'invention concerne l'utilisation d'une séquence et/ou d'un vecteur selon l'invention pour la préparation d'antibiotiques hybrides. En effet, les

polynucléotides selon l'invention peuvent servir à l'obtention de microorganismes exprimant une ou plusieurs protéines mutantes donnant lieu à une modification dans la spécificité du substrat ou encore être exprimés dans de multiples microorganismes producteurs d'antibiotiques dans le but de générer des antibiotiques hybrides. Ainsi les

5 polynucléotides selon l'invention peuvent permettre, par transfert des gènes entre microorganismes producteurs, de fabriquer des antibiotiques hybrides ayant des propriétés pharmacologiquement intéressantes (Hopwood *et al.*, 1985a, Hopwood *et al.*, 1985b, Hutchinson *et al.*, 1989). Le principe par lequel l'ingénierie génétique peut conduire à la production d'antibiotiques hybrides a été tout d'abord proposé par

10 Hopwood (Hopwood 1981). Il a ainsi été proposé que les enzymes participant à la biosynthèse d'antibiotiques acceptent souvent des substrats reliés structurellement, mais différents de leur substrat naturel. Il est généralement admis (Hopwood 1981, Hutchinson 1988, Robinson 1988) que les enzymes codées par les gènes de la voie de biosynthèse d'antibiotiques ont une spécificité de substrat moins stricte que les enzymes

15 du métabolisme primaire. Il a ainsi pu être montré qu'un grand nombre de substrats non-naturels sont transformés par des microorganismes producteurs d'antibiotique, leur mutants ou des enzymes purifiées de la voie de biosynthèse de ces antibiotiques (Hutchinson 1988). En utilisant cet enseignement, l'homme du métier pourra construire des microorganismes exprimant une ou plusieurs protéines mutantes donnant lieu à une

20 modification dans la spécificité du substrat dans le but de générer des antibiotiques hybrides.

L'invention concerne également l'utilisation d'au moins un polynucléotide et/ou au moins un ADN recombinant et/ou au moins un vecteur d'expression et/ou au moins un polypeptide et/ou au moins une cellule hôte selon l'invention pour réaliser une ou

25 plusieurs bioconversions. Ainsi l'invention permet de construire des souches bactériennes ou fongiques dans lesquelles on exprime, sous le contrôle de signaux d'expression appropriés, une ou plusieurs protéines selon l'invention. De telles souches peuvent alors être utilisées pour réaliser une ou des bioconversions. Ces bioconversions peuvent se faire soit à l'aide de cellules entières, soit à l'aide d'extraits acellulaires des

30 dites cellules. Ces bioconversions peuvent permettre de transformer une molécule en une

forme dérivée, avec une enzyme d'une voie de biosynthèse. Par exemple, Carreras *et al.* décrit l'utilisation de souche de *Saccharopolyspora erythraea* et de *Streptomyces coelicolor* pour la production de nouveaux dérivés d'érythromycines (Carreras *et al.*, 2002). Walczar *et al.* décrit l'utilisation d'une P450 mono-oxygénase de *Streptomyces* pour la bioconversion de désacétyladriamycine (un analogue de l'anthracycline) en de nouvelles anthracyclines (Walczar *et al.*, 2001). Olonao *et al.* décrit l'utilisation d'une souche de *Streptomyces lividans* modifiée pour la bioconversion de rhodomycinone-epsilon en rhodomycine D (Olonao *et al.*, 1999). L'homme du métier pourra appliquer ce principe à tout intermédiaire de biosynthèse.

10 L'invention concerne également un ADN recombinant caractérisé en ce qu'il comprend :

- un polynucléotide susceptible d'être obtenu par amplification en chaîne par polymérase en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, plus préférentiellement il s'agit d'un polynucléotide de séquence SEQ ID N° 141,
- ou un fragment d'au moins 10, 12, 15, 18, 20 à 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1460, 1470, 1480, 1490 ou 1500 nucléotides consécutifs de ces polynucléotides.

Préférentiellement ; cet ADN recombinant est un vecteur. Encore plus préférentiellement, le vecteur est choisi parmi les bactériophages, les plasmides, les phagemides, les vecteurs intégratifs, les fosmides, les cosmides, les vecteurs navettes, les BAC (Bacterial Artificial Chromosome) ou les PAC (P1-derived Artificial Chromosome). A titre illustratif on peut citer comme bactériophages le phage lambda et le phage M13. Comme plasmides on peut citer les plasmides se répliquant chez *E. coli* par exemple pBR322 et ses dérivés, pUC18 et ses dérivés, pUC19 et ses dérivés, pGB2 et ses dérivés (Churchward G. *et al.*, 1984), pACYC177 (numéro d'accès

GenBank : X06402) et ses dérivés, pACYC184 (numéro d'accès GenBank : X06403) et ses dérivés. On peut également citer les plasmides se répliquant chez *Streptomyces* comme par exemple pIJ101 et ses dérivés, pSG5 et ses dérivés, SLP1 et ses dérivés, SCP2* et ses dérivés (Kieser *et al.* 2000). Comme phagemides on peut citer à titre
5 illustratif pBluescript II et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)), pGEM-T et ses dérivés (commercialisé par la société Promega (Madison, Wisconsin, USA)), λ ZAPII et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)). Comme vecteurs intégratifs on peut citer à titre illustratif, les vecteurs intégratifs
10 chez *Streptomyces* comme par exemple ceux dérivés de SLP1 (Kieser *et al.*, 2000), ceux dérivés de pSAM2 (Kieser *et al.*, 2000), les vecteurs utilisant les systèmes d'intégration des phages PhiC31 (Kieser *et al.*, 2000) (par exemple pSET152 (Bierman *et al.*, 1992)) ou de VWB (Van Mellaert L, *et al.* 1998), ainsi que les vecteurs utilisant le système d'intégration de IS117 (Kieser *et al.* 2000). Comme
15 fosmides on peut citer à titre illustratif le fosmide pFOS1 (commercialisé par la société New England Biolabs Inc., Beverly, Massachusetts, USA) et ses dérivés. Comme cosmides on peut citer à titre illustratif le cosmide SuperCos et ses dérivés (commercialisés notamment par la société Stratagene (LaJolla, Californie, USA)), le cosmide pWED15 (Wahl *et al.*, 1987) et ses dérivés. Comme vecteurs navettes on
20 peut citer à titre illustratif les plasmides navettes *E. coli* / *Streptomyces* comme par exemple pIJ903 et ses dérivés, la série des plasmides pUWL, pCAO106, pWHM3, pOJ446 et leurs dérivés (Kieser *et al.* 2000), les BAC navettes *E. coli* / *Streptomyces* comme par exemple ceux décrits dans la demande de brevet WO 01/40497. Comme BAC (Bacterial Artificial Chromosome) on peut citer à titre illustratif le BAC
25 pBeloBAC11 (numéro d'accès GenBank : U51113). Comme PAC (P1-derived Artificial Chromosome) on peut citer à titre illustratif le vecteur pCYPAC6 (numéro d'accès GenBank : AF133437). Plus préférentiellement, cet ADN recombinant est un vecteur d'expression. Les vecteurs d'expression utilisables dans différents systèmes d'expression sont bien connus de l'homme du métier, en ce qui concerne les cellules
30 procaryotes, on peut citer à titre illustratif les vecteurs d'expression chez *E. coli* par

exemple, de la famille pET commercialisés par la société Stratagene (LaJolla, Californie, USA), les vecteurs de la famille GATEWAY commercialisés par la société Invitrogen (Carlsbad, Californie, USA), les vecteurs de la famille pBAD commercialisés par la société Invitrogen (Carlsbad, Californie, USA), les vecteurs de la famille pMAL commercialisés par la société New England Biolabs Inc. (Beverly, Massachusetts, USA), les vecteurs d'expression inductibles par le rhamnose cités dans la publication Wilms B *et al.*, 2001 et leurs dérivés, on peut également citer les vecteurs d'expression chez *Streptomyces* comme par exemple les vecteurs pIJ4123, pIJ6021, pPM927, pANT849, pANT 850, pANT 851, pANT1200, pANT1201, pANT1202 et leurs dérivés (Kieser *et al.*, 2000). En ce qui concerne les cellules de levures, on peut citer à titre illustratif le vecteur pESC commercialisé par la société Stratagene (LaJolla, Californie, USA). En ce qui concerne le système d'expression baculovirus permettant une expression dans des cellules d'insectes on peut citer à titre illustratif le vecteur BacPAK6 commercialisé par la société BD Biosciences Clontech, (Palo Alto, Californie, USA). En ce qui concerne les cellules de mammifères, on peut citer à titre illustratif des vecteurs comprenant le promoteur des gènes précoces du virus CMV (Cytomegalovirus) (par exemple le vecteur pCMV et ses dérivés commercialisés par la société Stratagene (LaJolla, Californie, USA), ou le promoteur des gènes immédiats (SV40 early promoter) du virus simien vacuolisant SV40 (par exemple le vecteur pSG5 commercialisé par la société Stratagene (LaJolla, Californie, USA). Un autre aspect de l'invention concerne des cellules hôtes dans laquelle a été introduit au moins un ADN recombinant décrit dans ce paragraphe.

Un autre aspect de l'invention concerne un procédé de production d'un polypeptide caractérisé en ce que ledit procédé comprend les étapes suivantes:

- a) transformer une cellule hôte par au moins un vecteur d'expression tel que décrit dans le paragraphe ci-dessus;
- b) cultiver, dans un milieu de culture approprié, ladite cellule hôte;
- c) récupérer le milieu de culture conditionné ou un extrait cellulaire;
- d) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape c), ledit polypeptide ;

e) le cas échéant, caractériser le polypeptide recombinant produit.

Le polypeptide recombinant ainsi produit peut être purifié par passage sur une série appropriée de colonnes de chromatographie, selon les méthodes connues de l'homme de l'art et décrites par exemple dans F. Ausubel et al (2002). On peut citer à titre illustratif la technique du « Tag-Histidine », qui consiste à ajouter une courte séquence poly-histidine au polypeptide à produire, ce dernier pouvant ensuite être purifié sur colonne de nickel. Ce polypeptide peut également être préparé par les techniques de synthèse *in vitro*. A titre illustratif de telles techniques, le polypeptide pourra être préparé grâce au système « rapid translation system (RTS) », commercialisé notamment par la société Roche Diagnostics France S.A, Meylan, France.

Un autre aspect de l'invention concerne un microorganisme producteur d'au moins une spiramycine caractérisé en ce qu'il surexprime :

-un gène susceptible d'être obtenu par amplification en chaîne par polymérase (PCR) en utilisant le couple d'amorces de séquence suivante : 5'

15 AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5'

GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme

matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, plus préférentiellement il s'agit du gène de séquence codante SEQ ID N° 141,

- ou un gène dérivé de ceux-ci en raison de la dégénérescence du code génétique.

20 Un exemple de séquence d'un tel gène est donné en SEQ ID N° 111 (ADN), cependant cette séquence est partielle puisqu'elle ne comprend pas la partie 3' de la séquence codante correspondante. La traduction en protéine de cette partie de séquence codante est donnée en SEQ ID N° 112. L'homme du métier pourra aisément la compléter notamment en utilisant l'enseignement présenté dans l'exemple 24. Il est ainsi présenté dans cet exemple une méthode pour cloner le gène *orf28c* et pour fabriquer un vecteur d'expression permettant l'expression de *orf28c*. Il est également montré dans cet exemple que la surexpression du gène *orf28c* dans la souche OSC2 conduit à l'amélioration de la production en spiramycines de cette souche. De façon préférée, le microorganisme surexprimant

- un gène susceptible d'être obtenu par amplification en chaîne par polymérase (PCR) en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme
5 matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, plus préférentiellement il s'agit du gène de séquence codante SEQ ID N° 141,
- ou un gène dérivé de celui-ci en raison de la dégénérescence du code génétique
est une bactérie du genre *Streptomyces*, de manière encore plus préférée, il s'agit d'une
bactérie de l'espèce *Streptomyces ambofaciens*. De façon préférée, la surexpression
10 dudit gène est obtenue par transformation dudit microorganisme par un vecteur d'expression, de manière tout à fait préférée, la souche de microorganisme est la souche OSC2/pSPM75(1) ou de la souche OSC2/pSPM75(2) déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 6 octobre 2003 sous le numéro d'enregistrement
15 I-3101.

Un autre aspect de l'invention concerne un procédé de production de spiramycine(s) utilisant les microorganismes décrits au paragraphe précédent. Ce procédé consiste à cultiver dans un milieu de culture approprié un microorganisme défini dans le paragraphe précédent, récupérer le milieu de culture conditionné ou un
20 extrait cellulaire, séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape précédente le ou les spiramycines produites. Les conditions de culture de tels microorganismes pourront être déterminées selon des techniques bien connues de l'homme du métier. Le milieu de culture pourra par exemple être le milieu MP5 ou le milieu SL11 pour les *Streptomyces* et notamment
25 pour *Streptomyces ambofaciens* (Pernodet et al., 1993). L'homme du métier pourra notamment se référer à l'ouvrage de Kieser et al. 2000 en ce qui concerne la culture des *Streptomyces*. Le ou les spiramycines produites peuvent être récupérés par toutes techniques connues de l'homme du métier. L'homme du métier pourra se référer, par exemple, aux techniques enseignées dans le brevet américain US 3,000,785 et plus
30 particulièrement aux méthodes d'extraction des spiramycines décrites dans ce brevet.

Préférentiellement, les microorganismes utilisés dans un tel procédé sont des bactéries du genre *Streptomyces*. Plus préférentiellement, les microorganismes sont des souches de *S. ambofaciens*.

Un autre aspect de l'invention concerne un vecteur d'expression caractérisé en ce que le polynucléotide de séquence SEQ ID N° 47 ou un polynucléotide dérivé de celui-ci en raison de la dégénérescence du code génétique est placé sous le contrôle d'un promoteur permettant l'expression de la protéine codée par le dit polynucléotide chez *Streptomyces ambofaciens*. Des exemples de vecteurs d'expression utilisables chez *Streptomyces* ont été donnés ci-dessus.

Préférentiellement un tel vecteur d'expression est caractérisé en ce qu'il s'agit du plasmide pSPM524 ou pSPM525.

Un autre aspect de l'invention concerne une souche de *Streptomyces ambofaciens* transformée par un vecteur défini dans le paragraphe précédent.

Un autre aspect de l'invention concerne un polypeptide caractérisé en ce que sa séquence comprend la séquence SEQ ID N° 112. L'invention est également relative à un polypeptide caractérisé en ce que sa séquence correspond à la séquence traduite de la séquence codante :

- d'un gène susceptible d'être obtenu par amplification en chaîne par polymérase (PCR) en utilisant le couple d'amorces de séquence suivante : 5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et 5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, plus préférentiellement il s'agit du gène de séquence codante SEQ ID N° 141,

-ou d'un gène dérivé de celui-ci en raison de la dégénérescence du code génétique.

Préférentiellement ces polypeptides sont exprimés à l'état naturel par une bactérie du genre *Streptomyces*, plus préférentiellement, ces polypeptides sont impliqués dans la biosynthèse des spiramycines.

Un autre aspect de l'invention concerne un vecteur d'expression permettant l'expression d'un polypeptide tel que défini dans le paragraphe précédent chez *Streptomyces ambofaciens*. Des exemples de vecteurs d'expression utilisables chez *Streptomyces* ont été donnés ci-dessus. Préférentiellement, le vecteur d'expression en question est le plasmide pSPM75.

LISTE DES FIGURES

Figure 1 : Structure chimique des spiramycines I, II et III.

Figure 2 : Cosmides utilisés pour le séquençage de la région.

Figure 3 : Organisation d'un groupe de gènes impliqués dans la voie de biosynthèse des spiramycines.

Figure 4 : Voie de biosynthèse proposée du mycarose.

Figure 5 : Voie de biosynthèse proposée du mycaminose.

Figure 6 : Voie de biosynthèse proposée de la forosamine.

Figure 7 : Ordre préférentiel d'ajout des sucres dans la molécule de spiramycine et intermédiaires.

Figure 8 : Voie de biosynthèse proposée du méthoxymalonyl chez *S. ambofaciens*. Cette voie est proposée par analogie avec la biosynthèse du méthoxymalonyl chez *Streptomyces hygroscopicus* var. *ascomyceticus* (Wu, K. et al., 2000).

Figure 9 : Etapes conduisant à l'inactivation d'un gène:

- A) Clonage du gène cible dans un vecteur se répliquant chez *E. coli* mais pas chez *Streptomyces*;
- B) Insertion de la cassette de résistance dans le gène cible (par clonage ou recombinaison entre courtes séquences identiques);
- C) Introduction du plasmide chez *Streptomyces ambofaciens* (par transformation ou conjugaison avec *E. coli*) et sélection des clones ayant intégré la cassette puis criblage de clones ayant perdu la partie vecteur pour avoir un remplacement de gène;
- D) Région du chromosome de la souche mutante dans laquelle le gène cible est inactivé par remplacement de gène.

Figure 10 : Etapes optionnelles suivant l'inactivation d'un gène selon la méthode décrite en figure 9, pouvant être conduites si une cassette excisable a été employée pour interrompre le gène cible :

5 E) Introduction dans la souche mutante du plasmide pOSV508 portant les gènes *xis* et *int* de pSAM2 dont les produits vont permettre l'excision efficace par recombinaison spécifique de sites entre les séquences attL et attR bordant la cassette;

F) Obtention de clones ayant perdu la cassette excisable et sensibles à l'antibiotique à laquelle la cassette conférait la résistance;

10 G) Après croissance et sporulation sur milieu solide sans antibiotique le plasmide pOSV508 est perdu à haute fréquence. On peut ainsi obtenir des clones sensibles au thiostrepton dans lesquels le gène cible contient une délétion en phase. La séquence du gène cible délété peut être contrôlée par amplification PCR et séquençage du produit PCR.

15 Figure 11 : Amplification de la cassette excisable dans le but de l'utiliser pour une expérience de recombinaison homologue. La technique de recombinaison homologue grâce à de courtes séquences homologues a été décrite par (Chaveroche *et al.*, 2000). Les 39 ou 40 déoxy-nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence du gène à inactiver et les 20 déoxy-nucléotides situés le plus en 3' correspondent à la séquence d'une des extrémités de la cassette excisable.

Figure 12 : Obtention d'une construction pour l'inactivation d'un gène cible grâce à la technique décrite par (Chaveroche *et al.*, 2000).

25 Figure 13 : Carte du plasmide pWHM3. Strepto ori : origine de répllication *Streptomyces*.

Figure 14 : Carte du plasmide pOSV508.

Figure 15 : Exemple de structure d'une cassette excisable. Celle-ci est constituée de la cassette *Ωhyg* (Blondelet-Rouault *et al.*, 1997) bordée des sites attR et attL (Raynal *et*

al., 1998), entre lesquels un événement de recombinaison permettra l'excision de la cassette, grâce à l'expression des gènes *xis* et *int*.

Figure 16 : Carte du plasmide pBXL1111.

Figure 17 : Carte du plasmide pBXL1112.

- 5 Figure 18 : Test microbiologique de production de spiramycine, basé sur la sensibilité d'une souche de *Micrococcus luteus* à la spiramycine. La souche de *Micrococcus luteus* utilisée est une souche naturellement sensible à la spiramycine mais résistante à la congocidine. Les différentes souches de *Streptomyces* à tester ont été cultivées en erlenmeyers de 500 ml contenant 70 ml de milieu MP5, inoculés à une concentration
- 10 initiale de $2.5 \cdot 10^6$ spores/ml et mises à pousser à 27°C sous agitation orbitale de 250 rpm. Des prélèvements de moûts de fermentation ont été réalisés après 48, 72 et 96 heures de culture et centrifugés. Une dilution au dixième de ces surnageants dans du milieu de culture stérile est utilisée pour le test. La souche indicatrice de *Micrococcus luteus* résistante à la congocidine mais sensible à la spiramycine (Gourmelen *et al.*,
- 15 1998) a été cultivée dans une boîte carrée de 12x12 cm. Des disques en papier Whatman AA ont été imbibés de 70 µl de la dilution au dixième de chaque surnageant et déposés sur la surface de la boîte. Des disques imbibés d'une solution de spiramycine de différentes concentrations (2-4-8 µg/ml dans du milieu de culture MP5) sont utilisés comme gamme étalon. Les boîtes sont incubées à 37°C pendant 24 à 48h. Si le disque
- 20 contient de la spiramycine, celle-ci diffuse dans l'agar et inhibe la croissance de la souche indicatrice de *Micrococcus luteus*. Cette inhibition crée un « halo » autour du disque, ce halo reflétant la zone où la souche de *Micrococcus luteus* n'a pas poussé. La présence de ce halo est donc une indication de la présence de spiramycine et permet de déterminer si la souche de *S. ambofaciens* en question est productrice ou non productrice
- 25 de spiramycine. Une comparaison avec les diamètres d'inhibition obtenus pour la gamme étalon permet d'obtenir une indication de la quantité de spiramycine produite par cette souche.

Figure 19 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche OSC2.

Figure 20 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche SPM501.

Figure 21 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche SPM502.

5 Figure 22 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche SPM507.

Figure 23 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche SPM508.

10 Figure 24 : Chromatogramme CLHP du surnageant filtré du milieu de culture de la souche SPM509.

Figure 25 : Alignement de la protéine Orf3 (SEQ ID N° 29) avec la protéine TylB (SEQ ID N° 87) de *S. fradiae* réalisé grâce au programme FASTA.

Figure 26 : Alignement de la protéine MdmA de *S. mycarofaciens* (SEQ ID N° 88) avec la protéine SrmD (SEQ ID N° 16) réalisé grâce au programme FASTA.

15 Figure 27 : Exemple de séquences des sites résiduels après excision de la cassette excisable. En gras est indiqué le site att26 minimum tel que défini dans Raynal *et al.*, 1998. La séquence de la phase 1 (att1) de 33 nucléotides est présentée en SEQ ID N° 104, la séquence de la phase 2 (att2) de 34 nucléotides est présentée en SEQ ID N° 105, la séquence de la phase 3 (att3) de 35 nucléotides est présentée en SEQ ID N° 95.

20 Figure 28 : Représentation schématique du gène *orf10*, localisation des amorces PCR utilisées et construction obtenue avec chaque couple d'amorces.

Figure 29 : Construction de la cassette *pac-oriT*.

Figure 30 : Carte du cosmide pWED2.

25 Figure 31 : Représentation schématique du groupe de gènes impliqués dans la voie de biosynthèse des spiramycines et de la localisation des trois sondes utilisées pour isoler les cosmides de la banque d'ADN génomique de la souche OSC2 de *Streptomyces ambofaciens* chez *E. coli* (cf. exemple 19). Figure 32 : Localisation des inserts des cosmides isolés de la banque d'ADN génomique de la souche OSC2 de *Streptomyces ambofaciens* chez *E. coli* (cf. exemple 19).

Nouvelle banque d'ADN cosmidique. Figure 33 : Sous clonage du fragment *PstI-PstI* du cosmide pSPM36 (insert du plasmide (pSPM58), sous clonage du fragment *StuI-StuI* du cosmide pSPM36 (insert du plasmide pSPM72) et sous clonage d'un fragment *EcoRI-StuI* (insert du plasmide pSPM73).

- 5 Figure 34 : Localisation des phases ouvertes de lecture indentifiées dans l'insert *PstI-PstI* du plasmide pSPM58 et dans l'insert *EcoRI-StuI* du plasmide pSPM73.

Figure 35 : Superposition des chromatogrammes CLHP du surnageant filtré du milieu de culture de la souche OS49.67 réalisés à 238 et 280nm (haut) et spectres UV des molécules élués à 33,4 minutes et 44,8 minutes (bas).

- 10 Figure 36 : Structure moléculaire des molécules platénolide A et platénolide B.

Figure 37 : Organisation du groupe de gènes impliqués dans la voie de biosynthèse des spiramycines.

Figure 38 : Structure moléculaire d'un intermédiaire de biosynthèse produit par la souche SPM507.

- 15 Figure 39 : Structure d'un intermédiaire de biosynthèse produit par une souche de *S. ambofaciens* de génotype *orf6*::att1Ωhyg+* obtenue à partir d'une souche surproduisant les spiramycines. L'insertion de la cassette *att1Ωhyg+* dans *orf6** exerce un effet polaire empêchant l'expression de *orf5**.

- Figure 40 : Structure moléculaire des molécules platénolide A + mycarose et platénolide B + mycarose produites par la souche OS49.67
- 20

Figure 41 : Sous clonage d'un fragment *PstI-PstI* du cosmide pSPM36 (insert du plasmide (pSPM79)), localisation des phases ouvertes de lecture indentifiées dans l'insert *PstI-PstI* du plasmide pSPM79 et localisation de la séquence SEQ ID N° 140.

La présente invention est illustrée à l'aide des exemples qui suivent, qui doivent être considérés comme illustratifs et non limitatifs.

En résumé, les gènes de la voie de biosynthèse des spiramycines ont été isolés à partir d'une banque d'ADN génomique de *Streptomyces ambofaciens*. Cette banque a été
5 obtenue par digestion partielle de l'ADN génomique de *Streptomyces ambofaciens*, par l'enzyme de restriction *Bam*HI. De larges fragments d'ADN, de 35 à 45 kb en moyenne, ont été clonés dans le cosmide pWED1 (Gourmelen *et al.*, 1998) dérivé du cosmide pWED15 (Wahl *et al.*, 1987). Ces cosmides ont été introduits chez *E. coli* grâce à des
10 particules phagiques. La banque ainsi obtenue a été hybridée avec une sonde (de séquence SEQ ID N° 86) correspondant à une partie du gène *tylB* de *S. fradiae* (Merson-Davies & Cundliffe, 1994, numéro d'accès GenBank : U08223). Après hybridation, un cosmide sur les 4 s'hybridant avec la sonde a été plus particulièrement sélectionné. Ce cosmide nommé pOS49.1 a ensuite été digéré par *Sac*I et un fragment de 3,3 kb contenant la région s'hybridant avec la sonde a été sous cloné dans le vecteur pUC19 et
15 séquencé. Quatre phases ouvertes de lectures ont été identifiées et l'une d'entre elles code une protéine (SEQ ID N° 29) présentant de fortes similitudes de séquence avec la protéine TylB de *S. fradiae* (SEQ ID N° 87) (cf. figure 25). Ce gène a été nommé *orf3* (SEQ ID N° 28) et a été inactivé chez *S. ambofaciens*. Il a pu être démontré que les clones inactivés dans le gène *orf3* ne produisent plus de spiramycines. Ceci montre
20 l'implication du gène *orf3* ou de gènes situés en aval dans la biosynthèse des spiramycines.

Une fois cette confirmation obtenue, une plus grande région du cosmide pOS49.1 a été séquencée, de part et d'autre du fragment *Sac*I précédemment étudié. Ainsi, à partir du cosmide pOS49.1, la séquence d'une région comportant sept phases ouvertes de
25 lecture entières et deux autres incomplètes, situées de part et d'autre de ces sept phases ouvertes complètes, a pu être obtenue. Grâce à une recherche dans les bases de données, il a pu être montré que l'une des phases ouvertes de lecture incomplète correspondait au locus *srmG* (région codant une enzyme appelée « polyketide synthase » (PKS)). Les gènes correspondants ont été clonés par Burgett S. *et al.* en 1996 (Brevet américain US

5,945,320). Par ailleurs, les autres phases ouvertes de lecture, les sept ORFs entières nommées : *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7* (SEQ ID N° 23, 25, 28, 30, 34, 36, 40) et le début de la huitième ORF nommé *orf8* (séquence SEQ ID N° 43) n'ont pas été retrouvées dans les bases de données.

5 Dans un deuxième temps et dans le but de cloner d'autres gènes impliqués dans la biosynthèse des spiramycines, d'autres cosmides comportant des fragments du génome de *S. ambofaciens* dans cette même région ont été isolés. Pour cela une nouvelle série d'hybridations sur colonies a été effectuée en utilisant trois sondes. La première sonde correspond à un fragment d'ADN de 3,7kb contenant *orf1*, *orf2* et le début de *orf3*, sous
10 cloné à partir de pOS49.1. La deuxième sonde correspond à un fragment de 2 kb d'ADN contenant une partie de *orf7* et une partie de *orf8* et sous cloné à partir de pOS49.1. Une troisième sonde a également été utilisée. Cette dernière correspond à un fragment d'ADN de 1.8 kb contenant le gène *srmD*. Le gène *srmD* est un gène isolé de *S. ambofaciens* capable de conférer la résistance à la spiramycine. En effet, des travaux antérieurs
15 avaient permis le clonage de plusieurs déterminants de résistance de *S. ambofaciens*, conférant la résistance à la spiramycine à une souche de *S. griseofuscus* (souche sensible à la spiramycine) (Pernodet et al., 1993) (Pernodet et al., 1999). Pour l'isolement de gènes de résistances, une banque cosmidique de l'ADN génomique de la souche *S. ambofaciens* avait été réalisée dans le cosmide pKC505 (Richardson MA et al., 1987).
20 Ce pool de cosmides avait été introduit chez *S. griseofuscus*, naturellement sensible à la spiramycine. Cinq cosmides capables de conférer la résistance à l'apramycine et la spiramycine chez *S. griseofuscus* avaient ainsi été obtenus. Parmi ces 5 cosmides, un cosmide nommé pOS44.1 possède dans son insert le gène *srmD* qui code une protéine présentant une certaine similitude avec la protéine codée par le gène *mdmA* de
25 *Streptomyces mycarofaciens* et impliquée dans la résistance à la midécamycine chez cet organisme producteur (Hara et al., 1990, numéro d'accès Genbank : A60725) (figure 26). La troisième sonde utilisée pour localiser les gènes de biosynthèse de spiramycine a été un insert d'environ 1.8kb comprenant le gène *srmD*.

Ces trois sondes ont été utilisées pour hybrider la banque d'ADN génomique décrite ci-dessus et ont permis de choisir deux cosmides (pSPM7 et pSPM5) susceptibles de contenir les inserts les plus longs et n'ayant pas de bandes communes. Le cosmide pSMP5 hybridait avec la première et la deuxième sonde, mais n'hybridait pas
5 avec la troisième sonde, alors que le cosmide pSPM7 hybridait avec la troisième sonde seulement. Ces deux cosmides ont été séquencés en totalité par la technique du séquençage en aveugle (« shotgun sequencing »). Les séquences des inserts de ces deux cosmides : pSPM7 et pSPM5 ont pu être assemblées car bien que ne se chevauchant pas, chacun des inserts comprenait une séquence connue à l'une de ses extrémités. En effet,
10 chacun de ces inserts comportait un fragment de la séquence d'un des gènes codant une enzyme appelée « polyketide synthase » (PKS). Ces 5 gènes ont été clonés par Burgett S. et al. en 1996 (Brevet américain US 5,945,320) (cf. figure 2). Ainsi, il a pu être déterminé une séquence unique d'ADN génomique de *S. ambofaciens*. Une séquence de 30943 nucléotides partant, en 5', d'un site *EcoRI* situé dans le premier gène PKS et
15 allant jusqu'à un site *BamHI* en 3' est présentée en SEQ ID N° 1. Cette séquence correspond à la région amont des gènes de la PKS (cf. figure 2 et 3). Une deuxième région de 11171 nucléotides, partant d'un site *PstI* en 5' et allant jusqu'à un site *BstEII* en 3' situé dans le cinquième gène de la PKS est présentée en SEQ ID N° 2. Cette région est la région aval des gènes de la PKS (l'aval et l'amont étant définis par l'orientation
20 des 5 gènes de PKS tous orientés dans le même sens) (cf. figures 2 et 3).

Dans un troisième temps et dans le but de cloner d'autres gènes impliqués dans la biosynthèse des spiramycines, d'autres cosmides comportant des fragments du génome de *S. ambofaciens* dans cette même région ont été isolés (cf. exemple 18 et 19).

EXEMPLE 1 : Construction d'une banque d'ADN génomique de la souche ATCC23877 de *Streptomyces ambofaciens* chez *E. coli*

1.1. Extraction de l'ADN génomique de la souche ATCC23877 de *Streptomyces ambofaciens*.

5 La souche ATCC23877 de *Streptomyces ambofaciens* (accessible notamment auprès de l'American Type Culture Collection (ATCC) (Manassas, Virginie, USA), sous le numéro 23877) a été cultivée en milieu YEME (Yeast Extract-Malt Extract) ((Kieser, T, *et al.*, 2000) et l'ADN génomique de cette souche a été extrait et purifié selon les techniques standards de lyse et précipitation (Kieser T. *et al.*, 2000).

10 1.2. Construction de la banque d'ADN génomique

15 L'ADN génomique de la souche ATCC23877 de *Streptomyces ambofaciens* tel qu'isolé ci-dessus a été digéré partiellement par l'enzyme de restriction *Bam*HI de manière à obtenir des fragments d'ADN de taille comprise entre environ 35 et 45 kb. Ces fragments ont été clonés dans le cosmide pWED1 (Gourmelen *et al.*, 1998) digéré au préalable par *Bam*HI. Le cosmide pWED1 est dérivé du cosmide pWE15 (Wahl, *et al.*, 1987) par délétion du fragment *Hpa*I-*Hpa*I de 4.1 kb contenant le module d'expression actif chez les mammifères (Gourmelen *et al.*, 1998). Le mélange de ligation a ensuite été encapsidé *in vitro* dans des particules de phages lambda grâce au système « Packagene® Lambda DNA packaging system » commercialisé par la société Promega en suivant les recommandations du fabricant. Les particules phagiques obtenues ont été utilisées pour infecter la souche SURE® d'*E. coli* commercialisée par la société Stratagene (LaJolla, Californie, USA). La sélection des clones a été effectuée sur milieu LB + ampicilline (50 µg/ml) puisque le cosmide pWED1 confère la résistance à l'ampicilline.

EXEMPLE 2 : Isolement et caractérisation de gènes impliqués dans la biosynthèse de spiramycines chez *Streptomyces ambofaciens*

2.1 Hybridation sur colonie des clones d'*E. coli* de la banque génomique de *Streptomyces ambofaciens* ATCC23877

5 Environ 2000 clones d'*E. coli* de la banque obtenue ci-dessus, ont été transférés sur filtre pour hybridation sur colonies. La sonde utilisée pour l'hybridation est un fragment *NaeI-NaeI* d'ADN (SEQ ID N° 86) comportant une partie du gène *tylB* de *Streptomyces fradiae*. Ce fragment correspond aux nucléotides 2663 à 3702 du fragment d'ADN décrit par Merson-Davies, L.A. & Cundliffe E. (Merson-Davies, L.A. & Cundliffe E., 1994, numéro d'accès GenBank : U08223), dans lequel la séquence codante du gène *tylB* correspond aux nucléotides 2677 à 3843.

15 Le fragment *NaeI-NaeI* d'ADN portant une partie du gène *tylB* de *Streptomyces fradiae* (SEQ ID N° 86) a été marquée au ³²P par la technique du « random priming » (Kit commercialisé par la société Roche) et utilisé comme sonde pour l'hybridation des 2000 clones de la banque, transférés sur filtre. Les membranes utilisées sont des membranes nylon Hybond N commercialisées par la société Amersham (Amersham Biosciences, Orsay, France) et l'hybridation a été effectuée à 55°C dans le tampon décrit par Church & Gilbert (Church & Gilbert, 1984). Un lavage a été effectué en 2X SSC à 55°C pendant 15 minutes et deux lavages successifs en 0.5X SSC ont ensuite été effectués à 55°C, d'une durée de 15 minutes chacun. Dans ces conditions d'hybridation et de lavage, 4 clones parmi les 2000 hybridés présentaient un fort signal d'hybridation. Ces 4 clones ont été cultivés en milieu LB+ ampicilline (50 µg/ml) et les 4 cosmides correspondants ont été extraits par lyse alcaline standard (Sambrook *et al.*, 1989). Il a ensuite été vérifié que l'hybridation était bien due à un fragment d'ADN présent dans l'insert de ces quatre cosmides. Pour cela, les cosmides ont été digérés indépendamment par plusieurs enzymes (*Bam*HI, *Pst*I et *Sac*I). Les produits de digestions ont été séparés sur gel d'agarose, transférés sur membrane Nylon et hybridés par le fragment *NaeI-NaeI* d'ADN comportant une partie du gène *tylB* de *Streptomyces fradiae* (cf. ci-dessus) dans

les mêmes conditions que ci-dessus. Les quatres cosmides ont pu être validés et un de ces cosmides a été plus particulièrement retenu et a été nommé pOS49.1.

2.2 Vérification de l'implication de la région identifiée et séquençage de l'insert du cosmide pOS49.1

5 Plusieurs fragments de l'insert du cosmide pOS49.1 ont été sous-clonés et leurs séquences ont été déterminées. Le cosmide pOS49.1 a été digéré par l'enzyme *SacI* et il a été montré par Southern blot, dans les conditions décrites précédemment, qu'un fragment de 3,3 kb contient la région s'hybridant avec la sonde *tylB*. Ce fragment de 3,3 kb a été isolé par électroélution à partir d'un gel d'agarose 0.8% puis cloné dans le
10 vecteur pUC19 (numéro d'accès GenBank M77789) et séquencé. Le plasmide ainsi obtenu a été nommé pOS49.11. Quatre phases ouvertes de lecture présentant un usage des codons typique de *Streptomyces* ont pu être identifiées dans ce fragment (deux complètes et deux tronquées), grâce au programme FramePlot (Ishikawa J & Hotta K. 1999). Des comparaisons de séquences grâce au programme FASTA (cf. (Pearson W. R
15 & D. J. Lipman, 1988) et (Pearson W. R., 1990), accessible notamment auprès du centre de ressources INFOBIOGEN, Evry, France) ont permis de montrer que la protéine déduite de l'une de ces 4 phases ouvertes de lecture présente de fortes similitudes de séquence avec la protéine TylB (SEQ ID N° 87; Numéro d'accès GenBank : U08223) de *S. fradiae* (cf. figure 25). Cette protéine a été nommée Orf3 (SEQ ID N° 29).

20 Dans le but de tester si le gène correspondant (le gène *orf3* (SEQ ID N° 28)) était impliqué dans la biosynthèse des spiramycines chez *S. ambofaciens*, ce gène a été interrompu par une cassette Ω_{hyg} (Blondelet-Rouault M-H. *et al.*, 1997, numéro d'accès GenBank : X99315). Pour cela, le plasmide pOS49.11 a été digéré par l'enzyme *XhoI* et le fragment contenant les quatre phases ouvertes de lecture (deux complètes et deux
25 tronquées, dont *orf3* en entier) a été sous-cloné au niveau du site *XhoI* du vecteur pBC SK+ commercialisé par la société Stratagene (LaJolla, Californie, USA). Le plasmide ainsi obtenu a été nommé pOS49.12. Dans le but d'inactiver *orf3*, un fragment *PmlI*-*BstEII* interne à *orf3* a été remplacé par la cassette Ω_{hyg} par clonage bout franc dans ce

dernier plasmide. Pour cela, le plasmide pOS49.12 a été digéré par les enzymes *PmlI* et *BstEII* dont le site unique se trouve dans la séquence codante du gène *orf3*. Les extrémités du fragment correspondant au vecteur ont été rendues bouts francs par un traitement par l'enzyme de Klenow (grand fragment de la DNA polymérase I). La cassette *Ωhyg* a été obtenue par digestion par l'enzyme *BamHI* du plasmide pHP45 *Ωhyg* (Blondelet-Rouault *et al.*, 1997, numéro d'accès GenBank : X99315). Le fragment correspondant à la cassette *Ωhyg* a été récupéré sur gel d'agarose et ses extrémités ont été rendues bouts francs par un traitement par l'enzyme de Klenow. Les deux fragments bouts francs ainsi obtenus (la cassette *Ωhyg* et le plasmide pOS49.12) ont été ligués et le produit de ligation a été utilisé pour la transformation de bactéries *E. coli*. Le plasmide ainsi obtenu a été nommé pOS49.14 et contient le gène *orf3* interrompu par la cassette *Ωhyg*.

L'insert du plasmide pOS49.14, sous la forme d'un fragment *XhoI-XhoI* dont les extrémités ont été rendues non cohésives par un traitement par l'enzyme de Klenow, a été cloné au niveau du site *EcoRV* du plasmide pOJ260 (le plasmide pOJ260 est un plasmide conjugatif capable de se répliquer chez *E. coli* mais incapable de se répliquer chez *S. ambifaciens* (Bierman M *et al.*, 1992). Ce plasmide confère la résistance à l'apramycine chez *E. coli* et *Streptomyces*). Le plasmide obtenu (insert du plasmide pOS49.14 cloné dans le plasmide pOJ260) a été nommé pOS49.16. Ce dernier a été transféré dans la souche ATCC23877 de *S. ambifaciens* par conjugaison à l'aide de la souche conjugative *E. coli* S17-1, comme décrit par Mazodier *et al.* (Mazodier *et al.*, 1989). La souche *E. coli* S17-1 est dérivée de la souche *E. coli* 294 (Simon *et al.*, 1983) (Simon, *et al.*, 1986). Des clones transconjugants possédant le marqueur de résistance à l'hygromycine porté par la cassette *Ωhyg* et ayant perdu le marqueur de résistance à l'apramycine porté par le vecteur pOJ260 ont pu être obtenus. Pour cela, après conjugaison, les clones sont sélectionnés pour leur résistance à l'hygromycine. Les clones résistants à l'hygromycine sont ensuite repiqués respectivement sur milieu avec hygromycine (antibiotique B) et sur milieu avec apramycine (antibiotique A) (cf. figure 9). Les clones résistants à l'hygromycine (HygR) et sensibles à l'apramycine (ApraS) sont en principe ceux où un double événement de recombinaison s'est produit et qui

possèdent donc le gène *orf3* interrompu par la cassette *Ωhyg*. Le remplacement de la copie sauvage de *orf3* par la copie interrompue a été vérifié par deux hybridations successives. Ainsi, l'ADN total des clones obtenus, a été digéré par différentes enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde
5 correspondant à la cassette *Ωhyg* (cf. ci-dessus)) pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde l'insert *XhoI-XhoI* du plasmide pOS49.11 contenant les quatre phases ouvertes de lecture (deux complètes et deux tronquées, dont *orf3* en entier). La vérification du génotype peut également être réalisée par toute méthode connue de
10 l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un des clones *orf3::Ωhyg* ainsi obtenu et dont le génotype a été vérifié, a été choisi et a été appelé OS49.16.

La production en spiramycine du clone OS49.16 ainsi obtenu a été testée grâce au test de production décrit plus bas (cf. exemple 15). Il a ainsi pu être démontré que cette
15 souche ne produit plus de spiramycine, confirmant l'implication de *orf3* et/ou des gènes situés en aval comme par exemple *orf4* dans la biosynthèse des spiramycines.

Une fois cette confirmation obtenue, une plus grande région du cosmide pOS49.1 a été séquencée, de part et d'autre du fragment *SacI* précédemment étudié. Ainsi, à partir du cosmide pOS49.1, la séquence d'une région comportant sept phases ouvertes de
20 lecture entières et deux autres incomplètes, situées de part et d'autre de ces sept phases ouvertes complètes, a pu être obtenue. Grâce à une recherche dans les bases de données, il a pu être montré que l'une des phases ouvertes de lecture incomplète correspondait au locus *srmG* (région codant une enzyme appelée « polyketide synthase » (PKS)). Les gènes correspondants ont été clonés par Burgett S. *et al.* en 1996 (Brevet américain US
25 5,945,320). Par ailleurs, les autres phases ouvertes de lecture: les sept ORFs entières nommées : *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7* (SEQ ID N° 23, 25, 28, 30, 34, 36 et 40) et le début de la huitième ORF nommée *orf8* (la séquence entière de cette orf est donnée en SEQ ID N° 43) n'ont pas été retrouvées dans les bases de données.

EXEMPLE 3 : Isolement et caractérisation d'autres gènes impliqués dans la biosynthèse de spiramycines chez *Streptomyces ambofaciens*

Dans un deuxième temps et dans le but de cloner d'autres gènes impliqués dans la biosynthèse des spiramycines, d'autres cosmides comportant des fragments du génome de *S. ambofaciens* dans cette même région ont été isolés. Pour cela il a été procédé à une nouvelle série d'hybridations sur colonies en utilisant trois sondes :

- La première sonde utilisée correspond à un fragment d'ADN *Bam*HI-*Pst*I de 3,7kb contenant un fragment du gène de la PKS (Les gènes correspondant à la PKS ont été clonés par Burgett S. *et al.* en 1996 (Brevet américain US 5,945,320)), *orf1*, *orf2* et le début de *orf3*, sous cloné à partir de pOS49.1 et allant d'un site *Bam*HI situé 1300 paires de bases en amont du site *Eco*RI définissant la position 1 de SEQ ID N° 1, jusqu'au site *Pst*I situé en position 2472 (SEQ ID N° 1). Ce fragment *Bam*HI-*Pst*I a été sous cloné à partir de pOS49.1 dans le plasmide pBC SK+, ce qui a permis d'obtenir le plasmide pOS49.28.

- La deuxième sonde utilisée correspond à un fragment d'ADN *Pst*I-*Bam*HI d'environ 2kb contenant un fragment d'*orf7* et d'*orf8*, sous cloné à partir de pOS49.1 et allant d'un site *Pst*I situé en position 6693 de SEQ ID N° 1 jusqu'au site *Bam*HI situé en position 8714 de SEQ ID N° 1. Ce fragment *Pst*I-*Bam*HI a été sous cloné à partir de pOS49.1 dans le plasmide pBC SK+, ce qui a permis d'obtenir le plasmide pOS49.76.

- Une troisième sonde a également été utilisée. Cette dernière correspond à un fragment d'ADN de 1.8kb *Eco*RI-*Hind*III contenant le gène *srmD*. Le gène *srmD* est un gène isolé de *S. ambofaciens* capable de conférer la résistance à la spiramycine. En effet, des travaux antérieurs avaient permis le clonage de plusieurs déterminants de résistance de *S. ambofaciens*, conférant la résistance à la spiramycine à une souche de *S. griseofuscus* (souche sensible à la spiramycine) (Pernodet *et al.*, 1993) (Pernodet *et al.*, 1999). Pour isoler des gènes de résistance, une banque cosmidique de l'ADN génomique de la souche *S. ambofaciens* ATCC23877 avait été réalisée dans le cosmide pKC505

(Richardson MA *et al.*, 1987). Pour cela l'ADN génomique de la souche *S. ambofaciens* ATCC23877 avait été digéré partiellement par *Sau3AI* de manière à obtenir des fragments de taille comprise entre environ 30 et 40 kb. L'ADN génomique ainsi digéré (3µg) avait été ligué avec 1µg du pKC505 préalablement digéré par l'enzyme *BamHI* (Pernodet *et al.*, 1999). Le mélange de ligation avait ensuite été encapsidé *in vitro* dans des particules phagiques. Les particules phagiques obtenues avaient été utilisées pour infecter la souche d'*E. coli* HB101 (accessible notamment auprès de l'American Type Culture Collection (ATCC) (Manassas, Virginie, USA), sous le numéro 33694). Environ 20 000 clones d'*E. coli* résistant à l'apramycine avaient été poolés et les cosmides de ces clones avaient été extraits. Ce pool de cosmides avait été introduit par transformation de protoplastes dans la souche DSM 10191 de *S. griseofuscus* (Cox KL & Baltz RH. 1984), naturellement sensible à la spiramycine (Rao RN *et al.*, 1987, cette souche est disponible notamment auprès de la German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikro-organismen und Zellkulturen GmbH, DSMZ), (Braunschweig, Allemagne), sous le numéro DSM 10191). Les transformants avaient été sélectionnés sur un milieu contenant de l'apramycine. 1300 des clones poussant sur milieu contenant de l'apramycine avaient été transférés sur milieu contenant 5µg/ml de spiramycine. Plusieurs clones résistants à l'apramycine avaient également poussé sur milieu contenant de la spiramycine et les cosmides de ces colonies avaient été extraits et utilisés pour transformer *E. coli* et *S. griseofuscus* (Pernodet *et al.*, 1999). Cinq cosmides capables de conférer la résistance à l'apramycine à *E. coli* et la co-résistance à l'apramycine et la spiramycine chez *S. griseofuscus* avaient ainsi été obtenus. Parmi ces 5 cosmides, il a été déterminé qu'un cosmide nommé pOS44.1 possède dans son insert un gène (SEQ ID N° 15) qui code une protéine (SEQ ID N° 16) présentant une certaine similitude avec une protéine codée par le gène *mdmA* de *Streptomyces mycarofaciens* (SEQ ID N° 88), ce gène a été nommé *srmD* (cf. alignement présenté en figure 26 réalisé grâce au programme FASTA (cf. (Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990), accessibles notamment auprès du centre de ressources INFOBIOGEN, Evry, France).

Pour isoler le déterminant de résistance contenu dans le plasmide pOS44.1, celui-ci a été digéré partiellement par l'enzyme de restriction *Sau3AI* de manière à obtenir des fragments de taille d'environ 1,5 à 3 kb, ces fragments ont été ligués dans le vecteur pIJ486 linéarisé par l'enzyme *BamHI* (Ward *et al.*, 1986). Un plasmide a été sélectionné pour sa capacité à conférer la résistance à la spiramycine chez la souche DSM 10191 de *S. griseofuscus* (Rao RN *et al.*, 1987), naturellement sensible à la spiramycine (cf. ci-dessus). Pour cela, le pool de plasmides correspondant aux fragment *Sau3AI* de pOS44.1 ligués dans le vecteur pIJ486 (cf. ci-dessus), a été introduit par transformation de protoplastes dans la souche DSM 10191 et les transformants ont été sélectionnés pour leur résistance au thiostrepton (due au gène *tsr* porté par pIJ486). Les clones poussant sur milieu contenant du thiostrepton ont été transférés sur milieu contenant de la spiramycine. Plusieurs clones résistants au thiostrepton ont également poussé sur milieu contenant de la spiramycine et les plasmides de ces colonies ont été extraits. Un plasmide conférant la résistance et contenant un insert d'environ 1.8kb a été sélectionné et nommé pOS44.2. Cet insert de 1.8kb peut être sorti facilement grâce à un site *HindIII* et un site *EcoRI* présents dans le vecteur de part et d'autre de l'insert. Cet insert de 1.8kb *HindIII-EcoRI* a été séquencé et le gène de résistance qu'il renferme a été nommé *srmD*. Ce fragment contenant le gène *srmD* a ainsi pu être facilement sous-cloné dans le vecteur pUC19 (numéro d'accès GenBank M77789) ouvert par *EcoRI-HindIII*, le plasmide obtenu a été nommé pOS44.4. L'insert de 1.8kb *HindIII-EcoRI*, contenant le gène *srmD*, de ce plasmide a été utilisé comme sonde pour localiser les gènes de biosynthèse de spiramycine (cf. ci-dessous).

Un échantillon d'une souche *Escherichia Coli* DH5α contenant le plasmide pOS44.4 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2918.

Environ 2000 clones de la banque, obtenus ci-dessus (cf. exemple 1), ont été transférés sur filtre pour hybridation sur colonies selon les techniques classiques (Sambrook *et al.*, 1989).

Les trois sondes décrites ci-dessus ont été marquées au ^{32}P par la technique du « random priming » (Kit commercialisé par Roche) et utilisées pour l'hybridation des 2000 clones de la banque, transférés sur filtre. L'hybridation a été effectuée à 65°C dans le tampon décrit par Church & Gilbert (Church & Gilbert, 1984). Un lavage a été effectué en 2X SSC à 65°C pendant 15 minutes et deux lavages successifs ont ensuite été effectués en 0.5X SSC à 65°C, d'une durée de 15 minutes chacun. Dans ces conditions d'hybridation et de lavage, 16 clones parmi les 2000 hybridés présentaient un fort signal d'hybridation avec au moins une des sondes. Cependant, aucun cosmide n'hybridait avec les trois sondes. Les 16 cosmides ont été extraits et digérés par l'enzyme de restriction *Bam*HI. La comparaison des profils de restriction de ces différents cosmides entre eux, a conduit à choisir deux cosmides susceptibles de contenir les inserts les plus longs de la région et n'ayant pas de bandes communes. Ainsi deux cosmides l'un nommé pSPM5 et l'autre pSPM7 ont été choisis. Le cosmide pSPM5 hybridait avec les sondes *orf1* à *orf4* et la sonde *orf8*, mais n'hybridait pas avec la sonde *srmD*. pSPM7 hybridait seulement avec la sonde *srmD* et pas avec les deux autres sondes.

Ces deux cosmides ont été séquencés en totalité par la technique du séquençage en aveugle (« shotgun sequencing »). Les séquences des inserts de ces deux cosmides : pSPM7 et pSPM5 ont pu être assemblées car bien que ne se chevauchant pas, chacun des inserts comprenait une séquence connue à l'une de ses extrémités. En effet, chacun de ces inserts comportait un fragment de la séquence d'un des gènes codant une enzyme appelée « polyketide synthase » (PKS). Ces 5 gènes ont été clonés par Burgett S. *et al.* en 1996 (Brevet américain US 5,945,320) (cf. figure 2). Ainsi, il a pu être déterminé une séquence unique d'ADN génomique de *S. ambofaciens*. Une séquence de 30943 nucléotides partant en 5' d'un site *Eco*RI situé dans le premier gène PKS et allant jusqu'à un site *Bam*HI en 3' est présentée en SEQ ID N° 1. Cette séquence correspond à la région amont des gènes de la PKS (cf. figure 2 et 3). Une deuxième région de 11171 nucléotides, partant d'un site *Pst*I en 5' et allant jusqu'à un site *Nco*I en 3' situé dans le cinquième gène de la PKS est présentée en SEQ ID N° 2. Cette deuxième région est la

région aval des gènes de la PKS (l'aval et l'amont étant défini par l'orientation des 5 gènes de PKS tous orientés dans le même sens) (cf. figure 2 et 3).

EXEMPLE 4 : Analyse des séquences nucléotidiques, détermination des phases ouvertes de lecture et caractérisation des gènes impliqués dans la biosynthèse des spiramycines.

Les séquences obtenues ont été analysées grâce au programme FramePlot (Ishikawa J & Hotta K. 1999). Ceci a permis d'identifier, parmi les phases ouvertes de lecture, les phases ouvertes de lecture présentant un usage des codons typique de *Streptomyces*. Cette analyse a permis de déterminer que cette région comporte 35 ORFs localisées de part et d'autre de cinq gènes codant l'enzyme « polyketide synthase » (PKS). Respectivement 10 et 25 ORFs ont été identifiées en aval et en amont de ces gènes (l'aval et l'amont étant défini par l'orientation des 5 gènes de PKS tous orientés dans le même sens) (cf. figure 3). Ainsi, les 25 phases ouvertes de lecture de ce type, occupant une région d'environ 31 kb (SEQ ID N° 1 et figure 3) ont été identifiées en amont des 5 gènes codant la PKS et 10 occupant une région d'environ 11,1 kb (SEQ ID N° 2 et figure 3), ont été identifiées en aval des gènes de la PKS. Les gènes de la région amont ont été nommés *orf1*, *orf2*, *orf3*, *orf4*, *orf5*, *orf6*, *orf7*, *orf8*, *orf9c*, *orf10*, *orf11c*, *orf12*, *orf13c*, *orf14*, *orf15c*, *orf16*, *orf17*, *orf18*, *orf19*, *orf20*, *orf21c*, *orf22c*, *orf23c*, *orf24c* et *orf25c* (SEQ ID N° 23, 25, 28, 30, 34, 36, 40, 43, 45, 47, 49, 53, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82 et 84). Les gènes de la région aval ont été nommés *orf1*c*, *orf2*c*, *orf3*c*, *orf4*c*, *orf5**, *orf6**, *orf7*c*, *orf8**, *orf9**, *orf10** (SEQ ID N° 3, 5, 7, 9, 11, 13, 15, 17, 19 et 21). Le « c » ajouté dans le nom du gène signifiant pour l'ORF en question que la séquence codante est dans l'orientation inverse (le brin codant est donc le brin complémentaire de la séquence donnée en SEQ ID N° 1 ou SEQ ID N° 2 pour ces gènes) (cf. figure 3).

Les séquences protéiques déduites de ces phases ouvertes de lecture ont été comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et al.*, 1997), CD-search, COGs (Cluster of Orthologous Groups) (ces trois programmes sont accessibles notamment

auprès du National Center for Biotechnology Information (NCBI) (Bethesda, Maryland, USA)), FASTA ((Pearson W. R. & D. J. Lipman, 1988) et (Pearson W. R., 1990), BEAUTY (Worley K. C. *et al.*, 1995)), (ces deux programmes sont accessibles notamment auprès du centre de ressources INFOBIOGEN, Evry, France). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycine.

EXEMPLE 5 : Inactivation de gène : principe de la construction d'une souche de *Streptomyces ambofaciens* interrompue

Les méthodes utilisées consistent à effectuer un remplacement de gène. Le gène cible à interrompre est remplacé par une copie de ce gène interrompue par une cassette conférant la résistance à un antibiotique (par exemple l'apramycine ou l'hygromycine), comme l'illustre la figure 9. Les cassettes utilisées sont bordées de part et d'autre par des codons de terminaison de la traduction dans toutes les phases de lecture et par des terminateurs de transcription actifs chez *Streptomyces*.

L'insertion de la cassette dans le gène cible peut s'accompagner ou non d'une délétion dans ce gène cible. La taille des régions flanquant la cassette peut aller de quelques centaines à plusieurs milliers de paires de bases.

Les constructions nécessaires à l'inactivation du gène par la cassette ont été obtenues chez *E. coli*, organisme de référence pour l'obtention de constructions d'ADN recombinant. Le gène interrompu a été obtenu dans un plasmide se répliquant chez *E. coli* mais ne pouvant pas se répliquer chez *Streptomyces*.

Les constructions ont ensuite été sous-clonées dans des vecteurs pour permettre la transformation et l'inactivation du gène voulu chez *S. ambofaciens*. Pour cela, deux plasmides ont été utilisés :

- pOJ260 (Bierman M. *et al.*, 1992) (cf. exemple 2) qui confère la résistance à l'apramycine chez *E. coli* et *Streptomyces* et qui a été employé quand le gène cible a été interrompu par une cassette conférant la résistance à l'hygromycine.

- 5 - pOSK1205 (4726pb). Ce plasmide dérive du plasmide pBK-CMV (commercialisé par la société Stratagene (LaJolla, Californie, USA)) dans lequel un fragment *AvrII* contenant la séquence codant la résistance à la Néomycine/Kanamycine a été remplacé par une séquence codant la résistance à l'hygromycine, tout en conservant le promoteur P SV40. Pour cela, le plasmide pHP45-*Ω*hyg (Blondelet-Rouault *et al.*, 1997) a été digéré par les enzymes *NotI* et *PflmI* et le fragment conférant la résistance à l'hygromycine a été sous-cloné au niveau du site *AvrII* du vecteur pBK-CMV après que toutes les extrémités aient été rendues bout franc par traitement à l'enzyme de Klenow. Dans pOSK1205, la cassette qui confère la résistance à l'hygromycine est précédée du promoteur pSV40. Ce plasmide confère la résistance à l'hygromycine chez *E. coli* et *Streptomyces* et a été employé quand le gène cible a été interrompu par une cassette conférant la résistance à l'apramycine.

15 Les cassettes ont été introduites dans le gène cible soit par clonage en utilisant des sites de restriction présents dans le gène cible, soit par recombinaison entre de courtes séquences identiques comme décrit par exemple par Chaverroche *et al* (Chaverroche MK *et al.*, 2000).

20 Le plasmide portant le gène interrompu par la cassette peut ensuite être introduit chez *Streptomyces ambofaciens*, par exemple par conjugaison entre *E. coli* et *Streptomyces* (Mazodier P, *et al.*, 1989). Cette technique a été utilisée dans le cas où le vecteur de base est le vecteur pOJ260. Une deuxième technique peut être utilisée : la technique de la transformation de protoplastes après dénaturation par traitement alcalin de l'ADN (Kieser, T *et al.*, 2000), pour augmenter la fréquence de recombinaison comme décrit par exemple par Oh & Chater (Oh & Chater, 1997). Cette technique a été utilisée dans le cas où le vecteur de base est pOJ260 ou pOSK1205 (cf. ci-dessous). Les transformants sont ensuite sélectionnés avec l'antibiotique correspondant à la cassette présente dans le gène cible (cf. figure 9, antibiotique B). On sélectionne ainsi un mélange de clones parmi lesquels il y a eu intégration par un seul ou par deux événements de recombinaison. Dans un deuxième temps on recherche les clones sensibles à

l'antibiotique pour lequel le gène de résistance est présent dans le vecteur (hors de la cassette de recombinaison) (cf. figure 9, antibiotique A). On peut ainsi sélectionner les clones pour lesquels il y a eu en principe deux événements de recombinaison aboutissant au remplacement du gène sauvage par la copie interrompue par la cassette. Ces étapes
5 sont schématisées figure 9.

Plusieurs cassettes peuvent être utilisées pour l'interruption des gènes ciblés. On peut par exemple utiliser la cassette Ω_{hyg} qui confère la résistance à l'hygromycine (Blondelet-Rouault *et al.*, 1997, numéro d'accès GenBank : X99315).

10 **EXEMPLE 6 : Construction d'une souche de *Streptomyces ambofaciens* interrompue en phase dans le gène *orf3***

Le gène *orf3* avait été interrompu par la cassette Ω_{hyg} (cf. exemple 2.2) et il a pu être démontré qu'une souche *orf3::\Omega_{hyg}* ne produit plus de spiramycine, confirmant l'implication d'un ou plusieurs gènes de la région clonée dans la biosynthèse des
15 spiramycines (cf. exemple 2.2). Au vu de leur orientation, une cotranscription des ORFs 1 à 7 est probable (cf. figure 3) et le phénotype observé (non producteur de spiramycines) peut être dû à l'inactivation d'un ou plusieurs des gènes co-transcrits avec *orf3*. Pour confirmer l'implication de *orf3* dans la biosynthèse de spiramycines, une nouvelle inactivation du gène *orf3* a été effectuée, cette dernière étant réalisée en phase.
20 Pour cela, un fragment *DraIII* interne à *orf3* de 504 paires de bases a été délété. Un fragment d'ADN obtenu à partir de pOS49.1, allant du site *EcoRI* situé en position 1 (SEQ ID N° 1) jusqu'au site *SacI* situé en position 5274 (SEQ ID N° 1) et qui comporte une délétion entre les deux sites *DraIII* aux positions 2563 et 3067 (504 nucléotides retirés) a été cloné dans le plasmide pOJ260 (Bierman M. *et al.*, 1992). Le plasmide
25 ainsi obtenu a été nommé pOS49.67.

L'insert de pOS49.67 est donc constitué par un fragment d'ADN de *S. ambofaciens* contenant les gènes *orf1*, *orf2*, *orf3* avec la délétion en phase, *orf4* et une partie de *orf5*. Le vecteur dans lequel cet insert a été sous-cloné est pOJ260, le plasmide pOS49.67

confère donc la résistance à l'apramycine et a été introduit par transformation de protoplastes dans la souche OS49.16 (cf. exemple 2). La souche OS49.16 étant résistante à l'hygromycine, des transformants hygR et apraR ont été obtenus. Après deux passages de tels clones sur milieu non sélectif, les clones sensibles à l'apramycine et à l'hygromycine (apraS et hygS) ont été recherchés. Dans certains de ces clones, on s'attend en effet à ce qu'un événement de recombinaison entre séquences homologues conduise au remplacement de la copie de *orf3* interrompue par la cassette Ω hyg (contenu dans le génome de la souche OS49.16) par la copie de *orf3* avec la délétion en phase présente sur le vecteur. Les clones résultant de cette recombinaison sont attendus comme apraS et hygS après l'élimination des séquences du vecteur. Le génotype des souches ainsi obtenues peut être vérifié par hybridation ou par PCR et séquençage du produit de PCR (pour vérifier que seule une copie de *orf3* délétée en phase est présente dans le génome des clones obtenus). Des clones n'ayant que la copie de *orf3* avec une délétion en phase ont ainsi été obtenus et leur génotype a été vérifié. Un clone présentant les caractéristiques recherchées a été plus particulièrement sélectionné et a été nommé OS49.67.

Un échantillon de la souche OS49.67 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2916.

EXEMPLE 7 : Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf8*

Pour réaliser l'inactivation du gène *orf8*, une construction dans laquelle la cassette Ω hyg a été introduite dans la séquence codante de *orf8* a été obtenue. Pour cela, le plasmide pOS49.88 a tout d'abord été construit. Le plasmide pOS49.88 est dérivé du plasmide pUC19 (numéro d'accès GenBank M77789) par insertion d'un fragment de 3.7 kb (fragment *Pst*I-*Eco*RI obtenu à partir du cosmide pSPM5), contenant la fin de *orf7*,

orf8 et le début de *orf9*, cloné dans les sites *Pst*I-*Eco*RI de pUC19. La cassette Ω hyg (sous forme d'un fragment *Bam*HI, rendu bout franc par traitement à l'enzyme de Klenow) a été clonée au niveau du site unique *Sal*I de pOS49.88 situé dans *orf8* après que toutes les extrémités aient été rendues bout franc par traitement à l'enzyme de Klenow.

Le clonage étant bout franc, deux types de plasmides ont été obtenus selon le sens d'insertion de la cassette : pOS49.106 dans lequel les gènes *hyg* et *orf8* sont dans la même orientation et pOS49.120 dans lequel les gènes *hyg* et *orf8* sont dans des orientations opposées. L'insert du plasmide pOS49.106 a ensuite été sous-cloné dans le plasmide pOJ260 pour donner pOS49.107. Pour cela, le plasmide pOS49.106 a été digéré par l'enzyme *Asp*718I et les extrémités ont été rendues bout franc par traitement à l'enzyme de Klenow, ce produit de digestion a été redigéré par l'enzyme *Pst*I et le fragment contenant le gène *orf8* dans lequel la cassette Ω hyg a été insérée, a été cloné dans le vecteur pOJ260 (cf. ci dessus). Pour cela, le vecteur pOJ260 a été digéré par les enzymes *Eco*RV et *Pst*I et utilisé pour la ligation. Cette manipulation permet donc d'obtenir une ligation orientée puisque chacun des deux fragments est bout franc d'un côté et *Pst*I de l'autre. Le plasmide obtenu a été nommé pOS49.107.

Le plasmide pOS49.107 a été introduit dans la souche ATCC23877 de *S. ambifaciens* par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation des protoplastes, les clones sont sélectionnés pour leur résistance à l'hygromycine. Les clones résistants à l'hygromycine sont ensuite repiqués respectivement sur milieu avec hygromycine (antibiotique B) et sur milieu avec apramycine (antibiotique A) (cf. figure 9). Les clones résistants à l'hygromycine (HygR) et sensibles à l'apramycine (ApraS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf8* interrompu par la cassette Ω hyg. Le remplacement de la copie sauvage de *orf8* par la copie interrompue par la cassette Ω hyg a été vérifié par Southern Blot. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec

une sonde correspondant à la cassette *Ωhyg* pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde l'insert *PstI-EcoRI* contenant la fin de *orf7*, *orf8* et le début de *orf9* d'une taille d'environ 3,7 kb du plasmide pOS49.88. La vérification du génotype
5 peut être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone *orf8::Ωhyg* a été choisi et nommé OS49.107. Un échantillon de la souche OS49.107 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex
10 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2917.

EXEMPLE 8 : Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf10*

Un fragment d'ADN de 1,5 kb interne au gène *orf10* a été obtenu par PCR en
15 utilisant comme matrice, l'ADN génomique de *S. ambofaciens* et les amorces suivantes :

SRMR1 : 5' CTGCCAGTCCTCTCCCAGCAGTACG 3' (SEQ ID N° 89)

SRMR2 : 5' TGAAGCTGGACGTCTCCTACGTCGG 3' (SEQ ID N° 90)

Ce fragment d'ADN issu de la PCR a été cloné dans le vecteur pCR2.1 commercialisé par la société Invitrogen (Carlsbad, Californie, USA). Le plasmide ainsi obtenu a été
20 nommé pOS49.32. La cassette *Ωhyg* (sous forme d'un fragment *Bam*HI, cf. ci-dessus) a été clonée au niveau du site unique *Bst*EII interne au fragment du gène *orf10*, après que toutes les extrémités aient été rendues bout franc par traitement à l'enzyme de Klenow. Le clonage étant bout franc, deux types de plasmides ont été obtenus selon le sens d'insertion de la cassette : pOS49.43 dans lequel les gènes *hyg* et *orf10* sont dans la
25 même orientation et pOS49.44 dans lequel les gènes *hyg* et *orf10* sont dans des

orientations opposées. L'insert du plasmide pOS49.43 a été transféré (sous forme d'un fragment *Asp718I-XbaI* dont les extrémités ont été rendues bout franc par traitement à l'enzyme de Klenow) dans le site *EcoRV* du plasmide pOJ260 ce qui a permis d'obtenir le plasmide pOS49.50. Le plasmide pOS49.50 contenant le fragment du gène *orf10* interrompu par la cassette *Ωhyg* a été introduit dans la souche ATCC23877 de *Streptomyces ambofaciens*. Après transformation, les clones sont sélectionnés pour leur résistance à l'hygromycine. Les clones résistants à l'hygromycine sont ensuite repiqués respectivement sur milieu avec hygromycine (antibiotique B) et sur milieu avec apramycine (antibiotique A) (cf. figure 9). Les clones résistants à l'hygromycine (HygR) et sensibles à l'apramycine (ApraS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf10* interrompu par la cassette *Ωhyg*. Des clones qui possédaient le marqueur de résistance à l'hygromycine porté par la cassette et qui avaient perdu le marqueur de résistance à l'apramycine porté par le vecteur pOJ260 ont ainsi été obtenus. L'événement de remplacement de la copie sauvage de *orf10* par la copie interrompue *orf10::Ωhyg* a été vérifié par Southern Blot. Ainsi, l'ADN génomique total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *Ωhyg* pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde le produit PCR de 1,5 kb interne au gène *orf10* (cf. ci-dessus).

Un clone présentant les caractéristiques attendues (*orf10::Ωhyg*) a été plus particulièrement sélectionné et nommé OS49.50. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *Ωhyg* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *Ωhyg* dans le génome de ce clone. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

EXEMPLE 9 : Inactivation de gènes : principe de la construction d'une souche de *Streptomyces ambofaciens* interrompue selon la technique des « cassettes excisables » (cf. Figure 9 et 10)

5

Un deuxième type de cassettes peut être utilisé pour l'inactivation de gènes : les cassettes dites « cassettes excisables ». Ces cassettes présentent l'avantage de pouvoir être excisées chez *Streptomyces* par un événement de recombinaison spécifique de site après avoir été introduites dans le génome de *S. ambofaciens*. Le but est d'inactiver certains gènes dans des souches de *Streptomyces* sans laisser dans la souche finale de marqueurs de sélection ou de grandes séquences d'ADN n'appartenant pas à la souche. Après excision il subsiste uniquement une courte séquence d'environ une trentaine de paires de bases (appelé site « cicatriciel ») dans le génome de la souche (cf. figure 10).

15

La mise en oeuvre de ce système consiste, dans un premier temps, au remplacement de la copie sauvage du gène cible (grâce à deux événements de recombinaison homologue, cf. figure 9) par une construction dans laquelle une cassette excisable a été insérée dans ce gène cible. L'insertion de cette cassette est accompagnée d'une délétion dans le gène cible (cf. figure 9). Dans un deuxième temps, l'excision de la cassette excisable du génome de la souche est provoquée. La cassette excisable fonctionne grâce à un système de recombinaison spécifique de site et a pour avantage de permettre l'obtention de mutants de *Streptomyces* ne portant finalement pas de gène de résistance. On s'affranchit également d'éventuels effets polaires sur l'expression des gènes situés en aval du ou des gènes inactivés (cf. figure 10).

20

L'emploi de cassettes excisables a été décrit et utilisé chez de nombreux organismes dont les cellules de mammifère, de levures et chez *E. coli* (Bayley *et al.*, 1992 ; Brunelli et Pall, 1993 ; Camilli *et al.*, 1994 ; Dale et Ow, 1991 ; Russell *et al.*, 1992; Lakso *et al.*, 1992). Ces cassettes excisables utilisent toutes la recombinase spécifique de site Cre agissant sur les sites lox. Ce système de recombinaison provient du bactériophage P1.

25

Pour la construction d'un système de type « cassette excisable » chez *Streptomyces*, il a été tiré partie du système de recombinaison spécifique de site décrit pour l'élément génétique mobile pSAM2 de *Streptomyces ambofaciens* (Boccard *et al.*, 1989a et b). Le système mis en place consiste dans un premier temps à construire un vecteur recombinaison comportant le gène à interrompre dans lequel est insérée une cassette excisable. L'insertion dans le gène cible de la cassette excisable s'accompagne d'une délétion dans le gène cible. Elle peut être réalisée par clonage en utilisant des sites de restriction présents dans le gène cible ou par recombinaison entre de courtes séquences identiques comme décrit par exemple par Chaveruche *et al* (Chaveruche MK *et al.*, 2000). La cassette excisable peut être construite en utilisant par exemple la cassette Ω_{hyg} (Blondelet Rouault *et al.*, 1997). Cette cassette a été bordée par des séquences *attR* et *attL* qui flanquent normalement la copie intégrée de pSAM2 (cf. figure 15). Les séquences *attL* et *attR* contiennent tous les sites nécessaires à la recombinaison site-spécifique permettant l'excision de pSAM2 ou de tout fragment d'ADN situé entre ces deux régions (Sezonov *et al.*, 1997, Raynal *et al.*, 1998). La construction d'une telle cassette n'est évidemment pas limitée à l'utilisation d'une cassette Ω_{hyg} , mais d'autres cassettes de résistances peuvent servir de base à la construction de cette cassette (par exemple la cassette Ω_{aac} ou Ω_{vph} (Blondelet Rouault *et al.*, 1997)).

Après l'obtention de cette construction, la souche de *Streptomyces* est transformée avec le plasmide recombinaison. Les transformants sont ensuite sélectionnés avec l'antibiotique correspondant à la cassette présente dans le gène cible (cf. figure 9, antibiotique B, il s'agit par exemple d'une sélection par l'hygromycine si la cassette excisable dérive de la cassette Ω_{hyg}). On sélectionne ainsi un mélange de clones parmi lesquels il y a eu intégration par un seul ou par deux événements de recombinaison. Dans un deuxième temps on recherche les clones sensibles à l'antibiotique pour lesquels le gène de résistance est présent dans le vecteur (hors de la cassette de recombinaison) (cf. figure 9, antibiotique A). On peut ainsi sélectionner les clones pour lequel il y a eu en principe deux événements de recombinaison aboutissant au remplacement du gène sauvage par la copie interrompue par la cassette. Ces étapes sont schématisées figure 9, le génotype des clones ainsi obtenus est vérifié par Southern blot et une souche

présentant les caractéristiques voulues (le remplacement du gène sauvage par la copie interrompue par la cassette excisable) est sélectionnée.

Dans un deuxième temps, la souche sélectionnée ci-dessus, est transformée par un plasmide permettant l'expression des gènes *xis* et *int* qui sont tous deux nécessaires à la recombinaison spécifique de site entre les sites *attR* et *attL*. Cette recombinaison entraîne le départ de la cassette excisable du génome de la souche, grâce à un événement de recombinaison (cf. figure 10) (Raynal *et al.*, 1998). Il est intéressant de choisir le vecteur portant les gènes *xis* et *int* parmi les vecteurs relativement instables chez *Streptomyces* (par exemple dérivé du vecteur *Streptomyces* pWHM3, (Vara *et al.*, 1989)), ceci permet d'obtenir une souche ayant perdu ce dernier vecteur après quelques cycles de sporulation en absence de pression de sélection.

Pour exciser la cassette, on peut par exemple utiliser le plasmide pOSV508 (cf. figure 14) qui est introduit par transformation de protoplastes dans la souche de *S. ambofaciens* contenant un gène interrompu par la cassette excisable. Le plasmide pOSV508 est dérivé du plasmide pWHM3 (Vara J *et al.*, 1989) (cf. figure 13) dans lequel ont été ajoutés les gènes *xis* et *int* de pSAM2 (Boccard F. *et al.*, 1989b) placés sous le contrôle du promoteur ptrc (Amann, E. *et al.*, 1988). Les gènes *xis* et *int* placés sous le contrôle du promoteur ptrc ont été sous clonés dans le plasmide pWHM3 à partir du plasmide pOSint3 (Raynal *et al.*, 1998) (cf. figure 14). L'introduction dans la souche mutante du plasmide pOSV508 portant les gènes *xis* et *int* de pSAM2 va permettre l'excision efficace par recombinaison spécifique de site de la cassette excisable entre les sites *attL* et *attR* flanquant cette cassette (Raynal A. *et al.*, 1998) (figure 10). Parmi les transformants, sélectionnés pour leur résistance au thiostrepton due au gène *tsr* porté par pOSV508, on choisit ceux devenus sensibles à l'antibiotique auquel la présence de la cassette confère la résistance (cf. figure 10). L'excision est efficace et il a été observé que plus de 90% des transformants sont de ce type. Après un ou plusieurs cycles de croissance et sporulation sur un milieu solide dépourvu de thiostrepton, on obtient des clones ayant perdu le plasmide pOSV508. Ces clones sont repérables par leur sensibilité

au thiostrepton. La séquence du gène cible délété peut être vérifiée par PCR et séquençage du produit PCR.

5 Finalement, la souche résultante porte au niveau du gène inactivé (délétion interne par exemple) un site att « cicatriciel » correspondant au site *attB* minimal (Raynal *et al.*, 1998), issu de la recombinaison entre les sites *attR* et *attL*. Ce site *attB* minimal qui subsiste est semblable à celui présent naturellement dans les souches de *Streptomyces ambofaciens*, *Streptomyces pristinaespiralis* et *Streptomyces lividans* (Sezonov *et al.*, 1997).

10 Le gène que l'on veut inactiver peut être cotranscrit avec d'autres gènes situés en aval. Pour éviter que l'inactivation d'un des gènes ait un effet polaire sur l'expression des gènes en aval dans l'opéron, il est important d'obtenir une délétion en phase après excision de la cassette. Le système des cassettes excisables tel que décrit ci-dessus permet de répondre à une telle exigence. L'homme du métier pourra, en effet, aisément
15 construire trois cassettes excisables distinctes, celles-ci laissant après excision une séquence de 33, 34 ou 35 nucléotides respectivement, sans codon stop quelle que soit la phase de lecture. Connaissant la séquence du gène cible et la taille de la délétion associée à l'insertion de la cassette, il est possible de choisir entre ces trois cassettes excisables de façon à ce que l'excision conduise à une délétion en phase. Sur les 33, 34 ou 35 nucléotides rajoutés, 26 correspondent à la séquence *attB* minimale (cf. figure 27).

20 Dans le cas de la présente demande, deux cassettes excisables ont été utilisées. Ces deux cassettes sont les suivantes : *att1Ωhyg+* (SEQ ID N° 91) et *att3Ωaac-* (SEQ ID N° 92), ces cassettes laissent respectivement 33 et 35 nucléotides après excision. Elles comportent respectivement la cassette *Ωhyg* ou la cassette *Ωaac*, les signes + et - correspondant à l'orientation de la cassette de résistance. Ces deux cassettes ont été
25 construites et clonées au niveau du site *EcoRV* du vecteur pBC SK+ dont le site *HindIII* a été supprimé au préalable. Les plasmides obtenus ont été nommés *patt1Ωhyg+* et *patt3Ωaac-* respectivement. Les cassettes excisables peuvent être facilement sorties par une digestion du plasmide par *EcoRV*.

EXEMPLE 10 : Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf2*

L'inactivation du gène *orf2* a été réalisée grâce à la technique des cassettes excisables (cf. ci-dessus). La souche de départ utilisée est la souche *Streptomyces ambofaciens* OSC2 qui dérive de la souche ATCC23877. Cependant, la souche OSC2 diffère de la souche ATCC23877 en ce qu'elle a perdu l'élément génétique mobile pSAM2 (Boccard *et al.*, 1989a et b). Cet élément mobile peut être perdu de manière spontanée au cours de la protoplastisation (action du lysozyme pour digérer la paroi bactérienne et fragmenter le mycélium (Kieser *et al.*, 2000)) et de la régénération des protoplastes de la souche ATCC23877. Pour sélectionner les clones ayant perdu l'élément pSAM2, il a été mis en place un crible, basé sur la répression du gène *pra* par le répresseur de transcription KorSA (Sezonov *et al.*, 1995) (Sezonov G. *et al.*, 2000). Pour cela, un fragment d'ADN contenant le promoteur du gène *pra* placé en amont du gène *aph* (conférant la résistance à la kanamycine et dépourvu de son propre promoteur) a été cloné dans le vecteur instable pWHM3Hyg, ce dernier dérivant du plasmide pWHM3 (Vara *et al.*, 1989) dans lequel le gène *tsr* a été remplacé par le gène *hyg* (conférant la résistance à l'hygromycine). Le plasmide ainsi obtenu a été nommé pOSV510. La souche ATCC23877 a été transformée après protoplastisation par le plasmide pOSV510. Le promoteur Pra est un promoteur réprimé par le répresseur KorSA, le gène codant ce dernier étant situé au sein de l'élément mobile pSAM2 (Sezonov G. *et al.*, 2000). Après transformation par le plasmide pOSV510, les bactéries transformées sont sélectionnées pour leur résistance à la kanamycine (due au gène *aph* porté par pOSV510). Les clones ayant perdu l'élément intégratif pSAM2 ont perdu le répresseur KorSA et le promoteur Pra n'est donc plus réprimé et permet l'expression du gène de résistance à la kanamycine *aph*. Une sélection par la kanamycine après transformation par le plasmide pOSV510 permet donc de sélectionner les clones ayant perdu l'élément intégratif pSAM2 (et donc KorSA) et possédant le plasmide pOSV510. Le plasmide pOSV510 étant instable, après quelques cycles de sporulation sans antibiotique, des clones isolés sont repiqués sur milieu avec kanamycine, sur milieu avec hygromycine et sur milieu sans antibiotique. Les clones sensibles à la kanamycine et à

l'hygromycine ont perdu pOSV510. La perte de l'élément pSAM2 a été vérifiée par hybridation et PCR. Un clone présentant les caractéristiques souhaitées a été sélectionné et a été nommé OSC2.

Un échantillon de la souche OSC2 a été déposé auprès de la Collection Nationale
 5 de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux
 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-
 2908.

L'inactivation du gène *orf2* a été réalisée grâce à la technique des cassettes
 excisables (cf. ci-dessus et figure 10). Pour cela, un insert de 4,5 kb dont la séquence
 10 part du site *EcoRI* situé en position 1 jusqu'au site *BamHI* situé en position 4521 (SEQ
 ID N° 1) a été sous-cloné au niveau des sites *EcoRI* et *BamHI* du plasmide pUC19
 (numéro d'accès GenBank M77789) à partir du cosmide pSPM5. Le plasmide ainsi
 obtenu a été nommé pOS49.99.

Ce plasmide a été introduit dans la souche *E. coli* KS272 qui contenait déjà le
 15 plasmide pKOBEG (Chaveroche *et al.*, 2000) (cf. figure 12).

En parallèle, la cassette excisable *att3 Ω aac-* (SEQ ID N° 92, cf. ci dessus) a été
 amplifiée par PCR en utilisant comme matrice le plasmide pOSK1102 (Le plasmide
 pOSK1102 est un plasmide dérivé du vecteur pGP704Not (Chaveroche *et al.*, 2000)
 (Miller VL & Mekalanos JJ, 1988) dans lequel la cassette *att3 Ω aac-* a été clonée comme
 20 un fragment *EcoRV* dans le site unique *EcoRV* de pGP704Not) et à l'aide des amorces
 suivantes :

ORF2A

5' CCCGCGCGGCAGCCTCTCCGTGATCGAGTCCGGCGTGACCATCGCGCGCGCTTCGTTTCGG-3' (SEQ
 25 ID N° 93)

ORF2B

5' GCTCCGTGCGTCATGCAGGAAGGTGTCGTAGTCGCGGTAGATCTGCCTCTTCGTCCCGAA-3' (SEQ
 ID N° 94)

Les 40 déoxy-nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène cible (*orf2* dans le cas présent) et les 20 déoxy-nucléotides situés le plus en 3' (figurés en gras ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3 Ω aac-* (cf. figure 11).

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli* contenant les plasmides pKOBEG et pOS49.99 comme décrit (Chaveroche *et al.*, 2000) (cf. figure 12). Ainsi, les bactéries ont été transformées par électroporation et sélectionnées pour leur résistance à l'apramycine. Les plasmides des clones obtenus ont été extraits et digérés par plusieurs enzymes de restriction, dans le but de vérifier que le profil de digestion obtenu correspond au profil attendu s'il y a eu insertion de la cassette (*att3 Ω aac-*) dans le gène cible (*orf2*), c'est à dire si il y a bien eu recombinaison homologue entre les extrémités du produit PCR et le gène cible (Chaveroche *et al.*, 2000). La vérification de la construction peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un clone dont le plasmide possède le profil attendu a été sélectionné et le plasmide correspondant a été nommé pSPM17. Ce plasmide dérive de pOS49.99 dans lequel *orf2* est interrompue par la cassette apramycine (cf. figure 12).

L'insertion de la cassette s'accompagne d'une délétion dans *orf2*, entre les nucléotides 211 et 492 de la partie codante de *orf2*.

Le plasmide pSPM17 a été digéré par l'enzyme *EcoRI* puis les extrémités ont été rendues bout franc par traitement à l'enzyme de Klenow, ce produit de digestion a été ensuite digéré par l'enzyme *XbaI* et l'insert contenant le gène *orf2* délété a été cloné dans le vecteur pOSK1205 (cf. ci dessus). Pour cela, le vecteur pOSK1205 a été digéré par l'enzyme *BamHI* puis les extrémités ont été rendues bout franc par traitement à l'enzyme de Klenow, ce produit a été ensuite digéré par l'enzyme *XbaI*, et utilisé pour la ligation avec l'insert obtenu à partir de pSPM17 comme ci-dessus. Cette manipulation permet donc d'obtenir une ligation orientée puisque chacun des deux fragments est bout franc d'un côté et *XbaI* de l'autre. Le plasmide ainsi obtenu a été nommé pSPM21, il

porte un gène de résistance à l'hygromycine (partie vecteur) et un insert dans lequel le gène *orf2* délété est remplacé par la cassette *att3Ωaac-*.

Le vecteur pSPM21 a été introduit dans la souche *Streptomyces ambofaciens* OSC2 (cf. ci dessus) par transformation de protoplastes (Kieser, T *et al.*, 2000). Après
5 transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec apramycine (antibiotique B) et sur milieu avec hygromycine (antibiotique A) (cf. figure 9). Les clones résistants à l'apramycine (ApraR) et sensibles à l'hygromycine (HygS) sont en principe ceux où un double événement de crossing over s'est produit et qui
10 possèdent le gène *orf2* interrompu par la cassette *att3Ωaac-*. Ces clones ont été sélectionnés et le remplacement de la copie sauvage de *orf2* par la copie interrompue par la cassette a été vérifié. La présence de la cassette *att3Ωaac-* a été vérifiée par PCR sur colonie. Une hybridation a également été effectuée. Pour cela, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur
15 membrane et hybridé avec une sonde de 3 kb correspondant à un fragment *EcoRI-BamHI* de l'insert du plasmide pOS49.99 (cf. ci-dessus). La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

20 Un clone présentant les caractéristiques attendues a été sélectionné et nommé SPM21. Il a pu être vérifié grâce à la PCR et à l'hybridation que la cassette *att3Ωaac-* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att3Ωaac-* dans
25 le génome de ce clone. Ce clone possède donc le génotype : *orf2::att3Ωaac-*.

Un échantillon de la souche SPM21 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2914.

- La souche SPM21 a été transformée par le vecteur pOSV508 par transformation de protoplastes pour provoquer l'excision de la cassette (cf. figure 14). Le plasmide pOSV508 est dérivé du plasmide pWHM3 (Vara J *et al.*, 1989) (cf. figure 13) dans lequel il a été ajouté les gènes *xis* et *int* de pSAM2 (Boccard F. *et al.*, 1989b) placé sous le contrôle du promoteur *ptrc* (Amann, E. *et al.*, 1988) (cf. figure 14). L'introduction dans la souche SPM21 du plasmide pOSV508 portant les gènes *xis* et *int* de pSAM2 permet l'excision efficace par recombinaison spécifique de site de la cassette excisable entre les sites *attL* et *attR* flanquant cette cassette (Raynal A. *et al.*, 1998) (figure 10). Parmi les transformants sélectionnés pour leur résistance au thiostrepton due au gène *tsr* porté par pOSV508, on choisit ceux devenus sensibles à l'apramycine dont le gène de résistance est porté par la cassette *att3Ωaac-*, l'excision entraîne en effet la perte de ce gène de résistance (cf. figure 10). Le plasmide pOSV508 est instable et après deux passages successifs sur milieu sans antibiotique des clones isolés sont repiqués sur milieu avec thiostrepton et sur milieu sans thiostrepton. Les clones sensibles au thiostrepton ont perdu pOSV508. Il a été vérifié que l'excision de la cassette aboutit bien à une délétion en phase dans le gène *orf2* par PCR et séquençage du produit PCR, l'excision de la cassette laisse en effet une séquence « cicatricielle » *att3* caractéristique (qui est similaire au site *attB* d'origine après recombinaison entre les sites *attL* et *attR*) :
- 5' ATCGCGCGCGCTTCGTTCGGGACGAAGAGGTAGAT 3' (SEQ ID N° 95).
- La souche ainsi obtenue et possédant le génotype voulu (*orf2::att3*) a été nommée SPM22.

Un échantillon de la souche SPM22 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2915.

EXEMPLE 11 : Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf12*

Pour l'inactivation de *orf12*, *orf13c* et *orf14* un même plasmide de départ (pSPM504) a été employé pour introduire à différentes positions une cassette de type
 5 « cassette excisable ». Ce plasmide possède un insert de 15,1 kb qui correspond à la région de *orf7* à *orf17*. Pour construire ce plasmide un fragment *Bgl*III de 15,1 kb provenant de la digestion du cosmide pSPM7 (cf. ci-dessus) a été cloné dans le plasmide pMBL18 (Nakano *et al.*, 1995) digéré par *Bam*HI. Les extrémités *Bam*HI et *Bgl*III étant compatibles, on obtient après ligation, le plasmide pSPM502. La totalité de l'insert de
 10 pSPM502 a ensuite été sous-cloné (sous forme d'un fragment *Hind*III/*Nhe*I) dans le plasmide pOSK1205 (digéré par *Hind*III/*Nhe*I) ce qui a permis d'obtenir le plasmide pSPM504.

Ce plasmide a été introduit dans la souche *E. coli* KS272 qui contenait déjà le plasmide pKOBEG (Chaveroche *et al.*, 2000) (cf. figure 12).

15 En parallèle, la cassette excisable *att3Ωaac-* a été amplifiée par PCR en utilisant comme matrice le plasmide pOSK1102 (le plasmide pOSK1102 est un plasmide dérivé du vecteur pGP704Not (Chaveroche *et al.*, 2000) (Miller VL & Mekalanos JJ, 1988) dans lequel la cassette *att3Ωaac-* a été clonée comme un fragment *Eco*RV dans le site unique *Eco*RV de pGP704Not), les amorces utilisées sont les suivantes :

20 EDR8 : 5' CGGGATGATCGCTTGTCCGGCGGCCGGATGCCTAGCCTCATCGCGCGCGCTTCGTTTCGG 3'
 (SEQ ID N° 96)

EDR9 : 5' CCCGATCCAGAACGTCTGGTCGGTGATCAGGTCGCTGTTTCATCTGCCTCTTCGTCCCGAA 3'
 (SEQ ID N° 97)

25 Les 40 (seulement 39 pour EDR8) déoxy-nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène

cible (*orf12* dans le cas présent) et les 20 déoxy-nucléotides situés le plus en 3' (figuré en gras ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3 Ω aac-* (cf. figure 11).

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli*
5 *KS272* contenant les plasmides pKOBEG et pSPM504 (cf. ci-dessus), comme décrit par
Chaverroche *et al.* (Chaverroche *et al.*, 2000) (cf. figure 12 pour le principe, le plasmide
pOS49.99 doit être remplacé par le plasmide pSPM504 et le plasmide obtenu n'est plus
pSPM17 mais pSPM507). Ainsi, les bactéries ont été transformées par électroporation
par ce produit de PCR et les clones ont été sélectionnés pour leur résistance à
10 l'apramycine. Les plasmides des clones obtenus ont été extraits et digérés par plusieurs
enzymes de restriction, dans le but de vérifier que le profil de digestion obtenu
correspond au profil attendu s'il y a eu insertion de la cassette (*att3 Ω aac-*) dans le gène
cible (*orf12*), c'est à dire si il y a bien eu recombinaison homologue entre les extrémités
du produit PCR et le gène cible (Chaverroche *et al.*, 2000). La vérification de la
15 construction peut également être réalisée par toute méthode connue de l'homme du
métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage
du produit de PCR. Un clone dont le plasmide possède le profil attendu a été sélectionné
et le plasmide correspondant a été nommé pSPM507. Ce plasmide dérive de pSPM504
dans lequel *orf12* est interrompue par la cassette *att3 Ω aac-* (cf. figure 12). L'insertion de
20 la cassette s'accompagne d'une délétion dans le gène *orf12*, l'interruption commence au
niveau du trentième codon de *orf12*. Il reste après la cassette les 46 derniers codons de
orf12.

Le vecteur pSPM507 a été introduit dans la souche *Streptomyces ambofaciens* OSC2
(cf. ci-dessus) par transformation de protoplastes (Kieser, T *et al.*, 2000). Après
25 transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les
clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec
apramycine (antibiotique B) et sur milieu avec hygromycine (antibiotique A) (cf. figure
9). Les clones résistants à l'apramycine (ApraR) et sensibles à l'hygromycine (HygS)
sont en principe ceux où un double événement de crossing over s'est produit et qui
30 possèdent le gène *orf12* interrompu par la cassette *att3 Ω aac-*. Ces clones ont été plus

particulièrement sélectionnés et le remplacement de la copie sauvage de *orf12* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3Ωaac-* pour
5 vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde un fragment d'ADN obtenu par PCR et correspondant à une très large partie de la séquence codante du gène *orf12*.

La vérification du génotype peut également être réalisée par toute méthode connue de
10 l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf12::att3Ωaac-*) a été plus particulièrement sélectionné et nommé SPM507. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *att3Ωaac-* était bien présente dans le génome de ce
15 clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att3Ωaac-* dans le génome de ce clone. Ce clone possède donc le génotype : *orf12::att3Ωaac-* et a été nommé SPM507. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf12*, au
20 vu de l'orientation des gènes (cf. figure 3). En effet, le fait que *orf13c* soit orienté en sens opposé à *orf12* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment, notamment par transformation par le plasmide pOSV508. Un échantillon de la souche SPM507 a été déposé auprès de la Collection
25 Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2911.

EXEMPLE 12 : Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf13c*

La cassette excisable *att3Ωaac-* a été amplifiée par PCR en utilisant comme matrice le plasmide pOSK1102 (cf. ci-dessus) à l'aide des amorces suivantes :

5

EDR3 : 5' ACCGGGGCGGTCCTCCCCTCCGGGGCGTCACGGCCGCGGAATCTGCCTCTTCGTCCCGAA 3' (SEQ ID N° 98)

EDR4 : 5' CACGCAGCGAGCCGACGCACTGATGGACGACACGATGGCCATCGCGCGCGCTTCGTTCGG 3' (SEQ ID N° 99)

10

Les 40 déoxy-nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène cible (*orf13c* dans le cas présent) et les 20 déoxy-nucléotides situés le plus en 3' (figurés en gras ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3Ωaac-* (cf. figure 11).

15

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli* KS272 contenant les plasmides pKOBEG et pSPM504 (cf. ci-dessus), comme décrit par Chaverroche *et al.* (Chaverroche *et al.*, 2000) (cf. figure 12 pour le principe, le plasmide pOS49.99 doit être remplacé par le plasmide pSPM504 et le plasmide obtenu n'est plus pSPM17 mais pSPM508). Ainsi, les bactéries ont été transformées par électroporation par le produit de PCR et les clones ont été sélectionnés pour leur résistance à l'apramycine. Les plasmides des clones obtenus ont été extraits et digérés par plusieurs enzymes de restriction, dans le but de vérifier que le profil de digestion obtenu correspond au profil attendu s'il y a eu insertion de la cassette (*att3Ωaac-*) dans le gène cible (*orf13c*), c'est à dire si il y a bien eu recombinaison homologue entre les extrémités du produit PCR et le gène cible (Chaverroche *et al.*, 2000). La vérification de la construction peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un clone dont le plasmide possède le profil attendu a été sélectionné et le plasmide correspondant a été nommé pSPM508. Ce plasmide dérive de pSPM504

30

dans lequel *orf13c* est interrompue par la cassette apramycine (cf. figure 12). L'insertion de la cassette s'accompagne d'une délétion dans le gène *orf13c*, l'interruption commence au niveau du sixième codon de *orf13c*. Il reste après la cassette les 3 derniers codons de *orf13c*.

5 Le vecteur pSPM508 a été introduit dans la souche *Streptomyces ambofaciens* OSC2 (cf. ci-dessus) par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec apramycine (antibiotique B) et sur milieu avec hygromycine (antibiotique A) (cf. figure
10 9). Les clones résistants à l'apramycine (ApraR) et sensibles à l'hygromycine (HygS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf13c* interrompu par la cassette *att3Ωaac-*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf13c* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des
15 clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3Ωaac-* pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde un produit de PCR correspondant à une séquence débordant d'une centaine de paires de base en amont et en
20 aval de la séquence codante de *orf13c*. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf13c::att3Ωaac-*) a été plus particulièrement sélectionné et nommé SPM508. Il a pu en effet être vérifié grâce aux
25 deux hybridations que la cassette *att3Ωaac-* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att3Ωaac-* dans le génome de ce clone. Ce clone possède donc le génotype : *orf13c::att3Ωaac-* et a été nommé SPM508. Il est inutile de

procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf13c*, au vu de l'orientation des gènes (cf. figure 3), le fait que *orf14* soit orienté en sens opposé à *orf13c* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment. Un échantillon de la souche SPM508 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2912.

10 **EXEMPLE 13 Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf14***

La cassette excisable *att3Ωaac-* a été amplifiée par PCR en utilisant comme matrice le plasmide pOSK1102 (cf. ci-dessus) à l'aide des amorces suivantes :

EDR5: 5' GGGCGTGAAGCGGGCGAGTGTGGATGTCATGCGAGTACTCATCGCGCGCGCTTCGTTCGG 3'
15 (SEQ ID N° 100)

EDR6: 5' CGGGAAACGGCGTCGCACTCCTCGGGGGCCGCGTCAGCCCATCTGCCTCTTCGTCCCGAA 3'
(SEQ ID N° 101)

Les 40 déoxy-nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène cible (*orf14* dans le cas présent) et les 20 déoxy-nucléotides situés le plus en 3' (figuré en gras ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3Ωaac-* (cf. figure 11).

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli* KS272 contenant les plasmides pKOBEG et pSPM504 (cf. ci-dessus), comme décrit par

Chaveroche *et al.* (Chaveroche *et al.*, 2000) (cf. figure 12 pour le principe, le plasmide pOS49.99 doit être remplacé par le plasmide pSPM504 et le plasmide obtenu n'est plus pSPM17 mais pSPM509). Ainsi, les bactéries ont été transformées par électroporation par le produit de PCR et les clones ont été sélectionnés pour leur résistance à l'apramycine. Les plasmides des clones obtenus ont été extraits et digérés par plusieurs enzymes de restriction, dans le but de vérifier que le profil de digestion obtenu correspond au profil attendu s'il y a eu insertion de la cassette (*att3Ωaac-*) dans le gène cible (*orf14*), c'est à dire si il y a bien eu recombinaison homologue entre les extrémités du produit PCR et le gène cible (Chaveroche *et al.*, 2000). La vérification de la construction peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un clone dont le plasmide possède le profil attendu a été sélectionné et le plasmide correspondant a été nommé pSPM509. Ce plasmide dérive de pSPM504 dans lequel *orf14* est interrompu par la cassette apramycine (cf. figure 12). L'insertion de la cassette s'accompagne d'une délétion dans le gène *orf14*, l'interruption commence au niveau du quatrième codon de *orf14*. Il reste après la cassette le dernier codon de *orf14*.

Le vecteur pSPM509 a été introduit dans la souche *Streptomyces ambifaciens* OSC2 (cf. ci-dessus) par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec apramycine (antibiotique B) et sur milieu avec hygromycine (antibiotique A) (cf. figure 9). Les clones résistants à l'apramycine (ApraR) et sensibles à l'hygromycine (HygS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf14* interrompu par la cassette *att3Ωaac-*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf14* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3Ωaac-* pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde un produit de PCR

correspondant à une séquence débordant d'une centaine de paires de base en amont et en aval de la séquence codante de *orf14*. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

5 Un clone présentant les caractéristiques attendues (*orf14::att3 Ω aac-*) a été plus particulièrement sélectionné et nommé SPM509. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *att3 Ω aac-* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un
10 remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att3 Ω aac-* dans le génome de ce clone. Ce clone possède donc le génotype : *orf14::att3 Ω aac-* et a été nommé SPM509. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf14*, au vu de l'orientation des gènes (cf. figure 3), le fait que *orf15c* soit orienté en sens opposé à *orf14* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette
15 excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment. Un échantillon de la souche SPM509 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2913.

20

EXEMPLE 14: Construction d'une souche de *Streptomyces ambofaciens* interrompue dans le gène *orf6**

L'inactivation du gène *orf6** a été réalisée grâce à la technique des cassettes excisables (cf. ci-dessus et figure 10). Pour cela, le cosmide pSPM7 a été utilisé comme
25 matrice pour amplifier un fragment du gène *orf6** grâce aux oligonucléotides suivants :

C9583 : 5' CTGCAGGTGCTCCAGCGCGTCGATCT 3' (oligo sens) (SEQ ID N° 102)

C9584 : 5' CTGCAGACGGAGGCGGACCTGCGGCT 3' (oligo antisens) (SEQ ID N° 103)

Les 20 déoxy-nucléotides situés à l'extrémité 3' de ces oligonucléotides correspondent à une séquence située dans la partie codante du gène *orf6** (SEQ ID N° 13) et les 6 déoxy-nucléotides situés les plus en 5' correspondent à la séquence d'un site *PstI* permettant de faciliter le clonage par la suite. Le fragment d'ADN amplifié fait une taille d'environ 1,11 kb. Ce produit de PCR est cloné dans le vecteur pGEM-T Easy (commercialisé par la société Promega (Madison, Wisconsin, USA)) ce qui a permis d'obtenir le plasmide nommé pBXL1111 (cf. figure 16).

La cassette excisable *att1Ωhyg+* a ensuite été introduite dans la séquence codante du gène *orf6**. Pour cela, le plasmide pBXL1111 a été digéré par les enzymes de restriction *SmaI* et *Asp718I* et le produit de digestion a été traité par l'enzyme de Klenow. Cette manipulation permet de réaliser une délétion interne de 120 pb dans la séquence codante du gène *orf6** (cf. figure 15). De plus, de part et d'autre des sites de restrictions, il reste respectivement 511 pb et 485 pb de la séquence de *orf6** qui permettront la recombinaison homologue pour l'inactivation du gène *orf6**. La cassette excisable *att1Ωhyg+* a été préparée à partir du plasmide *patt1Ωhyg+* (cf. ci-dessus) par digestion de ce plasmide par *EcoRV*. Cette dernière a ensuite été sous-clonée dans le vecteur pBXL1111 préalablement préparé comme décrit ci-dessus (digestion *SmaI* et *Asp718I* puis traitement à l'enzyme de Klenow). Le plasmide obtenu a été nommé pBXL1112 (cf. figure 17). Dans cette construction, le gène *orf6** comporte une délétion de 120pb et est interrompu par la cassette *att1Ωhyg+*.

Le plasmide pBXL1112 a été ensuite digéré par l'enzyme *PstI* (site bordant la cassette puisque présent dans les oligonucléotides PCR) et l'insert *PstI* de 3.7 kb comprenant une partie de *orf6** interrompu par la cassette *att1Ωhyg+* a ensuite été cloné au niveau du site *PstI* du plasmide pOJ260 (cf. ci-dessus). Le plasmide ainsi obtenu a été nommé pBXL1113.

Le vecteur pBXL1113 a été introduit dans la souche *Streptomyces ambofaciens* OSC2 (cf. ci-dessus) par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation, les clones sont sélectionnés pour leur résistance à l'hygromycine. Les clones résistants à l'hygromycine sont ensuite repiqués respectivement sur milieu avec
5 hygromycine (antibiotique B) et sur milieu avec apramycine (antibiotique A) (cf. figure 9). Les clones résistants à l'hygromycine (HygR) et sensibles à l'apramycine (ApraS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf6** interrompu par la cassette *att1Ωhyg+*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf6** par la
10 copie interrompue par la cassette a été vérifié par la technique du Southern blot. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant au gène *hyg* (obtenu par PCR) pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde
15 l'insert *PstI-PstI* contenant le gène *orf6**, d'une taille d'environ 1,1 kb et obtenu à partir du plasmide pBXL1111 (cf. ci-dessus et figure 16). La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf6*::att1Ωhyg+*) a été plus
20 particulièrement sélectionné et nommé SPM501. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *att1Ωhyg+* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att1Ωhyg+* dans le génome de ce clone. Ce clone
25 possède donc le génotype : *orf6*::att1Ωhyg+* et a été nommé SPM501. Un échantillon de la souche SPM501 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2909.

- La souche SPM501 a été transformée par le vecteur pOSV508 par transformation de protoplastes pour provoquer l'excision de la cassette (cf. figure 14). Le plasmide pOSV508 est dérivé du plasmide pWHM3 (Vara J *et al.*, 1989) (cf. figure 13) dans lequel les gènes *xis* et *int* de pSAM2 (Boccard F. *et al.*, 1989b) placés sous le contrôle du promoteur ptrc (Amann, E. *et al.*, 1988) ont été ajoutés (cf. figure 14). L'introduction dans la souche SPM501 du plasmide pOSV508 portant les gènes *xis* et *int* de pSAM2 permet l'excision efficace par recombinaison site spécifique de la cassette excisable entre les sites *attL* et *attR* flanquant cette cassette (Raynal A. *et al.*, 1998) (figure 10). Parmi les transformants, sélectionnés pour leur résistance au thiostrepton due au gène *tsr* porté par pOSV508, on choisit ceux devenus sensibles à l'hygromycine dont le gène de résistance est porté par la cassette *att1Ωhyg+*, l'excision entraîne en effet la perte de ce gène de résistance (cf. figure 10). Le plasmide pOSV508 est instable et après deux passages successifs sur milieu sans antibiotique des clones isolés sont repiqués sur milieu avec thiostrepton et sur milieu sans thiostrepton. Les clones sensibles au thiostrepton ont perdu pOSV508. Il a été vérifié que l'excision de la cassette a bien eu lieu en phase dans le gène *orf6** par PCR et séquençage du produit PCR. L'interruption commence au 158^{ième} codon, 40 codons sont délétés (120pb) et l'excision de la cassette laisse une séquence « cicatricielle » *att1* caractéristique de 33pb : 5' ATCGCGCGCTTCGTTTCGGGACGAAGAGGTAGAT 3' (SEQ ID N° 104).
- 20 La souche ainsi obtenue et possédant le génotype voulu (*orf6*::att1*) a été nommée SPM502.

Un échantillon de la souche SPM502 a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 10 juillet 2002 sous le numéro d'enregistrement I-2910.

EXEMPLE 15 : Analyse des souches de *Streptomyces ambofaciens* interrompue dans les gènes *orf2*, *orf3*, *orf8*, *orf10*, *orf12*, *orf13c*, *orf14*, *orf6**

Pour tester la production en spiramycine des diverses souches obtenues, un test microbiologique basé sur la sensibilité d'une souche de *Micrococcus luteus* a été mis au point (cf. (Gourmelen A. *et al.*, 1998)). La souche de *Micrococcus luteus* utilisée est une souche dérivée de la souche DSM1790 naturellement sensible à la spiramycine (cette souche est disponible notamment auprès de la German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikro-organismen und Zellkulturen GmbH, DSMZ), (Braunschweig, Allemagne), sous le numéro DSM1790), la souche utilisée dans le présent test diffère de la souche DSM1790 en ce qu'elle est résistante à la congocidine. Cette souche est un mutant spontané obtenu par sélection sur milieu contenant des doses croissantes de congocidine. Une telle souche a été sélectionnée du fait que *Streptomyces ambofaciens* produit à la fois de la spiramycine et de la congocidine. Le but étant de doser la production en spiramycine des diverses souches obtenues grâce à un test microbiologique basé sur la sensibilité d'une souche de *Micrococcus luteus*, il est nécessaire de disposer d'une souche résistante à la congocidine.

Les différentes souches de *Streptomyces* à tester ont été cultivées dans des erlenmeyers à baffles (erlenmeyers chicanés) de 500 ml contenant 70 ml de milieu MP5 (Pernodet *et al.*, 1993). Les erlenmeyers chicanés ont été inoculés à une concentration initiale de $2.5 \cdot 10^6$ spores/ml des différentes souches de *S. ambofaciens* et mises à pousser à 27°C sous agitation orbitale de 250 rpm. Des prélèvements de 2 ml de suspensions ont été réalisés après 48, 72 et 96 heures de culture et centrifugés. Les différents surnageants ont ensuite été congelés à -20°C. Une dilution au dixième de ces surnageants dans du milieu de culture stérile est utilisée pour le test (cf. figure 18).

La souche indicatrice de *Micrococcus luteus* résistante à la congocidine mais sensible à la spiramycine a été cultivée dans du milieu 2TY (Sambrook *et al.*, 1989) contenant de la congocidine à 5 µg/ml pendant 48h à 37°C. La densité optique (DO) de

la culture est mesurée et cette culture est diluée de façon à ajuster la densité optique à 4. 0,4 ml de cette pré-culture est dilué dans 40 ml de milieu DAM5 (Difco Antibiotic Medium 5, commercialisé par la société Difco), préalablement porté à une température d'environ 45°C. Ce milieu est ensuite coulé dans une boîte carrée de 12x12 cm et laissé
5 à reposer à température ambiante.

Une fois le milieu refroidi et solidifié, des disques en papier Whatman AA (cf. Gourmelen A. *et al.*, 1998) de 12 mm de diamètre ont été imbibés de 70 µl de la dilution au dixième de chaque surnageant et déposés sur la surface de la boîte. Des disques imbibés d'une solution de spiramycine de différentes concentrations (2-4-8 µg/ml dans
10 du milieu de culture MP5) sont utilisés comme gamme étalon. Les boîtes sont laissées à 4°C pendant 2 h de façon à permettre la diffusion des antibiotiques dans l'agar puis sont incubées à 37°C pendant 24 à 48h.

Si le disque contient de la spiramycine, celle-ci diffuse dans l'agar et inhibe la croissance de la souche indicatrice de *Micrococcus luteus*. Cette inhibition crée un
15 « halo » autour du disque, ce halo reflétant la zone où la souche de *Micrococcus luteus* n'a pas poussé. La présence de ce halo est donc une indication de la présence de spiramycine et permet de déterminer si la souche de *S. ambofaciens* correspondant au disque en question est productrice ou non productrice de spiramycine. Une comparaison avec les diamètres d'inhibition obtenus pour la gamme étalon permet d'obtenir une
20 indication de la quantité de spiramycine produite par cette souche.

Les différentes souches décrites dans les exemples précédents ont été utilisées dans ce test pour détecter leur production en spiramycine. Les résultats obtenus ont été regroupés dans le Tableau 38.

Tableau 38

Souche	Gène inactivé	Exemple dans lequel la souche est décrite	Phénotype : Producteur (+) ou non-producteur (-) de spiramycine
ATCC23877	Aucun	1	(+)
OS49.16	<i>orf3::Ωhyg</i>	2	(-)
OS49.67	<i>orf3</i> délétion en phase	6	(-)
OS49.107	<i>orf8::Ωhyg</i>	7	(-)
OS49.50	<i>orf10::Ωhyg</i>	8	(-)
OSC2	Aucun	10	(+)
SPM21	<i>orf2::att3Ωaac-</i>	10	(-)
SPM22	<i>orf2::att3</i> délétion en phase	10	(-)
SPM501	<i>orf6*::att1Ωhyg+</i>	14	(-)
SPM502	<i>orf6*::att1</i> délétion en phase	14	(+)
SPM507	<i>orf12::att3Ωaac-</i>	11	(-)
SPM508	<i>orf13c::att3Ωaac-</i>	12	(+)
SPM509	<i>orf14::att3Ωaac-</i>	13	(-)

Ces résultats permettent de tirer un certain nombre de conclusions en ce qui concerne la fonction des différents gènes impliqués dans la biosynthèse de la

spiramycine. Ainsi le gène *orf3* est essentiel à la biosynthèse de la spiramycine. En effet une inactivation en phase dans ce gène conduit à une souche (OS49.67, (cf. exemple 6)) ne produisant plus de spiramycine. L'inactivation en phase permet d'écarter l'hypothèse d'une éventuelle influence de la cassette introduite sur l'expression des gènes situés en aval de *orf3*.

De même, les gènes *orf8* et *orf10* codent des protéines essentielles à la biosynthèse de la spiramycine puisque les souches OS49.107 et OS49.50 ont un phénotype non producteur. De plus, dans ces deux dernières souches, c'est bien l'inactivation du gène correspondant qui est responsable de ce phénotype non producteur, puisqu'au vu de l'orientation des différentes orfs (cf. figure 3), la construction introduite ne peut avoir d'effet polaire.

L'étude des souches possédant une cassette excisable permet également de tirer un certain nombre de conclusion en ce qui concerne la fonction des gènes interrompus. La souche SPM507 possède le génotype : *orf12::att3 Ω aac-*. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf12*, au vu de l'orientation des gènes (cf. figure 3). Le fait que *orf13c* soit orienté en sens opposé à *orf12* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de selection à tout moment. Le phénotype de la souche SPM507 est non producteur, on peut donc en conclure que le gène *orf12* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

La souche SPM508 possède le génotype : *orf13c::att3 Ω aac-*. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf13c*, au vu de l'orientation des gènes (cf. figure 3). Le fait que *orf14* soit orienté en sens opposé à *orf13c* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de selection à tout moment. Le phénotype de la souche SPM508 est producteur, on peut

donc en conclure que le gène *orf13c* n'est pas un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

La souche SPM509 possède le génotype : *orf14::att3 Ω aac-*. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf14*, au vu de l'orientation des gènes (cf. figure 3), le fait que *orf15c* soit orienté en sens opposé à *orf14* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de selection à tout moment. Le phénotype de la souche SPM509 est non producteur, on peut donc en conclure que le gène *orf14* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

La souche SPM21 possède le génotype : *orf2::att3 Ω aac-*. Le phénotype de cette souche est non producteur de spiramycines. Cependant, l'orientation des gènes *orf1* à *orf8* (cf. figure 3) laisse penser que ces gènes sont cotranscrits. Aussi, le phénotype observé peut être dû à un effet polaire de la cassette introduite dans *orf2* sur l'expression de gènes situés en aval dans l'opéron. La souche SPM22 possède le génotype *orf2::att3* et a été obtenue après excision en phase de la cassette introduite. L'excision de la cassette laisse seulement une séquence « cicatricielle » caractéristique (cf. exemple 10). La souche SPM22 étant également de phénotype non producteur, on peut en conclure que le gène *orf2* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*. On observe ici uniquement l'effet dû à l'inactivation de *orf2*.

La souche SPM501 possède le génotype : *orf6*::att1 Ω hyg+*. Le phénotype de cette souche est non producteur de spiramycines. Cependant, les gènes *orf5** et *orf6** (cf. figure 3) ayant la même orientation, le phénotype observé peut être dû à un effet polaire de la cassette introduite dans *orf6** sur l'expression de *orf5**. La disposition de ces gènes laisse penser qu'ils peuvent être cotranscrits. Pour répondre à cette question, la souche SPM502 a été obtenue après excision en phase de la cassette introduite. Dans cette souche, on observe uniquement l'effet de l'inactivation de *orf6**. Cette souche

possède le génotype *orf6*::att1* (cf. exemple 14). L'excision de la cassette laisse seulement une séquence « cicatricielle » en phase (cf. exemple 14). La souche SPM502 possède un phénotype producteur (cette souche ne produit cependant que de la spiramycine I (cf*. exemple 16)). On peut donc conclure que le gène *orf5** est un gène
5 essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*, puisque son inactivation indirecte dans la souche SPM501 conduit à un phénotype non producteur. Par contre ; le gène *orf6** n'est pas un gène essentiel à la biosynthèse de la spiramycine I chez *S. ambofaciens* (par contre il est essentiel à la production de spiramycine II et III (cf. exemple 16)).

EXEMPLE 16. Dosage de la production des spiramycines I, II et III dans les souches mutantes obtenues

Les différentes souches à tester ont été cultivées chacune dans 7 erlenmeyers chicanés de 500 mL contenant 70 ml de milieu MP5 (Pernodet *et al.*, 1993). Les erlens
15 ont été inoculés par $2.5 \cdot 10^6$ spores/ml des différentes souches de *S. ambofaciens* et mises à pousser à 27°C sous agitation orbitale de 250 rpm pendant 72 heures. Les cultures correspondant au même clone sont rassemblées, éventuellement filtrées sur filtre plissé, et centrifugées 15 min à 7000 tr/min. Les différents surnageants ont ensuite été stockés à -30°C.

20 Les dosages ont été effectués par Chromatographie Liquide à Haute Performance (CLHP) par appariement d'ions. L'analyse CLHP du milieu de culture permet de déterminer précisément la concentration des trois formes de spiramycine. La colonne utilisée (Macherey-Nagel) est remplie avec une phase Nucleosil de silice greffée octyle. La granulométrie est de 5µm et la taille des pores 100Å. Le diamètre interne de la
25 colonne est de 4,6mm et sa longueur 25cm. La phase mobile est un mélange de tampon H_3PO_4 (pH2,2) et d'acétonitrile 70/30 (v/v) contenant 6,25g/L de perchlorate de $NaClO_4$. L'analyse est réalisée en régime isocratique avec un débit fixé à 1 ml/min. La

colonne est thermorégulée à 23°C. La détection est assurée par spectrophotométrie UV à 238nm. L'échantillon est réfrigéré à +10°C et la quantification est déterminée à partir de l'aire des pics (par étalonnage externe). Dans ces conditions, les temps de rétention de la spiramycine I, II et III sont respectivement d'environ 17; 21 et 30 minutes, comme
5 il a pu être vérifié grâce à un échantillon commercial comprenant les trois formes de spiramycine.

La souche OSC2 possède un phénotype producteur de spiramycine. C'est la souche parentale utilisée pour l'obtention des souches possédant une cassette excisable (cf. exemple 15). Cette souche a donc été utilisée comme témoin positif de production
10 des trois formes de spiramycine. Cette souche produit bien les trois formes de spiramycines comme il a été vérifié par CLHP (cf. figure 19).

L'étude des souches possédant une cassette excisable permet de tirer un certain nombre de conclusions en ce qui concerne la fonction des gènes interrompus. La souche SPM507 possède le génotype : *orf12::att3Ωaac-*. Le phénotype de la souche SPM507
15 est non producteur (cf. exemple 15), on peut donc en conclure que le gène *orf12* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*. Cette souche ne produit plus de spiramycines comme il a été vérifié par CLHP (cf. figure 22). Ce résultat confirme le caractère essentiel du gène *orf12* dans la biosynthèse de la spiramycine.

La souche SPM508 possède le génotype : *orf13c::att3Ωaac-*. Le phénotype de la
20 souche SPM508 est producteur de spiramycine (cf. exemple 15), on peut donc en conclure que le gène *orf13c* n'est pas un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*. Cette souche produit de la spiramycine I, II et III comme il a été vérifié par CLHP (cf. figure 23). Ce résultat confirme que le gène *orf13c* n'est pas un gène essentiel à la biosynthèse des spiramycines I, II et III ches *S.*
25 *ambofaciens*.

La souche SPM509 possède le génotype : *orf14::att3Ωaac-*. Le phénotype de la souche SPM509 est non producteur, on peut donc en conclure que le gène *orf14* est un

gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*. Cette souche ne produit plus de spiramycines comme il a été vérifié par CLHP (cf. figure 24). Ce résultat confirme l'essentialité du gène *orf14* dans la biosynthèse de la spiramycine.

La souche SPM501 possède le génotype : *orf6*::att1 Ω hyg+*. Le phénotype de
5 cette souche est non producteur de spiramycines. Cette souche ne produit plus de
spiramycines comme il a été vérifié par CLHP (cf. figure 20). Cependant, les gènes
*orf5** et *orf6** (cf. figure 3) ayant la même orientation, le phénotype observé peut être dû
à un effet polaire de la cassette introduite dans *orf6** sur l'expression de *orf5** dans
l'opéron. Ceci laisse penser que ces gènes sont cotranscrits. Pour répondre à cette
10 question, la souche SPM502 a été obtenue par excision de la cassette introduite,
produisant une délétion en phase dans le gène *orf6** et restaurant l'expression de *orf5**.
Cette souche possède le génotype *orf6*::att1* (cf. exemple 14 et 15). L'excision de la
cassette laisse seulement une séquence att « cicatricielle » en phase (cf. exemple 14). La
souche SPM502 possède un phénotype producteur de spiramycine. Cependant, comme il
15 a été prouvé par CLHP, cette souche ne produit que de la spiramycine I et ne produit pas
de spiramycine II et III (cf. figure 21). On peut donc conclure de ces résultats que le gène
*orf5** est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*,
puisque son inactivation indirecte dans la souche SPM501 conduit à un phénotype non
producteur de spiramycine (cf. figure 20). Par contre ; le gène *orf6** n'est pas un gène
20 essentiel à la biosynthèse de la spiramycine I chez *S. ambofaciens*, puisque l'inactivation
de ce gène conduit à un phénotype producteur de spiramycine I (cf. figure 21).
Cependant, *orf6** est essentiel à la production de spiramycine II et III (cf. exemple 16)).
Le gène *orf6** code donc bien une acyl-transférase responsable de la modification du
platénolide à la position 3 (cf. figure 1).

EXEMPLE 17 : Détermination du point de démarrage de la traduction de *orf10* et amélioration de la production en spiramycines

17.1 Construction des plasmides pSPM523, pSPM524 et pSPM525 :

Le gène *orf10* a été identifié chez *Streptomyces ambofaciens* et a été désigné *srnR* par Geistlich *et al.* (Geistlich, M., *et al.*, 1992). L'inactivation du gène *orf10* a été réalisée (cf. exemple 8). Il a pu ainsi être montré que la souche résultante ne produit plus de spiramycines (cf. exemple 15). Ceci confirme que le gène *orf10* est bien impliqué dans la biosynthèse des spiramycines. La protéine codée par ce gène est donc bien essentielle à la biosynthèse des spiramycines. Cependant, l'analyse de la séquence montre que deux codons ATG situés dans la même phase de lecture peuvent être utilisés pour la traduction de *orf10* (cf. Figure 28). Un des deux codons possible (le codons le plus en amont) démarre à la position 10656 de la séquence donnée en SEQ ID N° 1, alors que l'autre codon possible situé plus en aval démarre à la position 10809 (cf. SEQ ID N° 1). Avant de tester un éventuel effet de la surexpression de *srnR* sur la production de spiramycine, il est important de déterminer dans un premier temps le point de démarrage de la traduction.

Dans le but de déterminer le site de démarrage de la traduction, trois constructions comportant trois formes d'*orf10* ont été réalisées. Ces dernières ont été obtenues par PCR en utilisant des oligonucléotides comportant soit un site de restriction *HindIII* soit un site de restriction *BamHI*.

Le premier couple utilisé pour l'amplification correspond aux oligonucléotides suivants :

EDR39 :

5'CCCAAGCTTGAGAAGGGAGCGGACATTCATGGCCCGCGCCGAACGC3'

(SEQ ID N° 122) (le site *HindIII* figure en gras)

EDR42 :

5'CGGGATCCGGCTGACCATGGGAGACGGGCGCATCGCCGAGTTCAGC3'

(SEQ ID N° 123) (le site *BamHI* figure en gras)

Le couple d'amorces EDR39-EDR42 permet l'amplification d'un fragment d'*orf10* comportant l'ATG situé le plus en 3' (position 10809 dans la séquence donnée en SEQ ID N° 1) (cf. Figure 28). Le fragment obtenu fait une taille d'environ 2 kb et sera appelé par la suite "*orf10* court", il ne contient pas le promoteur de *orf10*. Ce
5 fragment de 2kb a été cloné dans le vecteur pGEM-T easy pour donner naissance au plasmide pSPM520.

Le deuxième couple utilisé pour l'amplification correspond aux oligonucléotides suivants :

EDR40 :

10 5'CCCAAGCTTGAGAAGGGAGCGGACATTCAATGCTTTGGTAAAGCAC3'
(SEQ ID N° 124) (le site *HindIII* figure en gras)

EDR42 :

5'CGGGATCCGGCTGACCATGGGAGACGGGCGCATCGCCGAGTTCAGC3'
(SEQ ID N° 123) (le site *BamHI* figure en gras)

15 Le couple d'amorces EDR40-EDR42 permet l'amplification d'un fragment d'*orf10* comportant l'ATG situé le plus en 5' (position 10656 dans la séquence donnée en SEQ ID N° 1) (cf. Figure 28). Ce fragment de 2,2 kb appelé par la suite "*orf10* long" a été cloné dans le vecteur pGEM-T easy pour donner naissance au plasmide pSPM521, ce plasmide ne contient pas le promoteur de *orf10*.

20 Le troisième couple utilisé pour l'amplification correspond aux oligonucléotides suivants :

EDR41 :

5'-CCCAAGCTTTCAAGGAACGACGGGGTGGTCAGTCAAGT-3'
(SEQ ID N° 125) (le site *HindIII* figure en gras)

25 EDR42 :

5'CGGGATCCGGCTGACCATGGGAGACGGGCGCATCGCCGAGTTCAGC3'
(SEQ ID N° 123) (le site *BamHI* figure en gras)

Le couple d'amorces EDR41-EDR42 permet l'amplification du gène *orf10* avec les deux ATG, ainsi que son propre promoteur (cf. Figure 28). Ce fragment de 2,8 kb

appelé par la suite "*orf10* pro" a été cloné dans le vecteur pGEM-T easy pour donner naissance au plasmide pSPM522.

Le fragment "*orf10* pro" a été obtenu en utilisant comme matrice l'ADN chromosomique de la souche OSC2. Les fragments "*orf10* court" et "*orf10* long" ont
5 quant à eux été obtenus en utilisant comme matrice l'ADN du fragment "*orf10* pro" préalablement purifié.

Les inserts *Hind*III-*Bam*HI des plasmides pSPM520, pSPM521 et pSPM522 ont ensuite été sous-clonés dans le vecteur pUWL201 (plasmide dérivé du plasmide pUWL199 (Wehmeier UF, 1995) dans lequel le fragment *Kpn*I-*Bam*HI de la région
10 du promoteur de *ermE* (cf. Bibb et al., 1985, notamment figure 2) portant une mutation augmentant la force du promoteur (promoteur *ermE**) (Bibb et al., 1994) a été introduit (cf. Doumith et al., 2000)) préalablement digéré par les enzymes *Hind*III-*Bam*HI. Ainsi, il a été obtenu trois plasmides : pSPM523 (dérivé du pUWL201 avec la forme "*orf10* court" comme insert), pSPM524 (dérivé du
15 pUWL201 avec la forme "*orf10* long" comme insert) et pSPM525 (dérivé du pUWL201 avec la forme "*orf10* pro") (Figure 28).

17.2 Transformation de la souche OS49.50 par les constructions pSPM523, pSPM524 et pSPM525 :

La souche OS49.50 (souche interrompue dans le gène *orf10*, cf. exemple 8) a été
20 transformée indépendamment par transformation de protoplastes (Kieser, T et al., 2000) par chacun des plasmides pSPM523, pSPM524 et pSPM525. Un témoin négatif a également été réalisé en transformant la souche OS49.50 par le plasmide pUWL201 sans insert. Après transformation des protoplastes, les clones sont sélectionnés pour leur résistance au thiostrepton. La transformation des clones par chacun des plasmides est
25 vérifiée par extraction de ces plasmides. Ainsi quatre nouvelles souches ont été obtenues : la souche OSC49.50 pUWL201, issue de la transformation par le plasmide pUWL201 sans insert, la souche OSC49.50 pSPM523, issue de la transformation par le plasmide pSPM523, la souche OSC49.50 pSPM524, issue de la transformation par le

plasmide pSPM524 et enfin la souche OS49.50 pSPM525, issue de la transformation par le plasmide pSPM525.

La production de spiramycines de chacune de ces quatre souches a été testée par CLHP. Pour cela, les différentes souches de *Streptomyces* à tester ont été
5 cultivées dans des erlenmeyers à baffles (erlenmeyers chicanés) de 500 ml contenant 70 ml de milieu MP5 (Pernodet *et al.*, 1993). Lorsque la souche contient le plasmide pUWL201 ou l'un de ses dérivés, il est ajouté 5 µg/ml de thiostrepton. Les erlenmeyers chicanés ont été inoculés à une concentration initiale de $2.5 \cdot 10^6$ spores/ml des différentes souches de *S. ambofaciens* et les cultures ont été incubées à
10 27°C sous agitation orbitale de 250 rpm pendant 96 heures. Les cellules ont ensuite été séparées du milieu par centrifugation et le surnageant a été analysé par CLHP (cf. exemple 16) pour déterminer la quantité de spiramycine produite. Grâce à un échantillon étalon et à la mesure de l'aire des pics, il a été possible de déterminer la quantité de chacune des spiramycines produites par ces souches. Les résultats de cette
15 analyse sont présentés dans le tableau 39, les données sont exprimées en mg par litre de surnageant. Les résultats correspondent à la production totale en spiramycines (obtenue en additionnant la production en spiramycine I, II et III).

Tableau 39. Production en spiramycines des souches dérivées de OS49.50, (résultats exprimés en mg/l).

Souche	Production en Spiramycines
OS49.50 pUWL201	0
OS49.50 pSPM523	0
OS49.50 pSPM524	93

OS49.50 pSPM525	149
--------------------	-----

Comme le montre les résultats donnés dans le tableau 39, le témoin négatif (souche OS49.50 transformée par le plasmide pUWL201) ne produit pas de spiramycine. Lorsque le plasmide pSPM523 (qui contient la forme "*orf10* court") est introduit dans la souche OS49.50, aucune production de spiramycine n'est observée.

- 5 En revanche, la présence du plasmide pSPM524 (qui contient la forme "*orf10* long") et du plasmide pSPM525 (qui contient la forme "*orf10* pro") restaure la production de spiramycine dans la souche hôte OS49.50. Ainsi, seuls les fragments de *orf10* contenant l'ATG le plus en amont permettent de restaurer la synthèse de spiramycine.

- 10 Dans le but de confirmer ces résultats, le plasmide pSPM521 (plasmide pGEM-T easy contenant la forme "*orf10* long") a été digéré par l'enzyme de restriction *Xho*I (cette enzyme possède un site unique dans ce plasmide, localisé entre les deux ATG (cf. Figure 28)). Les extrémités *Xho*I ont ensuite été rendues bout franc par traitement à l'enzyme de Klenow. Le plasmide a alors été refermé sur lui-même par action de la T4 DNA ligase pour donner naissance au plasmide pSPM527. Si
- 15 l'ATG le plus en amont (position 10656 dans la séquence donnée en SEQ ID N° 1) est bien utilisé comme site d'initiation de la traduction, ce traitement entraînera un décalage du cadre de lecture au niveau du site *Xho*I et aura pour effet la production d'une protéine ne présentant pas d'activité activatrice. Par contre si l'initiation de la traduction a lieu au niveau de l'ATG plus en aval (position 10809 dans la séquence
- 20 donnée en SEQ ID N° 1), ce traitement devrait avoir peu ou pas d'effet sur l'expression de Orf10 (compte tenu de la localisation du point de démarrage de la transcription) et aucun effet sur la protéine produite.

- L'insert *Bam*HI-*Hind*III de pSPM527 a ensuite été sous-cloné dans le vecteur pUWL201 pour donner naissance au plasmide pSPM528. Ce plasmide a été introduit
- 25 dans la souche OS49.50 et un clone possédant le plasmide voulu a plus particulièrement été sélectionné. La production de spiramycine de la souche

résultante a ensuite été testée par CLHP (Cf. exemple 16 et ci-dessus). Contrairement à ce qui avait été observé avec le plasmide pSPM524 (cf. tableau 39), la présence du plasmide pSPM528 dans la souche OS49.50 ne rétablit pas la production de spiramycine. Ceci confirme que le démarrage de la traduction du gène *orf10* est l'ATG situé le plus en aval (ATG 1 cf. Figure 28).

17.3 Amélioration de la production en spiramycines de la souche OSC2 de *S. ambofaciens* :

Afin de tester l'effet de la surexpression du gène *orf10* sur la production de spiramycines, les plasmides pSPM523, pSPM524, pSPM525 et pSPM528 ont été introduits dans la souche OSC2. Pour cela, des protoplastes de la souche OSC2 ont été transformés (Kieser, T *et al.*, 2000) indépendamment par chacun des plasmides pSPM523, pSPM524, pSPM525 et pSPM528. Un témoin négatif a également été réalisé en transformant la souche OSC2 par le plasmide pUWL201 sans insert. Après transformation des protoplastes, les clones sont sélectionnés pour leur résistance au thiostrepton. Ainsi cinq nouvelles souches ont été obtenues : la souche OSC2 pUWL201, issue de la transformation par le plasmide pUWL201 sans insert, la souche OSC2 pSPM523, issue de la transformation par le plasmide pSPM523, la souche OSC2 pSPM524, issue de la transformation par le plasmide pSPM524, la souche OSC2 pSPM525, issue de la transformation par le plasmide pSPM525 et enfin la souche OSC2 pSPM528, issue de la transformation par le plasmide pSPM528. La production en spiramycines de ces souches a alors été analysée par CLHP (de la même manière que dans l'exemple 17.2). L'analyse de la production en spiramycine de la souche OSC2 a également été effectuée en parallèle pour comparaison. Les résultats de cette analyse sont présentés dans le tableau 40, les données sont exprimées en mg par litre de surnageant. Les résultats correspondent à la production totale en spiramycines (obtenue en additionnant la production en spiramycine I, II et III).

Tableau 40. Production en spiramycines des souches dérivées de OSC2, (résultats exprimés en mg/l).

Souche	Production en spiramycines
OSC2	69
OSC2 pUWL201	103
OSC2 pSPM523	19
OSC2 pSPM524	135
OSC2 pSPM525	278
OSC2 pSPM528	68

Ainsi, il est observé que la présence du plasmide pSPM524 ou du plasmide
5 pSPM525 augmente significativement la production en spiramycines de la souche
OSC2. Ceci démontre bien que la surexpression de Orf10 a un effet positif sur la
production en spiramycines. Le plasmide pSPM528 n'a par contre pas d'effet sur la
production en spiramycines.

Les plasmides pSPM525 et pUWL201 ont de la même manière été introduit
10 dans la souche SPM502 (cf. exemple 14). Ainsi deux nouvelles souches ont été
obtenues : la souche SPM502 pUWL201, issue de la transformation par le plasmide

pUWL201 sans insert, et la souche SPM502 pSPM525, issue de la transformation par le plasmide pSPM525.

Un échantillon de la souche SPM502 pSPM525 (cette souche contient le plasmide pSPM525, cf. ci-dessus) a été déposé auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 26 février 2003 sous le numéro d'enregistrement I-2977.

La production en spiramycines des souches SPM502 pUWL201 et SPM502 pSPM525 a été analysée par CLHP (de la même manière que dans l'exemple 17.2). L'analyse de la production en spiramycines de la souche SPM502 a également été effectuée en parallèle pour comparaison. Les résultats de cette analyse sont présentés dans le tableau 41, les données sont exprimées en mg par litre de surnageant. Les résultats correspondent à la production en spiramycine I. En effet, aucune de ces souches ne produit de spiramycine II et III.

Tableau 41. Production en spiramycine I des souches dérivées de SPM502, (résultats exprimés en mg/l).

Souche	Spiramycine I
SPM502	47
SPM502 pUWL201	72
SPM502 pSPM525	130

Ainsi, il a pu être observé que la surexpression du gène *orf10* dans la souche SPM502 augmente de façon importante la production de spiramycine I.

EXEMPLE 18 : Construction d'une banque d'ADN génomique de la souche OSC2 de *Streptomyces ambofaciens* chez *E. coli* dans la cosmide pWED2.

18.1 Construction du cosmide pWED2 :

Dans le but de faciliter l'inactivation de gènes chez *Streptomyces*, un cosmide portant la séquence *oriT* du plasmide RK2 (ce qui permet son introduction par conjugaison dans *Streptomyces* à partir d'une souche appropriée de *E. coli*) et portant également un gène de résistance à un antibiotique conférant un phénotype repérable chez *Streptomyces*, a été construit. Un tel cosmide contenant de grands inserts d'ADN génomique de *Streptomyces ambofaciens* peut être utilisé dans des expériences d'inactivation de gènes.

Pour construire ce vecteur, une cassette *pac-oriT* (fragment *EcoRV*) a été introduite dans le cosmide pWED1 (Gourmelen *et al.*, 1998), dérivé du cosmide pWED15 (Wahl *et al.*, 1987), au niveau du site unique *HpaI*. La cassette *pac-oriT* a été obtenue par PCR. Pour cela, le gène *pac* a été amplifié par PCR à partir du plasmide pVF 10.4 (Vara *et al.*, 1985; Lacalle *et al.*, 1989) en utilisant comme première amorce, l'amorce A (de séquence 5'-CCAGTAGATATCCCGCCAACCCGGAGCTGCAC-3' (SEQ ID N° 126), le site de restriction *EcoRV* a été souligné et la séquence en gras de 20 nucléotides correspond à une région en amont du promoteur du gène *pac*) et comme deuxième amorce, l'amorce B (de séquence 5'-GAAAAGATCCGTCATGGGGTCGTGCGCTCCTT-3' (SEQ ID N° 127), qui comporte à son extrémité 5' une séquence de 12 nucléotides correspondant au début de la séquence *oriT* (souligné double) et une séquence de 20 nucléotides (en gras) correspondant à la fin du gène *pac* (cf. Figure 29, 1^{ère} PCR).

Le gène *oriT* a quant à lui été amplifié par PCR à partir du plasmide pPM803 (Mazodier, P. *et al.*, 1989) en utilisant comme première amorce, l'amorce C (de séquence 5'-CACGACCCCATGACGGATCTTTCCGCTGCAT-3' (SEQ ID N° 128)), qui comporte à son extrémité 5' une séquence de 12 nucléotides correspondant à une séquence en aval de la séquence codante du gène *pac* (en gras) et une séquence de

20 nucléotides correspondant au début de la séquence *oriT*) et comme deuxième
amorce, l'amorce D (de séquence 5'-
GAGCCGGATATCATCGGTCTTGCCTTGCTCGT-3' (SEQ ID N° 129)) qui
comporte le site de restriction *EcoRV* (souligné simple) et une séquence de 20
5 nucléotides correspondant à la fin de la séquence *oriT* (souligné double) (cf. Figure 29,
2^{ème} PCR).

Le produit d'amplification obtenu avec les amorces A et B et celui obtenu avec
les amorces C et D ont à une de leur extrémité, une séquence commune de 24
nucléotides. Une troisième PCR a été réalisée en mélangeant les deux produits
10 d'amplification obtenus précédemment et en utilisant comme amorces, les amorces A et
D (cf. figure 29, 3^{ème} PCR). Ceci a permis d'obtenir un produit d'amplification
correspondant à l'ensemble *pac+oriT*. Ce fragment *pac-oriT* a alors été cloné dans le
vecteur pGEM-T Easy (commercialisé par la société Promega (Madison, Wisconsin,
USA)), ce qui a permis d'obtenir le plasmide pGEM-T-*pac-oriT*. L'insert de ce
15 plasmide a ensuite été sous cloné dans le cosmide pWED1. Pour cela le plasmide
pGEM-*pac-oriT* a été digéré par l'enzyme *EcoRV* et l'insert *EcoRV* contenant
l'ensemble *pac+oriT* a été inséré dans le cosmide pWED1 ouvert au préalable par
l'enzyme *HpaI*. Le cosmide ainsi obtenu a été baptisé pWED2 (cf. Figure 30).

Ce cosmide permet de faciliter l'inactivation de gènes chez *Streptomyces*.
20 En effet, il porte la séquence *oriT* (ce qui permet son introduction par conjugaison
dans *Streptomyces* à partir d'une souche appropriée de *E. coli*) mais également un
gène de résistance à un antibiotique conférant un phénotype repérable chez
Streptomyces. Un tel cosmide contenant de grands inserts d'ADN génomique de
Streptomyces ambofaciens peut être utilisé dans des expériences d'inactivation de
25 gènes.

Ainsi un cosmide dérivé de pWED2 contenant le gène cible pourra par
exemple être introduit dans la souche de *E. coli* KS272 contenant le plasmide
pKOBEG (cf. Chaveroche *et al.* 2000) et une cassette sera introduite dans le gène
cible selon la technique décrite par Chaveroche *et al.* 2000. Le cosmide obtenu par
30 cette technique (cosmide dans lequel le gène cible est inactivé), pourra ensuite être

introduit dans une souche de *E. coli* telle que la souche S17.1 ou toute autre souche permettant de transférer par conjugaison des plasmides contenant la séquence *oriT* vers *Streptomyces*.

Après conjugaison entre *E. coli* et *Streptomyces*, des clones de *Streptomyces* dans lesquelles la copie sauvage du gène cible aura été remplacée par la copie interrompue pourront être obtenues comme décrit dans l'exemple 2.

Le gène de résistance exprimé chez *Streptomyces* présent sur ce nouveau cosmide est le gène *pac* de *Streptomyces alboniger* (Vara, J *et al.*, 1985 ; Lacalle *et al.*, 1989) qui code la puromycine N-acétyl transférase et qui confère la résistance à la puromycine. Lors d'expériences d'inactivation de gène, on cherchera les clones dans lesquels un double événement de recombinaison aura eu lieu. Ces clones auront éliminer le cosmide pWED2 et seront donc redevenus sensibles à la puromycine.

18.2 Construction d'une banque d'ADN génomique de la souche OSC2 de *Streptomyces ambofaciens* chez *E. coli* dans le cosmide pWED2

L'ADN génomique de la souche OSC2 de *Streptomyces ambofaciens* a été digéré partiellement par l'enzyme de restriction *Bam*HI de manière à obtenir des fragments d'ADN de taille comprise entre environ 35 et 45 kb. Ces fragments ont ensuite été clonés dans le cosmide pWED2, ce dernier ayant été au préalable digéré par *Bam*HI, puis traité par la phosphatase alcaline. Le mélange de ligation a ensuite été encapsidé *in vitro* dans des particules de phages lambda grâce au système « Packagene® Lambda DNA packaging system » commercialisé par la société Promega (Madison, Wisconsin, USA) en suivant les recommandations du fabricant. Les particules phagiques obtenues ont été utilisées pour infecter la souche SURE® d'*E. coli* commercialisée par la société Stratagene (LaJolla, Californie, USA). La sélection des clones a été effectuée sur milieu LB + ampicilline (50 µg/ml), le cosmide pWED2 conférant la résistance à l'ampicilline.

EXEMPLE 19 : Isolement de cosmides de la nouvelle banque couvrant la région de la voie de biosynthèse des spiramycines. Sous clonage et séquençage de fragments de ces cosmides.

5 19.1 Hybridation sur colonies de la banque génomique de *Streptomyces ambofaciens* OSC2 :

Des cosmides de la nouvelle banque de *Streptomyces ambofaciens* OSC2 (cf. exemple 18) couvrant les *orf1** à *orf10** ou une partie ou la totalité des *orf1* à *orf25c*, ou une région plus en amont de l'*orf25c* ont été isolés. Pour cela, des hybridations successives sur les colonies d'*E. coli* obtenues selon l'exemple 18 ont été réalisées en
10 utilisant les 3 sondes suivantes (cf. figure 31):

- La première sonde utilisée correspond à un fragment d'ADN d'environ 0,8kb amplifié par PCR en utilisant comme matrice le cosmide pSPM5 et les amorces suivantes :

15 ORF23c : 5'-ACGTGCGCGGTGAGTTCGCCGTTGC-3' (SEQ ID N° 130) et
ORF25c : 5'-CTGAACGACGCCATCGCGGTGGTGC-3' (SEQ ID N° 131).

Le produit de PCR ainsi obtenu contient un fragment du début de l'*orf23c*, l'*orf24c* en entier et la fin de l'*orf25c* (cf. figure 31, sonde I).

20 - La deuxième sonde utilisée correspond à un fragment d'ADN d'environ 0,7kb amplifié par PCR en utilisant comme matrice l'ADN total de la souche *S. ambofaciens* ATCC23877 et les amorces suivantes :

ORF1*c : 5'-GACCACCTCGAACCGTCCGGCGTCA-3' (SEQ ID N° 132) et
ORF2*c : 5'-GGCCCGGTCCAGCGTGCCGAAGC-3' (SEQ ID N° 133).

Le produit de PCR ainsi obtenu contient un fragment de la fin de l'*orf1*c* et du début de l'*orf2*c* (cf. figure 31, sonde II).

25 - La troisième sonde utilisée correspond à un fragment *EcoRI-BamHI* d'environ 3kb contenant les *orf1*, *orf2* et *orf3* et obtenu par digestion du plasmide pOS49.99 (cf. figure 31, sonde III).

Environ 2000 clones de la banque obtenue dans l'exemple 18.2 ont été transférés sur filtre pour hybridation sur colonies selon les techniques classiques (Sambrook *et al.*, 1989).

La première sonde (cf. figure 31, sonde I) a été marquée au ^{32}P par la technique
5 du « random priming » (Kit commercialisé par la société Roche) et utilisées pour
l'hybridation sur 2000 clones de la banque, après transfert sur filtre. L'hybridation a été
effectuée à 65°C dans le tampon décrit par Church & Gilbert (Church & Gilbert, 1984).
Deux lavages ont été effectués en 2x SSC, 0,1% SDS à 65°C, le premier pendant 10
minutes et le deuxième pendant 20 minutes et un troisième lavage d'une durée de 30
10 minutes a ensuite été effectué en 0.2X SSC, 0,1% SDS à 65°C. Dans ces conditions
d'hybridation et de lavage, 20 clones parmi les 2000 hybridés présentaient un fort signal
d'hybridation avec la première sonde. Ces 20 clones ont été cultivés en milieu LB+
ampicilline (50 µg/ml) et les 20 cosmides correspondants ont été extraits par lyse
alcaline (Sambrook *et al.*, 1989) puis ils ont été digérés par l'enzyme de restriction
15 *Bam*HI. Les produits de digestions ont ensuite été séparés sur gel d'agarose, transférés
sur membrane de Nylon et hybridés par la première sonde (cf. ci-dessus : le produit de
PCR ORF23c-ORF25c, sonde I) dans les mêmes conditions que ci-dessus. Douze
cosmides possédaient un fragment *Bam*HI hybridant fortement avec la sonde utilisée.
Ces 12 cosmides ont été nommés pSPM34, pSPM35, pSPM36, pSPM37, pSPM38,
20 pSPM39, pSPM40, pSPM41, pSPM42, pSPM43, pSPM44 et pSPM45. Les profils,
après digestion par *Bam*HI, de ces 12 cosmides ont été comparés entre eux et avec celui
du cosmide pSPM5. De plus des expériences d'amplification par PCR en utilisant
différentes amorces correspondant à différents gènes déjà identifiés dans la région *orf1-*
orf25c ont permis de positionner l'insert de certains de ces cosmides les uns par rapport
25 aux autres et de déterminer également la localisation de ces inserts dans la région *orf1-*
orf25c déjà connue (cf. Figure 32). Le cosmide pSPM36 a plus particulièrement été
choisi car il était susceptible de contenir une grande région en amont de l'*orf25c* (cf.
Figure 31 et 32).

Dans un deuxième temps, et en utilisant les mêmes conditions que celles décrites ci-dessus, les 2000 clones de la banque de *Streptomyces ambofaciens* OSC2 ont été hybridés avec la deuxième sonde correspondant au produit de PCR : ORF1*c-ORF2*c (cf. figure 31, sonde II). Cette hybridation a permis d'isoler des cosmides dont l'insert est situé dans la région de *orf1*c* à *orf10*c*. Dans les conditions d'hybridation utilisés, 16 clones parmi les 2000 hybridés présentaient un fort signal d'hybridation avec cette deuxième sonde. Ces 16 clones ont été cultivés en milieu LB+ ampicilline (50 µg/ml) et les 16 cosmides correspondants ont été extraits par lyse alcaline (Sambrook *et al.*, 1989) puis ils ont été digérés par l'enzyme de restriction *Bam*HI. Les profils de digestion (après digestion par *Bam*HI) de ces 16 cosmides ont été comparés entre eux et avec celui du cosmide pSPM7. Des expériences d'amplification par PCR avec les amorces ORF1*c et ORF2*c ont permis de choisir deux cosmides qui renfermaient bien les gènes *orf1*c* et *orf2*c* et dont les profils possédaient des bandes communes mais aussi des bandes différentes. De plus d'autres expériences d'amplification par PCR en utilisant différentes amorces correspondant à différents gènes déjà identifiés dans la région *orf1*c* à *orf10*c* ont permis de positionner l'insert de ces cosmides les uns par rapport aux autres et de déterminer également la localisation de ces inserts dans la région *orf1*c* à *orf10*c* déjà connue (cf. Figure 32). Les deux cosmides plus particulièrement sélectionnés ont été nommés pSPM47 et pSPM48 (cf. Figure 32).

En utilisant les mêmes conditions que celles décrites ci-dessus, les 2000 clones de la banque de *Streptomyces ambofaciens* OSC2 ont également été hybridés avec la troisième sonde correspondant au fragment d'ADN *Eco*RI-*Bam*HI du plasmide pOS49.99 (cf. figure 31 sonde III). Cette hybridation a permis d'isoler les cosmides contenant la région de l'*orf1* jusqu'à l'*orf3* et susceptibles de contenir soit une grande partie des gènes des PKS, soit une grande partie des gènes *orf1* à *orf25c* de la voie de biosynthèse des spiramycines. Dans ces conditions d'hybridation, 35 clones parmi les 2000 hybridés présentaient un fort signal d'hybridation avec la troisième sonde. Ces 35 clones ont été cultivés en milieu LB+ ampicilline (50 µg/ml) et les 35 cosmides correspondants ont été extraits par lyse alcaline (Sambrook *et al.*, 1989) puis digérés par

l'enzyme de restriction *Bam*HI. Les profils, après digestion par *Bam*HI, de ces 35 cosmides ont été comparés entre eux et avec celui du cosmide pSPM5. De plus des expériences d'amplification par PCR en utilisant différentes amorces correspondant à différents gènes déjà identifiés dans la région *orf1* à *orf25c* ont permis de vérifier que ces cosmides renfermaient bien des inserts provenant de la région *orf1* à *orf25c* et de positionner les inserts de ces cosmides les uns par rapport aux autres et par rapport à la région *orf1* à *orf25c* déjà connue (cf. Figure 32). Cinq cosmides ont été plus particulièrement sélectionnés, ils ont été nommés pSPM50, pSPM51, pSPM53, pSPM55 et pSPM56 (cf. Figure 32).

10 19.2 Sous clonage et séquençage d'une partie de l'insert du cosmide pSPM36.

La sonde d'environ 0,8 kb obtenue par PCR avec les amorces ORF23c et ORF25c (cf. ci-dessus et figure 31, sonde I) a également été utilisée dans des expériences de Southern Blot sur l'ADN total de *S. ambofaciens* OSC2 digéré par l'enzyme *Pst*I. Dans les conditions d'hybridation décrite ci-dessus, cette sonde révèle un fragment unique *Pst*I d'environ 6 kb lorsque hybridé sur l'ADN total de *S. ambofaciens* OSC2 digéré par l'enzyme *Pst*I. Il existe un site *Pst*I au niveau de l'*orf23c* (cf. SEQ ID N° 80) mais aucun autre site *Pst*I jusqu'à l'extrémité (site *Bam*HI) de la séquence connue (cf. SEQ ID N° 1). Ce fragment *Pst*I-*Bam*HI a une taille d'environ 1.4 kb. Le fragment *Pst*I de 6 kb hybridé sur l'ADN total de *S. ambofaciens* digéré par l'enzyme *Pst*I contient donc une région d'environ 4,6 kb située en amont de *orf25c*. Cette région est susceptible de contenir d'autres gènes dont les produits sont impliqués dans la voie de biosynthèse de la spiramycine. Il a été vérifié par digestion que le cosmide pSPM36 contenait bien ce fragment *Pst*I de 6 kb. Ce fragment a été isolé à partir de pSPM36, dans le but de déterminer la séquence plus en amont de *orf25c*. Pour cela, le cosmide pSPM36 a été digéré par l'enzyme de restriction *Pst*I. Le fragment *Pst*I-*Pst*I d'une taille d'environ 6kb a été isolé par électroélution à partir d'un gel d'agarose 0.8% puis cloné dans le vecteur pBK-CMV (4512bp) (commercialisé par la société Stratagene (La Jolla, Californie, USA)). Le plasmide ainsi obtenu a été nommé pSPM58 (cf. Figure 33) et la séquence de son insert a été déterminée. La séquence de cet insert est présentée en SEQ ID N° 134.

Toutefois, toute la séquence n'a pas été déterminée et il subsiste un trou d'environ 450 nucléotides, la partie de la séquence non déterminée a été notée par une succession de « N » dans la séquence correspondante.

19.3 Analyse des nouvelles séquences nucléotidiques, détermination des phases
5 ouvertes de lecture et caractérisation des gènes impliqués dans la biosynthèse des
spiramycines.

La séquence de l'insert du cosmide pSPM58 obtenue a été analysée grâce au programme FramePlot (Ishikawa J & Hotta K. 1999). Ceci a permis d'identifier, parmi les phases ouvertes de lecture, les phases ouvertes de lecture présentant un usage des
10 codons typique de *Streptomyces*. Cette analyse a permis de déterminer que cet insert comporte 4 nouvelles ORFs en amont de l'*orf25c* (cf. figure 34). Ces gènes ont été nommés *orf26* (SEQ ID N° 107), *orf27* (SEQ ID N° 109), *orf28c* (SEQ ID N° 111, la séquence de cette orf n'a pas été déterminée complètement puisqu'un trou d'environ 450 nucléotide subsiste lors du séquençage de l'insert de pSPM58, ces 450 nucléotides
15 figurent sous la forme d'une série de « N » dans la séquence SEQ ID N° 111) et *orf29* (la séquence de cette dernière orf était incomplète dans cet insert). Le « c » ajouté dans le nom du gène signifiant pour l'ORF en question que la séquence codante est dans l'orientation inverse (cf. figure 34).

Les séquences protéiques déduites de ces phases ouvertes de lecture ont été
20 comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et al.*, 1997), CD-search, (ces deux programmes sont accessibles notamment auprès du National Center for Biotechnology Information (NCBI) (Bethesda, Maryland, USA)), FASTA ((Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990) (ce programme est accessible
25 notamment auprès du centre de ressources INFOBIOGEN, Evry, France). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycines.

19.4 Sous clonage et séquençage d'une autre partie de l'insert du cosmide pSPM36.

La sonde d'environ 0,8 kb obtenue par PCR avec les amorces ORF23c et ORF25c (cf. ci-dessus et figure 31, sonde I) a également été utilisée dans des expériences de Southern Blot sur l'ADN total de la souche OSC2 digéré par l'enzyme *StuI*. Dans les conditions d'hybridation décrite ci-dessus pour cette sonde, cette sonde révèle un fragment unique *StuI* d'environ 10 kb lorsque hybridé sur l'ADN total de la souche OSC2 digéré par l'enzyme *StuI*. Compte tenu de la présence d'un site *StuI* dans l'*orf23c* (cf. SEQ ID N° 80) et de la localisation de ce site par rapport au site *PstI*, ce fragment de 10 kb inclut la totalité du fragment *PstI* précédemment étudié (insert de pSPM58) et permet d'avoir accès à une région non encore étudiée d'environ 4 kb. (cf figure 33). Il a été vérifié par digestion que le cosmide pSPM36 contenait bien ce fragment *StuI* de 10 kb. Ce fragment a été isolé à partir du cosmide pSPM36, dans le but de déterminer la séquence de la fin d'*orf29* et d'autres gènes plus en amont de *orf29*. Pour cela, le cosmide pSPM36 a été digéré par l'enzyme de restriction *StuI*. Le fragment *StuI-StuI* d'une taille d'environ 10 kb a été isolé par électroélution à partir d'un gel d'agarose 0.8% puis cloné dans le vecteur pBK-CMV (4512bp) (commercialisé par la société Stratagene (La Jolla, Californie, USA)). Le plasmide ainsi obtenu a été nommé pSPM72 (cf. Figure 33). Ce dernier a ensuite été digéré par *EcoRI* (site *EcoRI* dans l'insert de pSPM58) et *HindIII* (site *HindIII* dans le site de clonage multiple du vecteur, immédiatement après le site *StuI* de l'extrémité de l'insert) (cf. Figure 33). Le fragment d'ADN *EcoRI-HindIII* ainsi obtenu, correspond à un fragment de l'insert du plasmide pSPM72 (cf. figure 33) et a été sous-cloné dans le vecteur pBC-SK+ (commercialisé par la société Stratagene (La Jolla, Californie, USA)) digéré au préalable par *EcoRI* et *HindIII*. Le plasmide ainsi obtenu a été nommé pSPM73 et la séquence de son insert a été déterminée. La séquence de cet insert est présentée en SEQ ID N° 135.

Un assemblage des séquences des inserts de pSPM58 et pSPM73 est présenté en SEQ ID N° 106. Cette séquence part du site *PstI* au niveau de l'*orf23c* (cf. SEQ ID N° 80) et va jusqu'au site *StuI* au niveau de l'*orf32c* (cf. figure 34). La séquence de l'*orf28c* (SEQ ID N° 111), n'étant pas complète (cf. exemple 19.3), une région d'environ 450

nucléotides n'est pas séquencée, ces 450 nucléotides figurent sous la forme d'une série de « N » dans la sequence SEQ ID N° 106).

5 19.5 Analyse des nouvelles séquences nucléotidiques, détermination des phases
ouvertes de lecture et caractérisation des gènes impliqués dans la biosynthèse des
spiramycines.

La séquence partielle de l'insert du cosmide pSPM73 obtenue a été analysée grâce
au programme FramePlot (Ishikawa J & Hotta K. 1999). Ceci a permis d'identifier,
parmi les phases ouvertes de lecture, les phases ouvertes de lecture présentant un usage
10 des codons typique de *Streptomyces*. Cette analyse a permis de déterminer que cet insert
comporte 4 ORFs, une incomplète et trois complètes (cf. figure 34). L'ORF incomplète
correspond à la partie 3' de la séquence codante de *orf29* ce qui a permis de compléter la
séquence de ce gène grâce à la séquence partielle de ce même orf obtenue lors du
séquençage de l'insert du plasmide pSPM58 (cf. exemple 19.2 et 19.3), l'ensemble des
15 deux séquences a ainsi permis d'obtenir la séquence complète de *orf29*. Les 4 gènes ont
ainsi été nommés *orf29* (SEQ ID N° 113), *orf30c* (SEQ ID N° 115), *orf31* (SEQ ID N°
118) et *orf32c* (SEQ ID N° 120). Le « c » ajouté dans le nom du gène signifiant pour
l'ORF en question que la séquence codante est dans l'orientation inverse (cf. figure 34).

Les séquences protéiques déduites de ces phases ouvertes de lecture (SEQ ID N°
20 114 pour *orf29*, SEQ ID N° 116 et 117 pour *orf30c*, SEQ ID N° 119 pour *orf31* et SEQ
ID N° 121 pour *orf32c*) ont été comparées avec celles présentes dans différentes bases
de données grâce à différents programmes : BLAST (Altschul *et al.*, 1990) (Altschul *et*
al., 1997), CD-search, (ces deux programmes sont accessibles notamment auprès du
National Center for Biotechnology Information (NCBI) (Bethesda, Maryland, USA)),
25 FASTA ((Pearson W. R & D. J. Lipman, 1988) et (Pearson W. R., 1990) (ce programme
est accessible notamment auprès du centre de ressources INFOBIOGEN, Evry, France).
Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de

ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycines.

19.6 Sous clonage et séquençage d'une troisième partie de l'insert du cosmide pSPM36.

- 5 Une sonde (fragment d'ADN de 0,8 kb) correspondant à une séquence interne à *orf32c* a été obtenue par PCR en utilisant comme matrice l'ADN total de la souche de *Streptomyces ambofaciens* et les amorces suivantes

KF36: 5'- TTGCCGTAGCCGAGGACCAGCG-3' (SEQ ID N° 151) et

KF37: 5'- CACATGGCCCTGGAGGACCCTG-3' (SEQ ID N° 152).

- 10 Le produit PCR ainsi obtenu représente une séquence interne de l'*orf32c*. Cette sonde à été utilisée dans des expériences de Southern Blot sur l'ADN chromosomique de la souche OSC2 et sur l'ADN du cosmide pSPM36 digérés par l'enzyme *Pst*I. En utilisant les mêmes conditions d'hybridation que celles décrites ci-dessus (cf. exemple 19.1), cette sonde révèle deux fragments *Pst*I d'environ 3,4kb et 2,5kb lorsque hybridée
- 15 sur l'ADN total de la souche OSC2 et sur l'ADN du cosmide pSPM36 digérés par l'enzyme *Pst*I. Compte tenu de la présence d'un site *Pst*I dans la sonde utilisée on peut expliquer ces résultats. Le premier fragment d'ADN qui a une taille d'environ 3,4kb est un fragment dont la séquence est déjà connue en totalité. La séquence du fragment de 2,5 kb n'est connue que partiellement, sur une région de 700 pb. Ce fragment a été isolé
- 20 à partir du cosmide pSPM36 dans le but de déterminer la séquence de la fin de l'*orf32c* et d'autres gènes en amont de ce dernier. Pour cela, le cosmide pSPM36 a été digéré par l'enzyme de restriction *Pst*I. Le fragment *Pst*I-*Pst*I d'une taille d'environ 2,5kb a été isolé par purification à partir d'un gel d'agarose 0.8% puis cloné dans le vecteur pBK-CMV (4518bp) (commercialisé par la société Stratagene (La Jolla, Californie, USA)). Le
- 25 plasmide ainsi obtenu a été nommé pSPM79 (cf. Figure 41) et la séquence de son insert a été déterminée.

La séquence de l'*orf28c* (SEQ ID N° 111) n'était pas complète (cf. exemple 19.3). En effet, une région d'environ 450 nucléotides n'avait pas pu être déterminée, ces 450 nucléotides figurent sous la forme d'une série de « N » dans la séquence SEQ ID N° 106. La séquence de la totalité de la région manquante a été déterminée par reséquençage de
5 cette région. La séquence des inserts de pSPM58 et pSPM73 a donc été déterminée en entier. La séquence complète de la partie codante de l'*orf28c* est présentée en SEQ ID N° 141 et la protéine déduite de cette séquence en SEQ ID N° 142. La séquence de l'insert du plasmide pSPM79 est présentée en SEQ ID N° 161.

Un assemblage des séquences des inserts de pSPM58, pSPM73 et pSPM79 est
10 présentée en SEQ ID N° 140 (cf. figure 41). Cette séquence part du site *Pst*I au niveau de l'*orf23c* (cf. SEQ ID N° 80) et va jusqu'au site *Pst*I au niveau de l'*orf34c* (cf. figure 41).

15 19.7 Analyse des nouvelles séquences nucléotidiques, détermination des phases
ouvertes de lecture et caractérisation des gènes pouvant être impliqués dans la
biosynthèse des spiramycines.

La séquence de l'insert du plasmide pSPM79 obtenue a été analysée grâce au programme FramePlot (Ishikawa J & Hotta K. 1999). Ceci a permis d'identifier, parmi les phases ouvertes de lecture, les phases ouvertes de lecture présentant un usage des codons typique de *Streptomyces*. Cette analyse a permis de déterminer que cet insert
20 comporte 3 ORFs, deux incomplètes (*orf32c* et *orf34c*) et une complète (*orf33*) (cf. figure 41).

La première ORF incomplète correspond à la partie 5' de la séquence codante de *orf32c*. Ceci a permis de compléter la séquence de ce gène grâce à la séquence partielle de cette même orf obtenue lors du séquençage de l'insert du plasmide pSPM73 (cf.
25 exemple 19.4 et 19.5), l'ensemble des deux séquences a ainsi permis d'obtenir la séquence complète de *orf32c*. (SEQ ID N° 145). Le « c » ajouté dans le nom du gène

signifie pour l'ORF en question que la séquence codante est dans l'orientation inverse (cf. figure 41). L'orf complète a été nommée *orf33* (SEQ ID N° 147). La troisième ORF a été nommée *orf34c* (SEQ ID N° 149). Le « c » ajouté dans le nom du gène signifie pour l'ORF en question que la séquence codante est dans l'orientation inverse (cf. figure 37). Les comparaisons effectuées entre le produit de cette orf et les banques de données suggèrent que la partie C-terminale de cette protéine n'est pas dans le produit déduit de la séquence nucléotidique et donc que cette orf serait plus longue et continuerait en dehors de la région séquencée.

Les séquences protéiques déduites de ces phases ouvertes de lecture ont été comparées avec celles présentes dans différentes bases de données grâce à différents programmes : BLAST (Altschul et al., 1990) (Altschul et al., 1997), CD-search, (ces deux programmes sont accessibles notamment auprès du National Center for Biotechnology Information (NCBI) (Bethesda, Maryland, USA)), FASTA ((Pearson W. R. & D. J. Lipman, 1988) et (Pearson W. R., 1990) (ce programme est accessible notamment auprès du centre de ressources INFOBIOGEN, Evry, France). Ces comparaisons ont permis de formuler des hypothèses sur la fonction des produits de ces gènes et d'identifier ceux susceptibles d'être impliqués dans la biosynthèse de spiramycines.

EXEMPLE 20 : Analyse de la production d'intermédiaires de biosynthèse de spiramycines :

20.1 Préparation des échantillons :

Les différentes souches à tester ont été cultivées chacune dans 7 erlenmeyers chicanés de 500 ml contenant 70 ml de milieu MP5 (Pernodet et al., 1993). Les erlens ont été inoculés par $2.5 \cdot 10^6$ spores/ml des différentes souches de *S. ambofaciens* et mises à pousser à 27°C sous agitation orbitale de 250 rpm pendant 96 heures. Les

cultures correspondant au même clone sont rassemblées, éventuellement filtrées sur filtre plissé, et centrifugées 15 min à 7000 tr/min.

Le pH du moût est ensuite ajusté à 9 avec de la soude et le surnageant est extrait par de la méthyl iso-butyryl cétone (MIBK). La phase organique (MIBK) est alors
5 récupérée et évaporée. L'extrait sec est ensuite repris dans 1ml d'acétonitrile, puis dilué au 1/10 (100µl qsp 1ml avec de l'eau) avant d'être utilisé pour les analyses en chromatographie liquide couplée à la spectrométrie de masse (CL/SM).

20.2 Analyse des échantillons en CL/SM :

10 Les échantillons ont été analysés en CL/SM dans le but de déterminer la masse des différents produits synthétisés par les souches à tester.

La colonne de chromatographie liquide haute performance utilisée est une colonne Kromasil C8 150*4,6mm, 5mm, 100Å.

La phase mobile est un gradient constitué d'un mélange d'acétonitrile et d'une solution
15 aqueuse d'acide trifluoroacétique à 0.05%, le débit est fixé à 1 ml/min. La température du four colonne est maintenue à 30°C.

La détection UV en sortie de la colonne a été effectuée à deux longueurs d'onde différentes : 238 nm et 280 nm.

Le spectromètre de masse couplé à la colonne de chromatographie est un appareil
20 Simple Quadripole commercialisé par la société Agilent, avec des tensions de cone à 30 et à 70V.

20.3 Analyse des intermédiaires de biosynthèse produits par la souche OS49.67 :

La souche OS49.67 dans laquelle le gène *orf3* est inactivé par une délétion en phase ne produit pas de spiramycines (cf. exemple 6 et 15).

25 Un échantillon a été préparé selon la méthode décrite ci-dessus (cf. paragraphe 20.1) et a été analysé par CL/SM comme décrit ci-dessus (cf. le paragraphe 20.2).

Plus particulièrement, l'analyse par chromatographie a été réalisée en gradient de solvant avec comme phase mobile : 20% d'acétonitrile du temps $T=0$ à 5min puis montée linéaire à 30% à $T=35$ minutes, suivie d'un palier jusqu'à $T=50$ minutes.

Dans ces conditions, deux produits ont plus particulièrement été observés : un
5 absorbant à 238 nm (temps de rétention de 33.4 min) et un absorbant à 280 nm (temps de rétention de 44.8 min) (cf. figure 35). La figure 35 montre la superposition des chromatogrammes CLHP réalisés à 238 et 280nm (haut) ainsi que les spectres UV des molécules élués à 33,4 minutes et 44,8 minutes (bas).

Les conditions d'analyses en spectrométrie de masse couplée étaient les suivantes
10 : le balayage se fait en mode scan, couvrant une gamme de masse comprise entre 100 et 1000 Da. Le gain de l'électro-multiplicateur était de 1V. Concernant la source électrospray, la pression du gaz de nébulisation était de 35 psig, le débit du gaz séchant de 12.0 l.min^{-1} , la température du gaz séchant de 350°C , la tension du capillaire était portée à $\pm 3000 \text{ V}$. Ces expériences ont permis de déterminer la masse des deux
15 produits séparés. Ces masses sont respectivement de 370 g/mole pour le produit élué en premier ($[\text{M}-\text{H}_2\text{O}]^+=353$ produit majoritaire) et 368 g/mole pour le second produit ($[\text{M}-\text{H}_2\text{O}]^+=351$ produit majoritaire).

Afin d'obtenir la structure, les produits mentionnés ci-dessus ont été isolés et purifiés dans les conditions suivantes : la phase mobile est un mélange d'une solution
20 aqueuse d'acide trifluoroacétique à 0.05% et d'acétonitrile 70/30 (v/v). La chromatographie est réalisée en régime isocratique avec un débit fixé de 1 ml/min. Dans ces conditions, les temps de rétention des 2 produits sont respectivement de 8 et 13.3 minutes. De plus dans ce cas, l'échantillon préparé (cf. paragraphe 20.1) n'est pas dilué dans l'eau avant injecton de $10\mu\text{l}$.

25 Les 2 produits sont récupérés à la sortie de la colonne chromatographique et isolés dans les conditions suivantes : une cartouche Oasis HLB 1 cc 30mg (Waters) est conditionnée séquentiellement par 1 ml d'acétonitrile, puis 1 ml d'eau/acétonitrile (20v/80v) et 1ml d'eau/acétonitrile 80/20. L'échantillon est ensuite introduit et la cartouche lavée successivement par 1 ml d'eau/acétonitrile (95/5), 1ml

d'eau/acétonitrile deutéré (95/5), puis éluée avec 600µl d'eau/acétonitrile deutéré 40/60. La solution récupérée est ensuite directement analysée en RMN.

Les spectres RMN obtenus pour ces deux composés sont les suivants :

5 - Premier produit élué: Platenolide A: (Spectre 9646V)

Spectre 1H dans CD3CN (déplacements chimiques en ppm): 0,90 (3H, t, J=6Hz), 0,93 (3H, d, J=5Hz), 1,27 (3H, d, J=5Hz), entre 1,27 et 1,40 (3H, m), 1,51 (1H, m), 1,95 (1H, m), 2,12 (1H, m), 2,30 (1H, d, J=12Hz), 2,50 (1H, d, J=11Hz), 2,58 (1H, dd, J=9 et 12 Hz), 2,96 (1H, d, J=7Hz), 3,43 (3H, s), 3,70 (1H, d, J=9Hz), 3,86 (1H, d, J=7Hz),
10 4,10 (1H, m), 5,08 (1H, m), 5,58 (1H, dt, J=3 et 12Hz), 5,70 (1H, dd, J=8 et 12 Hz), 6,05 (1H, dd, J=9 et 12 Hz), 6,24 (1H, dd, J=9 et 12Hz).

-Deuxième produit élué: Platenolide B: (Spectre 9647V)

Spectre 1H dans CD3CN (déplacements chimiques en ppm): 0,81 (3H, t, J=6Hz), 0,89 (1H, m), 1,17 (3H, d, J=5Hz), 1,30 (4H, m), 1,47 (2H, m), 1,61 (1H, t, J=10Hz), 2,20 (1H, m), 2,38 (1H, d, J=13Hz), 2,52 (1H, m), 2,58 (1H, m), 2,68 (1H, dd, J=8 et 13Hz),
15 3,10 (1H, d, J=7Hz), 3,50 (3H, s), 3,61 (1H, d, J=8Hz), 3,82 (1H, d, J=7Hz), 5,09 (1H, m), 6,20 (1H, m), 6,25 (1H, dd, J=9 et 12 Hz), 6,58 (1H, d, J=12 Hz), 7,19 (1H, dd, J=9 et 12Hz).

20

Ces expériences ont ainsi permis de déterminer la structure de ces deux composés. Le premier produit élué est le platénolide A et le deuxième est le platénolide B, la structure déduite de ces deux molécules est présentée en figure 36.

25

Il a également pu être déterminé grâce à la technique de CL/SM associée à la RMN, dans des conditions légèrement différentes de celles décrites précédemment mais dont la mise au point est bien connue de l'homme du métier, que la souche OS49.67 produit en plus du platénolide A et B un dérivé de ces deux composés. Il s'agit du platénolide A+mycaorse et du platénolide B+mycarose (la structure de ces deux

composés est présenté en figure 40). Les résultats de l'analyse du moût de la souche OS49.67 sont présentés tableau 42.

Tableau 42. Les résultats de l'analyse CL/SM du moût de la souche OS49.67

Identité	[(mg/L)	Masses des ions	Max d'absorption
Platénolide A Masse exacte=370 Formule Moléculaire=C₂₀H₃₄O₆	16,1	[M + Na] ⁺ 393,0 [M + K] ⁺ 408,9 [M - H ₂ O + H] ⁺ 353,0 [2M + Na] ⁺ 763,2	231nm
Platénolide B Masse exacte=368 Formule Moléculaire=C₂₀H₃₂O₆	1.4	[M - H ₂ O + H] ⁺ 351,0 [M + Na] ⁺ 391,0 [M + K] ⁺ 406,9 [2M + Na] ⁺ 759,1	283nm
Platénolide A + 'mycarose' Masse exacte=514 Formule Moléculaire=C₂₇H₄₆O₉	4,27	[M + Na] ⁺ 537,0 [M + K] ⁺ 553,0	230nm
Platénolide B + 'mycarose' Masse exacte=512 Formule Moléculaire=C₂₇H₄₄O₉	ND	[M + Na] ⁺ 535,0 [M + K] ⁺ 550,9 [PlatB - H ₂ O + H] ⁺ 350,9	284nm

20.4 Analyse des intermédiaires de biosynthèse produits par la souche SPM509 :

La souche SPM509 dans laquelle le gène *orf14* est inactivé (*orf14::att3 Ω aac-*) ne produit pas de spiramycines (cf. exemple 13, 15 et 16). Un échantillon a été préparé selon la méthode décrite ci-dessus (cf. paragraphe 20.1) et a été analysé par CL/SM
5 comme décrit ci-dessus (cf. le paragraphe 20.2 et 20.3). L'analyse des intermédiaires de biosynthèse présents dans le surnageant de culture de la souche SPM509 cultivée en milieu MP5 a montré que cette souche ne produisait que la forme B du platénolide (« platénolide B », cf. figure 36) mais pas la forme A (« platénolide A », cf. figure 36).

10 **EXEMPLE 21: Interruption du gène *orf14* dans une souche interrompue dans le gène *orf3* (OS49.67)**

Le produit du gène *orf14* est indispensable à la biosynthèse des spiramycines (cf. exemple 13, 15 et 16 : la souche SPM509 dans laquelle ce gène est interrompu ne produit plus de spiramycines). L'analyse des intermédiaires de biosynthèse présents dans
15 le surnageant de culture de la souche SPM509 cultivée en milieu MP5 a montré que cette souche ne produisait que la forme B du platénolide mais pas la forme A (cf. exemple 20). Une des hypothèses pouvant expliquer cette observation est que le produit de *orf14* soit impliqué dans la conversion de la forme B du platénolide en forme A par une étape enzymatique de réduction. Pour tester cette hypothèse, le gène *orf14* a été
20 inactivé dans un mutant ne produisant plus de spiramycine, mais produisant les formes A et B du platénolide. C'est le cas de la souche OS49.67 (cf. exemple 6 et 20) dans laquelle le gène *orf3* est inactivé par une délétion en phase (Δ *orf3*). Pour inactiver le gène *orf14* dans cette souche le plasmide pSPM509 a été introduit par transformation de protoplastes de la souche OS49.67 (Kieser, T *et al.*, 2000). L'inactivation du gène *orf14*
25 a déjà été décrite dans le cas de la souche OSC2 (cf. exemple 13) et il a été procédé de la même manière pour l'inactivation du gène *orf14* dans la souche OS49.67. Après transformation par le plasmide pSPM509, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués

respectivement sur milieu avec apramycine et sur milieu avec hygromycine. Les clones résistants à l'apramycine (ApraR) et sensibles à l'hygromycine (HygS) sont en principe ceux où un double événement de crossing over s'est produit et dans lesquels le gène *orf14* a été remplacé par une copie de *orf14* interrompue par la cassette *att3Ωaac-*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf14* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3Ωaac-* pour vérifier que le remplacement de gène avait bien eu lieu. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit PCR.

Un clone présentant les caractéristiques attendues ($\Delta orf3$, *orf14::att3Ωaac-*) a été plus particulièrement sélectionné et nommé SPM510. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *att3Ωaac-* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage *orf14* par la copie interrompue par la cassette *att3Ωaac-* dans le génome de ce clone.

20 **EXEMPLE 22 : Complémentation fonctionnelle de l'interruption du gène *orf14***

22.1 Construction du plasmide pSPM519 :

Le gène *orf14* a été amplifié par PCR en utilisant le couple d'oligonucléotides suivant : EDR31 : 5' **CCCAAGCTT**CTGCGCCCGCGGGCGTGAA 3' (SEQ ID N° 136) et EDR37 : 5' GCTCTAGAACCGTG TAGCCGCGCCCCGG 3' (SEQ ID N° 137) et comme matrice l'ADN chromosomique de la souche OSC2. Les oligonucléotides EDR31 et EDR37 portent respectivement le site de restriction *HindIII* et *XbaI* (séquence en gras). Le fragment de 1,2 kb ainsi obtenu a été cloné dans le vecteur pGEM-T easy

(commercialisé par la société Promega (Madison, Wisconsin, USA)) pour donner naissance au plasmide pSPM515. Ce plasmide a ensuite été digéré par les enzymes de restriction *Hind*III et *Xba*I. L'insert *Hind*III/*Xba*I de 1,2 kb obtenu a été cloné dans le vecteur pUWL201 (cf. exemple 17.1) préalablement digéré par les mêmes enzymes. Le
5 plasmide ainsi obtenu a été nommé pSPM519.

22.1 Transformation des souches SPM509 et SPM510 par le plasmide pSPM519 :

Le plasmide pSPM519 a été introduit dans les souches SPM509 (cf. exemple 13) et SPM510 (cf. exemple 17) par transformation de protoplastes (Kieser, T *et al.*, 2000).
10 Après transformation, les clones sont sélectionnés pour leur résistance au thiostrepton. Les clones sont ensuite repiqués sur un milieu contenant du thiostrepton.

La souche SPM509 est une souche non productrice de spiramycines (cf. exemple 13, 15 et 16 et figure 24). La production en spiramycines de la souche SPM509 transformée par le vecteur pSPM519 (souche nommée SPM509 pSPM519) a été
15 analysée en cultivant cette souche dans du milieu MP5 en présence de thiostrepton. Les surnageants de culture ont ensuite été analysés par CLHP (cf. exemple 16 et 17). Les résultats de cette analyse sont présentés dans le tableau 43, les données sont exprimées en mg par litre de surnageant. Les résultats correspondent à la production totale en spiramycines (obtenue en additionnant la production en spiramycine I, II et III). Il a été
20 observé que la présence du vecteur pSPM519 dans la souche SPM509 restaure la production de spiramycines (cf. tableau 43).

Tableau 43. Production en spiramycines de la souche SPM509 transformée par le vecteur pSPM519, (résultats exprimés en mg/l de surnageant).

Souche	Production en spiramycines
SPM509 pSPM519	58

La souche SPM510 transformée par le plasmide pSPM519 a été nommée SPM510 pSPM509.

EXEMPLE 23 : Complémentation fonctionnelle de l'interruption du gène *orf3* par le gène *tylB* de *S. fradiae*

23.1 Construction du plasmide pOS49.52 :

Le plasmide pOS49.52 correspond à un plasmide permettant l'expression de la protéine TylB chez *S. ambofaciens*. Pour le construire, la séquence codante du gène *tylB* de *S. fradiae* (Merson-Davies & Cundliffe, 1994, numéro d'accès GenBank : U08223 (séquence de la région), SFU08223 (séquence ADN) et AAA21342 (séquence protéique)) a été introduite dans le plasmide pKC1218 (Bierman *et al.*, 1992, Kieser *et al.*, 2000, une souche de *E. coli* contenant ce plasmide est accessible notamment auprès de l'ARS (NRRL) Agricultural Research Service Culture Collection) (Peoria, Illinois, USA), sous le numéro B-14790). De plus cette séquence codante a été placée sous le contrôle du promoteur *ermE** (cf. ci-dessus, notamment exemple 17.1, Bibb *et al.*, 1985, Bibb *et al.*, 1994).

23.1 Transformation de la souches OS49.67 par le plasmide pOS49.52 :

La souche OS49.67 dans laquelle le gène *orf3* est inactivé par une délétion en phase ne produit pas de spiramycines (cf. exemple 6 et 15). Le plasmide pOS49.52 a été introduit dans la souche OS49.67 par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones sont ensuite repiqués sur un milieu contenant de l'apramycine. Un clone a été plus particulièrement sélectionné et a été nommé OS49.67 pOS49.52.

Comme il a été démontré ci-dessus, la souche OS49.67 ne produit pas de spiramycines (cf. exemple 6 et 15). La production en spiramycines de la souche OS49.67 transformée par le vecteur pOS49.52 a été analysée par la technique décrite dans l'exemple 15. Il a ainsi pu être démontré que cette souche possède un phénotype

producteur de spiramycines. Ainsi la protéine TylB permet la complémentation fonctionnelle de l'interruption du gène *orf3*.

EXEMPLE 24 : Amélioration de la production en spiramycines par surexpression du gène *orf28c*

24.1 Construction du plasmide pSPM75 :

Le gène *orf28c* a été amplifié par PCR en utilisant un couple d'oligonucléotides comportant un site de restriction *HindIII* ou un site de restriction *BamHI*. Ces amorces ont la séquence suivante :

10 KF30 : 5' **AAGCTT**GTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138)
avec un site de restriction *HindIII* (qui figure en gras)

KF31 : 5' **GGATC**CCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139)
avec le site de restriction *BamHI* (qui figure en gras)

15 Les amorces KF30 et KF31 portent respectivement le site de restriction *HindIII* et *BamHI* (séquence en gras). Le couple d'amorces KF30 et KF31 permet d'amplifier un fragment d'ADN d'une taille d'environ 1,5 kb contenant le gène *orf28c* en utilisant comme matrice le cosmide pSPM36 (cf. ci-dessus). Le fragment de 1,5 kb ainsi obtenu a été cloné dans le vecteur pGEM-T easy (commercialisé par la société Promega (Madison, Wisconsin, USA)) pour donner naissance au plasmide pSPM74. Le plasmide
20 pSPM74 a ensuite été digéré par les enzymes de restriction *HindIII* et *BamHI* et l'insert *HindIII/BamHI* d'environ 1,5 kb obtenu a été sous cloné dans le vecteur pUWL201 (cf. exemple 17.1) préalablement digéré par les mêmes enzymes. Le plasmide ainsi obtenu a été nommé pSPM75, il contient l'ensemble de la séquence codante de *orf28c* placée sous le contrôle du promoteur *ermE**.

25 24.2 Transformation de la souche OSC2 par le plasmide pSPM75 :

Le plasmide pSPM75 a été introduit dans les souches OSC2 par transformation de protoplastes (Kieser, T *et al.*, 2000). Après transformation des protoplastes, les clones sont sélectionnés pour leur résistance au thiostrepton. Les clones sont ensuite

repiqués sur un milieu contenant du thiostrepton et la transformation par le plasmide est vérifiée par extraction de plasmides. Deux clones ont plus particulièrement été sélectionnés et nommés OSC2/pSPM75(1) et OSC2/pSPM75(2).

Un échantillon de la souche OSC2/pSPM75(2) a été déposé auprès de la
5 Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue
du Docteur Roux 75724 Paris Cedex 15, France, le 6 octobre 2003 sous le numéro
d'enregistrement I-3101.

Afin de tester l'effet de la surexpression du gène *orf28c* sur la production de
spiramycines, la production en spiramycines des clones OSC2/pSPM75(1) et
10 OSC2/pSPM75(2) a été testée par la technique décrite dans l'exemple 15. L'analyse de
la production en spiramycines de la souche OSC2 a également été effectuée en parallèle
pour comparaison. Il a ainsi pu être observé que la présence du plasmide pSPM75
augmente significativement la production en spiramycines de la souche OSC2. Ceci
démontre que la surexpression de *orf28c* conduit à une augmentation de la production
15 en spiramycines et confirme son rôle en tant que régulateur.

La production en spiramycines des clones OSC2/pSPM75(1) et
OSC2/pSPM75(2) a également été analysée par CLHP (de la même manière que dans
l'exemple 17.2). L'analyse de la production en spiramycines de la souche OSC2 a
également été effectuée en parallèle pour comparaison. Les résultats de cette analyse sont
20 présentés dans le tableau 44, les données sont exprimées en mg par litre de surnageant.
Les résultats correspondent à la production totale en spiramycines (obtenue en
additionnant la production en spiramycine I, II et III).

Tableau 44. Production en spiramycines des souches dérivées de OSC2
transformé par pSPM75, (résultats exprimés en mg/l).

Souche	Spiramycines
OSC2	50

OSC2/pSPM75(1)	120
OSC2/pSPM75(2)	155

Ainsi, il est observé que la présence du plasmide pSPM75 augmente significativement la production en spiramycines de la souche OSC2. Ceci démontre bien que la surexpression de *orf28c* a un effet positif sur la production en spiramycines.

5

EXEMPLE 25 : Analyse de la production d'intermédiaires de biosynthèse de spiramycines d'une souche inactivée dans le gène *orf8* :

La souche OS49.107, dans laquelle le gène *orf8* est inactivé par insertion de la cassette *Ωhyg*, ne produit pas de spiramycines (cf. exemple 7 et 15). Le gène *orf8* code
10 une protéine présentant une relative forte similitude avec plusieurs aminotransférases et suggère fortement que le gène *orf8* code une 4 amino-transférase responsable de la réaction de transamination nécessaire à la biosynthèse de la forosamine (cf. figure 6). On s'attend donc à ce que la biosynthèse des spiramycines soit bloquée au stade de la forocidine (cf. figure 7). La souche OS49.107 qui est non productrice de spiramycine
15 devrait donc produire de la forocidine.

Un échantillon de surnageant de la souche OS49.107 a été préparé selon la méthode décrite ci-dessus (cf. exemple 16, sans extraction au MIBK) et a été analysé par CL/SM comme décrit ci-dessus (cf. le paragraphe 20.2 et 20.3). En mode SIM, la masse 558 relative à l'ion moléculaire de la forocidine a été sélectionné et plusieurs pics ont
20 été détectés. La présence de composés de masse 558 est compatible avec l'hypothèse du rôle de *orf8* dans synthèse de forosamine.

EXEMPLE 26 : Analyse de la production d'intermédiaires de biosynthèse de spiramycines d'une souche inactivée dans le gène *orf12* :

La souche SPM507, dans laquelle le gène *orf12* est inactivé, ne produit pas de spiramycines (cf. exemple 11 et 15). Le gène *orf12* coderait une 3,4 déshydratase
5 responsable de la réaction de déshydratation nécessaire à la biosynthèse de la forosamine (cf. figure 6). On s'attend donc à ce que la biosynthèse des spiramycines soit bloquée au stade de la forocidine (cf. figure 7). La souche SPM507 qui est non productrice de spiramycine devrait donc produire de la forocidine.

Un échantillon de surnageant de la souche SPM507 a été préparé selon la méthode
10 décrite ci-dessus (cf. exemple 16, sans extraction au MIBK) et a été analysé par CL/SM comme décrit ci-dessus (cf. le paragraphe 20.2 et 20.3). Dans ces conditions, le temps de rétention de la forocidine est d'environ 12,9 minutes. En mode SIM, la masse 558 relative à l'ion moléculaire $[M+H]^+$ de la forocidine a été sélectionné et un pic a été détecté. La présence d'un composé à 558 permet de valider l'hypothèse du rôle du
15 produit de *orf12* dans la biosynthèse des spiramycines.

Toutefois, la forocidine est présente en quantité relativement faible et dans ces conditions, un produit absorbant à 238 nm a plus particulièrement été observé (temps de rétention de 17,1 min). L'analyse en LC/SM a permis de déterminer la masse de ce composé qui est de 744,3 g/mole ($[M+H]^+=744,3$ produit majoritaire).

20 Afin d'obtenir la structure, le produits mentionné ci-dessus a été isolé et purifié dans les conditions décrites précédemment (cf. paragraphe 20.1)). La phase organique (MIBK) est alors récupérée et évaporée. L'extrait sec est repris avec de l'eau et extrait avec de l'heptane. La solution aqueuse est ensuite extraite par fixation sur une cartouche Oasis HLB 1 g (Waters SAS, St-Quentin en-Yvelines, France). Le composé est récupéré
25 par élution avec un mélange eau/acétonitrile 30/70. Cette solution est ensuite injectée (100µL) sur la colonne analytique et les fractions récupérées sur cartouche Oasis HLB 1 cc 30mg (Waters). Avant utilisation, les cartouches Oasis HLB 1 cc 30mg (Waters) sont conditionnées séquentiellement par de l'acétonitrile, puis un mélange eau/acétonitrile (20v/80v) et un mélange d'eau/acétonitrile 80/20.

La cartouche Oasis HLB 1 cc 30mg (Waters) est ensuite lavée successivement par 1 ml d'eau/acétonitrile (95/5), 1ml d'eau/acétonitrile deutéré (95/5), puis éluée avec 600µl d'eau/acétonitrile deutéré 40/60. La solution récupérée est ensuite directement analysée en RMN.

5 Le spectre RMN obtenu pour ce composé est le suivant (Spectre RMN 19312V) :

Spectre ¹H dans CD₃CN/D₂O (déplacements chimiques en ppm): 0,92 (3H, d, J=6Hz), 1,10 (1H, m), 1,14 (3H, s), 1,17 (3H, d, J=6Hz), 1,22 (3H, d, J=6Hz), 1,25 (3H, d, J=6Hz), 1,40 (1H, m), 1,75 (1H, dd, J=12 et 2Hz), 1,81 (1H, m), 1,90 (1H, d, J=12Hz), 2,05 (1H, m), 2,12 (3H, s), 2,15 (1H, m), 2,35 (2H, m), 2,45 (6H, s large), 2,53 (1H, m), 10 2,64 (1H, dd, J=12 et 9Hz), 2,80 (1H, dd, J=9 et 16Hz), 2,95 (1H, d, J=8Hz), 3,23 (2H, m), 3,34 (1H, d, J=7Hz), 3,45 (3H, s), 3,49 (1H, m), 3,93 (1H, dd, J=7 et 3Hz), 4,08 (1H, m), 4,37 (1H, d, J=6Hz), 4,88 (1H, m), 5,05 (2H, m), 5,65 (2H, m), 6,08 (1H, dd, J=8 et 12Hz), 6,40 (1H, dd, J=12 et 9Hz), 9,60(1H, s).

15 Ces expériences ont ainsi permis de déterminer la structure de ce composé, celle-ci est présentée en figure 38.

EXEMPLE 27 : Analyse de la production d'intermédiaires de biosynthèse de spiramycines d'une souche inactivée dans le gène *orf5 :**

La souche SPM501 possède le génotype *orf6*::att1Ωhyg+*. Grâce à l'effet polaire de l'insertion de la cassette *att1Ωhyg+* dans le gène *orf6**, il a pu être déterminé que le gène *orf5** est indispensable à la voie de biosynthèse des spiramycines. En effet, l'insertion de la cassette excisable dans la partie codante du gène *orf6** entraîne un arrêt total de la production en spiramycines par effet polaire sur l'expression du gène *orf5**. Cependant, une fois que la cassette insérée a été excisée (et donc lorsque seul le gène *orf6** est inactivé, cf exemples 14 et 15), on restaure une production de spiramycine I. Ceci démontre que le gène *orf5** est indispensable à la biosynthèse des spiramycines puisque son inactivation entraîne un arrêt total de la production en spiramycines.

Le gène *orf5** code une protéine présentant une relative forte similitude avec plusieurs O-méthyltransférases. Le gène *orf5** serait une O-méthyl transférase impliquée dans la biosynthèse du platénolide. Pour vérifier cette hypothèse, des expériences d'analyse CL/SM et de RMN ont été conduites sur une souche de *S. ambofaciens* de
5 génotype *orf6*::att1 Ω hyg+* obtenue à partir d'une souche surproduisant les spiramycines.

Un échantillon du surnageant de cette souche a été préparé selon la méthode décrite ci-dessus (cf. exemple 16, sans extraction au MIBK) et a été analysé par CL/SM comme décrit ci-dessus (cf. le paragraphe 20.2 et 20.3). Toutefois, la colonne utilisée est
10 une colonne X-Terra (Waters SAS, St-Quentin en-Yvelines, France), et la tension de cone du spectromètre est réglée à 380V pour obtenir la fragmentation du composé analysé. Dans ces conditions, un produit dont le temps de rétention est d'environ 13,1 minutes est observé. Le spectre de masse de ce composé a une allure similaire à celui de la spiramycine I mais l'ion moléculaire est à 829. La différence de masse de 14 par
15 rapport à la masse de la spiramycine peut s'expliquer par l'absence de groupement méthyl sur l'oxygène portée par le carbone n°4 du cycle lactone (la structure de ce composé est présenté en figure 39). La présence d'un composé à 829 permet de valider l'hypothèse du rôle de *orf5** dans la biosynthèse des spiramycines.

Par un test microbiologique effectué sur une souche sensible de *M. luteus* (cf. exemple
20 15 et figure 18) il a été démontré que la molécule intermédiaire (spiramycine moins groupement méthyle dont la structure est présentée en figure 39) produite par la souche *orf6*:: Ω atthyg+* est beaucoup moins active (d'un facteur 10) que la spiramycine d'origine avec le groupement méthyl en position 4.

25 **EXEMPLE 28 : Construction de nouvelles « cassettes excisables » :**

De nouvelles cassettes excisables ont été construites. Ces cassettes sont très semblables aux cassettes excisables déjà décrites dans l'exemple 9. La différence principale entre les anciennes et les nouvelles cassettes est l'absence dans ces dernières

des séquences correspondant aux extrémités de l'interposon Ω , séquences qui contiennent un terminateur de transcription provenant du phage T4.

Dans les cassettes sans terminateur, le gène qui confère la résistance à un antibiotique est flanqué des séquences *attR* et *attL* permettant l'excision. Le gène de
5 résistance est le gène *aac(3)IV* qui code une acétyltransferase qui confère la résistance à l'apramycine. Ce gène est présent dans la cassette Ωaac (numéro d'accès GenBank : X99313, Blondelet-Rouault, M.H. *et al.*, 1997) et a été amplifié par PCR en utilisant comme matrice le plasmide pOSK1102 (cf. ci-dessus) et comme amorces les
oligonucléotides KF42 et KF43 contenant chacun le site de restriction *HindIII* (en
10 gras)(AAGCTT) en 5'.

KF42: 5'-AAGCTTGTACGGCCACAGAATGATGTCAC-3' (SEQ ID N° 153)
et

KF43: 5'-AAGCTTCGACTACCTTGGTGATCTCGCCTT-3' (SEQ ID N° 154).

Le produit PCR obtenu d'environ 1kb a été cloné dans le vecteur *E. coli* pGEMT
15 Easy donnant naissance au plasmide pSPM83.

Le vecteur pSPM83 a été digéré par l'enzyme de restriction *HindIII*. Le fragment
HindIII-HindIII de l'insert a été isolé par purification à partir d'un gel d'agarose 0,8%
puis cloné dans le site *HindIII* situé entre les séquences *attL* et *attR* des différents
plasmides portant les différentes cassettes excisables possibles (cf. exemple 9 et figure
20 27) de façon à remplacer le fragment *HindIII* correspondant à Ωacc par le fragment
HindIII correspondant au gène *aac* seul. Ceci a permis d'obtenir les cassettes *att1aac*,
att2aac et *att3aac* (selon la phase souhaitée, cf. exemple 9). Selon l'orientation du gène
aac par rapport aux séquences *attL* et *attR*, on distingue *att1aac+*, *att1aac-*, *att2aac+*,
att2aac-, *att3aac+* et *att3aac-* (selon les mêmes conventions que celles adoptées dans
25 l'exemple 9).

EXEMPLE 29 : Construction d'une souche de *S. ambfaciens* interrompue dans le gène *orf28c* :

L'inactivation du gène *orf28c* a été réalisée grâce à la technique des cassettes excisables. La cassette excisable *att3aac+* (cf. exemple 28) a été amplifiée par PCR en utilisant comme matrice le plasmide pSPM101 (Le plasmide pSPM101 est un plasmide dérivé du vecteur pGP704Not (Chaveroche *et al.*, 2000) (Miller VL & Mekalanos JJ, 1988) dans lequel la cassette *att3aac+* a été clonée comme un fragment *EcoRV* dans le site unique *EcoRV* de pGP704Not) et à l'aide des amorces suivantes :

10 **KF32:**

5' CAACCGCTTGAGCTGCTCCATCAACTGCTGGGCCGAGGTATCGCGCGCG
CTTCGTTTCGGGACGAA 3' (SEQ ID N° 155)

et **KF33:**

5' TGGGTCCCGCCGCGCGGCACGACTTCGACTCGCTCGTCTATCTGCCTCTT
15 CGTCCCGAAGCAACT 3' (SEQ ID N° 156)

Les 39 nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène *orf28c* et les 26 nucléotides situés le plus en 3' (figuré en gras et soulignés ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3aac+*

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli* hyper-recombinante DY330 (Yu *et al.*, 2000) (cette souche contient les gènes *exo*, *bet* et *gam* du phage lambda intégrés dans son chromosome, ces gènes sont exprimés à 42°C, elle a été utilisée à la place de la souche *E. coli* KS272 (Chaveroche *et al.*, 2000)) contenant le cosmide pSPM36. Ainsi, les bactéries ont été transformées par électroporation par ce produit de PCR et les clones ont été sélectionnés pour leur résistance à l'apramycine. Les cosmides des clones obtenus ont été extraits et digérés par l'enzyme de restriction *BamHI*, dans le but de vérifier que le profil de digestion obtenu correspondait au profil attendu s'il y a eu insertion de la cassette (*att3aac+*) dans le gène *orf28c*, c'est à dire s'il y avait bien eu recombinaison homologue entre les extrémités du produit PCR et le gène

cible. La vérification de la construction peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un clone dont le cosmide possède le profil attendu a été sélectionné et le cosmide correspondant a été nommé pSPM107. Ce
5 cosmide est un dérivé de pSPM36 dans lequel *orf28c* est interrompue par la cassette *att3aac+*. L'insertion de la cassette s'accompagne d'une délétion dans le gène *orf28c*, l'interruption commence au niveau du 28ème codon de *orf28c*. Il reste après la cassette les 137 derniers codons de *orf28c*.

10 Le cosmide pSPM107 a dans un premier temps été introduit dans la souche *E. coli* DH5α puis dans la souche *Streptomyces ambofaciens* OSC2 par transformation de protoplastes. Après transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement
15 A) (cf. figure 9). Les clones résistants à l'apramycine (ApraR) et sensibles à la puromycine (PuroS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf28c* interrompu par la cassette *att3aac+*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf28c* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi,
20 l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3aac+* pour vérifier la présence de la cassette au locus attendu dans l'ADN génomique des clones obtenus. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les
25 oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf28c::att3aac+*) a été plus particulièrement sélectionné et nommé SPM107. Ce clone possède donc le génotype : *orf28c::att3aac+* et a été nommé SPM107. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf28c*, au vu de l'orientation des
30 gènes (cf. figure 3). En effet, le fait que *orf29* soit orienté en sens opposé à *orf28c*

montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment, notamment par transformation par le plasmide pOSV508.

Afin de tester l'effet de l'extinction du gène *orf28c* sur la production de spiramycines, la production en spiramycines de la souche SPM107 a été testée par la technique décrite dans l'exemple 15. Il a ainsi pu être démontré que cette souche possède un phénotype non producteur de spiramycines. Ceci démontre que le gène *orf28c* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

EXEMPLE 30 : Construction d'une souche de *S. ambofaciens* interrompue dans le gène *orf31* :

L'inactivation du gène *orf31* a été réalisée grâce à la technique des cassettes excisables. La cassette excisable *att3aac+* a été amplifiée par PCR en utilisant comme matrice le plasmide pSPM101 et les oligonucléotides EDR71 et EDR72.

EDR71:

5' CGTCATCGACGTGCGGGGAAGACAGAGGTGATACCGATGATCGCGCGC
GCTTCGTTCTCGGGACGAA 3' (SEQ ID N° 157)

EDR72:

5' GCCAGCACCTCGTCCAGCTGCTCGACGGAACCTACCCCCATCTGCCTCT
TCGTCCCGAAGCAACT 3' (SEQ ID N° 158)

Les 39 nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une séquence correspondant à une séquence dans le gène *orf31* et les 26 nucléotides situés le plus en 3' (figurés en gras et soulignés ci-dessus) correspondent à la séquence d'une des extrémités de la cassette excisable *att3aac+*.

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli* KS272 contenant le plasmide pKOBEG et le cosmide pSPM36, comme décrit par Chaveroche et al. (Chaveroche et al., 2000) (cf. figure 12 pour le principe, le plasmide

pOS49.99 doit être remplacé par le cosmide pSPM36 et le plasmide obtenu n'est plus pSPM17 mais pSPM543). Ainsi, les bactéries ont été transformées par électroporation par ce produit de PCR et les clones ont été sélectionnés pour leur résistance à l'apramycine. Les cosmides des clones obtenus ont été extraits et digérés par plusieurs enzymes de restriction, dans le but de vérifier que le profil de digestion obtenu correspond au profil attendu s'il y a eu insertion de la cassette (*att3aac+*) dans le gène *orf31*, c'est à dire s'il y a bien eu recombinaison homologue entre les extrémités du produit PCR et le gène cible. La vérification de la construction peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR. Un clone dont le cosmide possède le profil attendu a été sélectionné et le cosmide correspondant a été nommé pSPM543. Ce cosmide est un dérivé de pSPM36 dans lequel *orf31* est interrompue par la cassette *att3aac+* (cf. figure 12). L'insertion de la cassette s'accompagne d'une délétion dans le gène *orf31*, l'interruption commence au niveau du trente sixième codon de *orf31*. Il reste après la cassette les 33 derniers codons de *orf31*.

Le cosmide pSPM543 a été introduit dans la souche *Streptomyces ambifaciens* OSC2 (cf. ci-dessus) par transformation de protoplastes (Kieser, T et al., 2000). Après transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec apramycine (antibiotique B) et sur milieu avec puromycine (antibiotique A) (cf. figure 9). Les clones résistants à l'apramycine (ApraR) et sensibles à la puromycine (PuroS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf31* interrompu par la cassette *att3aac+*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf31* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3aac+* pour vérifier la présence de la cassette au locus attendu dans l'ADN génomique des clones obtenus. Une deuxième hybridation a été effectuée en utilisant comme sonde un

fragment d'ADN obtenu par PCR et correspondant à une très large partie de la séquence codante du gène *orf31*.

La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf31::att3aac+*) a été plus particulièrement sélectionné et nommé SPM543. Il a pu en effet être vérifié grâce aux deux hybridations que la cassette *att3aac+* était bien présente dans le génome de ce clone et que l'on obtient bien le profil de digestion attendu dans le cas d'un remplacement, à la suite d'un double événement de recombinaison, du gène sauvage par la copie interrompue par la cassette *att3aac+* dans le génome de ce clone. Ce clone possède donc le génotype : *orf31::att3aac+* et a été nommé SPM543. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf31*, au vu de l'orientation des gènes (cf. figure 3). En effet, le fait que *orf32c* soit orienté en sens opposé à *orf31* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment, notamment par transformation par le plasmide pOSV508.

Afin de tester l'effet de l'extinction du gène *orf31* sur la production de spiramycines, la production en spiramycines de la souche SPM543 a été testée par la technique décrite dans l'exemple 15. Il a ainsi pu être démontré que cette souche possède un phénotype non producteur de spiramycines. Ceci démontre que le gène *orf31* est un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

EXEMPLE 31 Construction d'une souche de *S. ambofaciens* interrompue dans le gène *orf32c* :

L'inactivation du gène *orf32c* a été réalisée grâce à la technique des cassettes excisables. La cassette excisable *att3aac+* a été amplifiée par PCR en utilisant comme matrice le plasmide pSPM101 et à l'aide des amorces suivantes :

KF52:

5' GATCCGCCAGCCTCACGTCACGCCGCGCCGCCTCCCTGACATCGCGCGC
GCTTCGTTCGGGACGAA 3' (SEQ ID N° 159).

et KF53:

5 5' GAGGCGGACGTCGGTACGCGGTGGGAGCCGGAGTTCGACAATCTGCCTC
TTCGTCCCGAAGCAACT 3' (SEQ ID N° 160).

Les 40 nucléotides situés à l'extrémité 5' de ces oligonucléotides comportent une
séquence correspondant à une séquence dans le gène *orf32c* et les 26 nucléotides situés
10 le plus en 3' (figuré en gras et soulignés ci-dessus) correspondent à la séquence d'une
des extrémités de la cassette excisable *att3aac+*.

Le produit de PCR ainsi obtenu a été utilisé pour transformer la souche *E. coli*
hyper-recombinante DY330 (Yu *et al.*, 2000) contenant le cosmide pSPM36. Ainsi, les
bactéries ont été transformées par électroporation par ce produit de PCR et les clones
15 ont été sélectionnés pour leur résistance à l'apramycine. Les cosmides des clones
obtenus ont été extraits et digérés par l'enzymes de restriction *Bam*HI, dans le but de
vérifier que le profil de digestion obtenu correspond au profil attendu s'il y a eu
insertion de la cassette (*att3aac+*) dans le gène *orf32c*, c'est à dire s'il y a bien eu
recombinaison homologue entre les extrémités du produit PCR et le gène cible. La
20 vérification de la construction peut également être réalisée par toute méthode connue de
l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés
et séquençage du produit de PCR. Un clone dont le cosmide possède le profil attendu a
été sélectionné et le cosmide correspondant a été nommé pSPM106. Ce cosmide est un
dérivé de pSPM36 dans lequel *orf32c* est interrompue par la cassette *att3aac+*.
25 L'insertion de la cassette s'accompagne d'une délétion dans le gène *orf32c*, l'interruption
commence au niveau du 112ème codon de *orf32c*. Il reste après la cassette les 91
derniers codons de *orf32c*.

Le cosmide pSPM106 a dans un premier temps été introduit dans la souche *E. coli*
30 DH5α puis, dans la souche *Streptomyces ambofaciens* OSC2 par transformation. Après

transformation, les clones sont sélectionnés pour leur résistance à l'apramycine. Les clones résistants à l'apramycine sont ensuite repiqués respectivement sur milieu avec apramycine (antibiotique B) et sur milieu avec puromycine (antibiotique A) (cf. figure 9). Les clones résistants à l'apramycine (ApraR) et sensibles à la puromycine (PuroS) sont en principe ceux où un double événement de crossing over s'est produit et qui possèdent le gène *orf32c* interrompu par la cassette *att3aac+*. Ces clones ont été plus particulièrement sélectionnés et le remplacement de la copie sauvage de *orf32c* par la copie interrompue par la cassette a été vérifié par hybridation. Ainsi, l'ADN total des clones obtenus, a été digéré par plusieurs enzymes, séparé sur gel d'agarose, transféré sur membrane et hybridé avec une sonde correspondant à la cassette *att3aac+* pour vérifier la présence de la cassette dans l'ADN génomique des clones obtenus. La vérification du génotype peut également être réalisée par toute méthode connue de l'homme du métier et notamment par PCR en utilisant les oligonucléotides appropriés et séquençage du produit de PCR.

Un clone présentant les caractéristiques attendues (*orf32c::att3aac+*) a été plus particulièrement sélectionné. Ce clone possède donc le génotype : *orf32c::att3aac+* et a été nommé SPM106. Il est inutile de procéder à l'excision de la cassette pour l'étude de l'effet de l'inactivation de *orf32c*, au vu de l'orientation des gènes (cf. figure 3). En effet, le fait que *orf33* soit orienté en sens opposé à *orf32c* montre que ces gènes ne sont pas co-transcrits. L'utilisation d'une cassette excisable permet en revanche d'avoir la possibilité de se débarrasser du marqueur de sélection à tout moment, notamment par transformation par le plasmide pOSV508.

Afin de tester l'effet de l'extinction du gène *orf32c* sur la production de spiramycines, la production en spiramycines de la souche SPM106 a été testée par la technique décrite dans l'exemple 15. Il a ainsi pu être démontré que cette souche possède un phénotype producteur de spiramycines. Ceci démontre que le gène *orf32c* n'est pas un gène essentiel à la biosynthèse de la spiramycine chez *S. ambofaciens*.

Liste des constructions décrites dans la présente demande

Liste des abréviations : Am : Ampicilline ; Hyg: Hygromycine ; Sp : Spiramycine ; Ts : Thiostrepton ; Cm : Chloramphénicol. Kn : Kanamycine, Apra : apramycine.

Nom de la construction	Marqueur de sélection	Principales caractéristiques	Référence
pWE15	Am		(Wahl, <i>et al.</i> , 1987)
pWED1	Am	pWE15 dans lequel un fragment <i>HpaI-HpaI</i> de 4.1 kb a été délété	(Gourmelen <i>et al.</i> , 1998)
pOJ260	Apra	Conjugatif, non répliatif chez <i>Streptomyces</i>	(Bierman <i>et al.</i> , 1992)
pHP45 Ω hyg	Hyg	Cassette Ω hyg dans pHP45.	(Blondelet-Rouault <i>et al.</i> , 1997)
pKC505	Apra	Cosmide	(Richardson MA <i>et al.</i> , 1987)
pIJ486	Ts	Plasmide répliatif multicopies <i>Streptomyces</i>	(Ward <i>et al.</i> , 1986)
pOSint3	Am	<i>ptrc-xis-int</i> dans pTrc99A	(Raynal <i>et al.</i> , 1998)
pWHM3	Am, Ts	Vecteur navette répliatif <i>E. coli/Streptomyces</i> .	(Vara <i>et al.</i> , 1989)
pKOBEG	Cm		(Chaveroche <i>et al.</i> , 2000)
pGP704Not	Am		(Chaveroche <i>et al.</i> , 2000)
pMBL18	Am		(Nakano <i>et al.</i> , 1995)
pGEM-T Easy	Am	vecteur <i>E. coli</i> pour clonage de produits PCR	Mezei <i>et al.</i> , 1994

pOS49.1	Am	pWED1 avec insert au niveau du site <i>Bam</i> HI	Exemple 2
pOS49.11	Am	Fragment <i>Sac</i> I de pOS49.1 dans pUC19.	Exemple 2
pOSC49.12	Ch	Fragment <i>Xho</i> I de pOS49.11 dans pBC SK ⁺ .	Exemple 2
pOS49.14	Cm , Hyg	pOSC49.12 avec le gène <i>orf3</i> interrompu par la cassette <i>Ω</i> <i>hyg</i>	Exemple 2
pOS49.16	Apra, Hyg	Insert de pOS49.14 dans pOJ260	Exemple 2
pOS49.28	Cm	Fragment <i>Bam</i> HI- <i>Pst</i> I de 3,7kb de pOS49.1 dans pBC SK ⁺	Exemple 3
pOS44.1	Apra, Sp	pKC505 contenant un insert conférant la résistance la spiramycine chez <i>S. griseofuscus</i>	(Pernodet <i>et al.</i> , 1999)
pOS44.2	Ts, Sp	Fragment <i>Sau</i> 3AI de 1,8 kb pOS44.1 dans pIJ486	Exemple 3
pOS44.4.	Am	Insert de pOS44.2 dans pUC19	Exemple 3
pSPM5	Am	pWED1 avec insert d'ADN de <i>S. ambofaciens</i> au niveau du site <i>Bam</i> HI	Exemple 3
pSPM7	Am	pWED1 avec insert d'ADN de <i>S. ambofaciens</i> au niveau du site <i>Bam</i> HI	Exemple 3

pOSK1205	Hyg	pBK-CMV dans lequel <i>hyg</i> remplace <i>neo</i>	Exemple 5
pOS49.67	Apra	Fragment <i>EcoRI-SacI</i> de pOS49.1, comportant une délétion interne de 504 nucléotides, dans pOJ260	Exemple 6
pOS49.88	Am	Fragment de 3.7 kb <i>PstI-EcoRI</i> de pOS49.1 dans pUC19	Exemple 7
pOS49.106	Am	pO49.88 avec <i>hyg</i> dans <i>orf8</i> (<i>hyg</i> et <i>orf8</i> dans la même orientation)	Exemple 7
pOS49.120	Am	pOS49.88 avec <i>hyg</i> dans <i>orf8</i> (<i>hyg</i> et <i>orf8</i> dans des orientations opposées)	Exemple 7
pOS49.107	Apra, Hyg	Insert de pOS49.106 dans pOJ260	Exemple 7
pOS49.32	Am, Kn	Fragment de 1,5 kb interne à <i>orf10</i> dans pCR2.1-TOPO	Exemple 8
pOS49.43	Am, Kn	pOS49.32 avec <i>hyg</i> dans <i>orf10</i> (<i>hyg</i> et <i>orf10</i> dans la même orientation)	Exemple 8
pOS49.44	Am, Kn	pOS49.32 avec <i>hyg</i> dans <i>orf10</i> (<i>hyg</i> et <i>orf10</i> dans des orientations opposées)	Exemple 8
pOS49.50	Apra, Hyg	Insert de pOS49.43 dans pOJ260	Exemple 8
pWHM3Hyg	Am, Hyg	pWHM3 dans lequel <i>tsr</i> est remplacé par <i>hyg</i>	Exemple 10

pOSV508	Am, Ts	<i>ptrc-xis-int</i> dans pWHM3	Exemple 9
patt1 Ω hyg+	Cm, Hyg	Cassette att1 Ω hyg+ dans pBC SK+ dont le site <i>HindIII</i> a été supprimé	Exemple 9
patt3 Ω aac-	Cm, Gn	Cassette att3 Ω aac- dans pBC SK+ dont le site <i>HindIII</i> a été supprimé	Exemple 9
pOSV510	Am, Hyg	pro <i>pra-Amh</i> dans pWHM3Hyg	Exemple 10
pOS49.99	Am	Fragment <i>EcoRI-BamHI</i> de 4,5 kb de pSPM5 dans pUC19	Exemple 10
pOSK1102	Am, Apra	pGP704Not contenant la cassette att3 Ω aac-	Exemple 10
pSPM17	Am, Apra	pOS49.99 dans lequel <i>orf2</i> est interrompue par la cassette att3 Ω aac-	Exemple 10
pSPM21	Hyg, Apra	pOSK1205 contenant l'insert <i>EcoRI-XbaI</i> de pSPM17 (dans lequel <i>orf2</i> est interrompue par la cassette att3 Ω aac-)	Exemple 10
pSPM502	Am	Fragment <i>BglII</i> de 15,1 kb de pSPM7 dans pMBL18	Exemple 11
pSPM504	Hyg	Insert de pSPM502 dans pOSK1205	Exemple 11
pSPM507	Hyg, Apra	pSPM504 dans lequel <i>orf12</i> est interrompue par la cassette att3 Ω aac-	Exemple 11
pSPM508	Hyg, Apra	pSPM504 dans lequel <i>orf13c</i> est interrompue par la cassette att3 Ω aac-	Exemple 12

pSPM509	Hyg, Apra	pSPM504 dans lequel <i>orf14</i> est interrompue par la cassette <i>att3Ωaac-</i>	Exemple 13
pBXL1111	Am	Fragment de 1,11 kb contenant <i>orf6*</i> amplifié par PCR à partir de pSPM7, dans le vecteur pGEM-T Easy	Exemple 14
pBXL1112	Am, Hyg	pBXL1111 dans lequel la cassette <i>att1Ωhyg+</i> a été introduite après délétion de 120 pb dans la séquence codante du gène <i>orf6*</i>	Exemple 14
pBXL1113	Apra, Hyg	Insert <i>Pst</i> I de 3.7 kb de pBXL1112 dans pOJ260	Exemple 14
pSPM520	Am	Fragment de PCR amplifié par les oligonucléotides EDR39-EDR42 dans pGEM-T Easy	Exemple 17
pSPM521	Am	Fragment de PCR amplifié par les oligonucléotides EDR40-EDR42 dans pGEM-T Easy	Exemple 17
pSPM522	Am	Fragment de PCR amplifié par les oligonucléotides EDR41-EDR42 dans pGEM-T Easy	Exemple 17
pUWL201	Am, Ts		(Doumith <i>et al.</i> , 2000)
pSPM523	Am, Ts	Fragment <i>Hind</i> III- <i>Bam</i> HI de l'insert du plasmide pSPM520 dans le vecteur pUWL201	Exemple 17

pSPM524	Am, Ts	Fragment <i>Hind</i> III- <i>Bam</i> HI de l'insert du plasmide pSPM521 dans le vecteur pUWL201	Exemple 17
pSPM525	Am, Ts	Fragment <i>Hind</i> III- <i>Bam</i> HI de l'insert du plasmide pSPM522 dans le vecteur pUWL201	Exemple 17
pSPM527	Am	pSPM521 avec décalage du cadre de lecture au niveau du site <i>Xho</i> I	Exemple 17
pSPM528	Am, Ts	Fragment <i>Hind</i> III- <i>Bam</i> HI de l'insert du plasmide pSPM527 dans le vecteur pUWL201	Exemple 17
pVF 10.4			(Vara <i>et al.</i> , 1985; Lacalle <i>et al.</i> , 1989)
pPM803	Ts		(Mazodier, P. <i>et al.</i> , 1989)
pGEM-T- <i>pac-oriT</i>	Am	Cassette <i>pac-oriT</i> (amplifiée par PCR à partir de pVF 10.4 et pPM803) dans dans pGEM-T Easy	Exemple 18
pWED2	Am	Cassette <i>pac-oriT</i> obtenue à partir de pGEM-T- <i>pac-oriT</i> insérée dans pWED1	Exemple 18
pSPM34	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM35	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM36	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19

pSPM37	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM38	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM39	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM40	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM41	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM42	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM43	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM44	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM45	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM47	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM48	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM50	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM51	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM52	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19

pSPM53	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM55	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM56	Am	pWED2 avec insert au niveau du site <i>Bam</i> HI	Exemple 19
pSPM58	Kn	Fragment <i>Pst</i> I- <i>Pst</i> I d'environ 6kb de l'insert de pSPM36 dans pBK-CMV	Exemple 19
pSPM72	Kn	Fragment <i>Stu</i> I- <i>Stu</i> I d'environ 10 kb de l'insert de pSPM36 cloné dans pBK-CMV	Exemple 19
pSPM73	Cm	Fragment <i>Eco</i> RI- <i>Hind</i> III de l'insert de pSPM72 dans pBC-SK+	Exemple 19
pSPM515	Am	Fragment de PCR amplifié par EDR31-EDR37 dans pGEM-T easy	Exemple 22
pSPM519	Am, Ts	Insert <i>Hind</i> III/ <i>Xba</i> I de pSPM515 dans pUWL201	Exemple 22
pOS49.52	Apra	Séquence codante de <i>tylB</i> sous contrôle du promoteur <i>ermE</i> * dans le plasmide pKC1218	Exemple 23
pSPM74	Am	Fragment de PCR amplifié par KF30-KF31 dans pGEM-T easy	Exemple 24
pSPM75	Am, Ts	Insert <i>Hind</i> III/ <i>Bam</i> HI de pSPM74 dans pUWL201	Exemple 24

pSPM79	Kn	Fragment <i>PstI-PstI</i> d'environ 2,5kb de l'insert de pSPM36 dans pBK-CMV	Exemple 19
pSPM83	Am	Fragment de PCR amplifié par KF42-KF43 dans pGEM-T easy	Exemple 28
pSPM107	Am, Apra	pSPM36 dans lequel <i>orf28c</i> est interrompue par la cassette <i>att3aac+</i>	Exemple 29
pSPM543	Am, Apra	pSPM36 dans lequel <i>orf31</i> est interrompue par la cassette <i>att3aac+</i>	Exemple 30
pSPM106	Am, Apra	pSPM36 dans lequel <i>orf32c</i> est interrompue par la cassette <i>att3aac+</i>	Exemple 31

Dépôt de matériel biologique

Les organismes suivants ont été déposés le 10 juillet 2002 auprès de la Collection Nationale de Cultures de Microorganismes (CNCM), 25 rue du Docteur Roux, 75724 Paris Cedex 15, France, selon les dispositions du Traité de Budapest.

- 5 - Souche OSC2 sous le numéro d'enregistrement I-2908.
- Souche SPM501 sous le numéro d'enregistrement I-2909.
- Souche SPM502 sous le numéro d'enregistrement I-2910.
- Souche SPM507 sous le numéro d'enregistrement I-2911.
- Souche SPM508 sous le numéro d'enregistrement I-2912.
- 10 - Souche SPM509 sous le numéro d'enregistrement I-2913.
- Souche SPM21 sous le numéro d'enregistrement I-2914.
- Souche SPM22 sous le numéro d'enregistrement I-2915.
- Souche OS49.67 sous le numéro d'enregistrement I-2916.
- Souche OS49.107 sous le numéro d'enregistrement I-2917.
- 15 - Souche *Escherichia coli* DH5 α contenant le plasmide pOS44.4 sous le numéro d'enregistrement I-2918.

Les organismes suivants ont été déposés le 26 février 2003 auprès de la Collection Nationale de Cultures de Microorganismes (CNCM), 25 rue du Docteur Roux, 75724 Paris Cedex 15, France, selon les dispositions du Traité de Budapest.

- 20 - Souche SPM502 pSPM525 sous le numéro d'enregistrement I-2977.

Les organismes suivants ont été déposés auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 6 octobre 2003 selon les dispositions du Traité de Budapest .

- Souche OSC2/pSPM75(2) sous le numéro d'enregistrement I-3101.

Toutes les publications et brevets cités sont incorporés à la présente demande par référence.

Bibliographie :

- 5 - Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990). **215** (3):403-410.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997). **25** (17):3389-402.
- 10 - Amann,E., Ochs,B. and Abel,K.J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli* *Gene* (1988) **69** (2), 301-315.
- Arisawa,A., Kawamura,N., Tsunekawa,H., Okamura,K., Tone,H. and Okamoto,R. Cloning and nucleotide sequences of two genes involved in the 4"-O-acylation of macrolide antibiotics from *Streptomyces thermotolerans*. *Biosci. Biotechnol. Biochem.* (1993). **57** (12) : 2020-2025.
- 15 - Arisawa A, Kawamura N, Takeda K, Tsunekawa H, Okamura K, Okamoto R. Cloning of the macrolide antibiotic biosynthesis gene *acyA*, which encodes 3-O-acyltransferase, from *Streptomyces thermotolerans* and its use for direct fermentative production of a hybrid macrolide antibiotic. *Appl Environ Microbiol.* (1994). **60** (7):2657-2660.
- 20 - August,P.R., Tang,L., Yoon,Y.J., Ning,Streptomyces, Mueller,R., Yu,T.W., Taylor,M., Hoffmann,D., Kim,C.G., Zhang,X., Hutchinson,C.R. et Floss,H.G. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amiclatopsis mediterranei* S699. *Chem. Biol.* (1998) **5** (2), 69-79.
- 25 - Ausubel Fred M., Brent Roger, Kingston Robert E., Moore David D., Seidman J.G., Smith John A., Struhl Kevin (Editeurs). Current Protocols in Molecular Biology, publié par John Wiley & Sons Inc. Current Protocols Customer Service, 605 Third Avenue, 9th Floor New York, NY 10158 USA (édition mise à jour de Mars 2002).
- 30 - Baltz RH, McHenney MA, Cantwell CA, Queener SW, Solenberg PJ. Applications of transposition mutagenesis in antibiotic producing streptomycetes. *Antonie Van Leeuwenhoek.* (1997). **71** (1-2): 179-187.
- 35 - Bate N, Butler AR, Gandecha AR, Cundliffe E. Multiple regulatory genes in the tylosin biosynthetic cluster of *Streptomyces fradiae*. *Chem Biol.* (1999). **6** (9): 617-624.

- Bate,N., Butler,A.R., Smith,I.P. et Cundliffe,E. The mycarose-biosynthetic genes of *Streptomyces fradiae*, producer of tylosin. *Microbiology* (2000). **146** (Pt 1), 139-146.
- 5 - Bayley C., Morgan, Dale E. C., Ow D. W. Exchange of gene activity in transgenic plants catalysed by the Cre-lox site specific system. *Plant Mol. Biol* (1992). **18** :353-361.
- 10 - Bentley,S.D., Chater,K.F., C  rdeno-Tarraga,A.M., Challis,G.L., Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H., Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M., Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S., Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S., Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G., Parkhill,J. et Hopwood,D.A. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* (2002). **417** (6885), 141-147.
- 15 - Bibb MJ, Findlay PR, Johnson MW. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene*. (1984) **30** (1-3):157-166.
- 20 - Bibb MJ, Janssen GR, Ward JM. Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene*. (1985). **38** (1-3) : 215-26.
- 25 - Bibb MJ, White J, Ward JM, Janssen GR. The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Mol Microbiol*. (1994). **14** (3): 533-45.
- Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schonher BE. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*. 1992;**116**(1):43-9.
- 30 - Blondelet-Rouault M-H., Weiser J., Lebrihi A., Branny P. and Pernodet J-L. Antibiotic resistance gene cassettes derived from the Ω interposon for use in *E. coli* and *Streptomyces*. *Gene* (1997) **190** : 315-317.
- Boccard F., Smokvina T., J.-L., Pernodet Fridmann A., and Gu  rineau M., Structural analysis of loci involved in pSAM2 site-specific integration in *Streptomyces*. *Plasmid*, (1989a). **21** : 59-70.
- 35 - Boccard F., Smokvina T., J.-L., Pernodet Fridmann A., and Gu  rineau M.. The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperate bacteriophages. *EMBO J*. (1989b) **8** : 973-980.

- Brunelli J.P., Pall M.L., A series of Yeast/ E. coli lambda expression vectors designed for directional cloning of cDNA and cre/lox mediated plasmid excision. *Yeast*. (1993) **9** : 1309-1318.
- 5 - Camilli A., Beattie D. T. et Mekalanos J. J. Use of genetic recombination as a reporter of gene expression. *Proc. Natl. Acad. Sci. USA* (1994) **91**(7), 2634-2638.
- Campelo, A.B. and Gil, J.A. The candidin gene cluster from *Streptomyces griseus* IMRU 3570. *Microbiology* (2002) **148** (Pt 1), 51-59.
- 10 - Carreras C, Frykman S, Ou S, Cadapan L, Zavala S, Woo E, Leaf T, Carney J, Burlingame M, Patel S, Ashley G, Licari P. Saccharopolyspora erythraea-catalyzed bioconversion of 6-deoxyerythronolide B analogs for production of novel erythromycins. *J Biotechnol.* (2002). **92**(3): 217-28.
- Chao K.-M., Pearson W.R., Miller W. Aligning two sequences within a specified diagonal band. *Comput. Appl. Biosci.* (1992) **8** : 481-487.
- 15 - Chater K. F. The improving prospects for yield increase by genetic engineering in antibiotic-producing Streptomyces. *Biotechnology*. (1990). **8** (2) : 115-121.
- Chaveroche MK, Ghigo JM, d'Enfert C. A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res.* 2000; **28**(22) :E97.
- 20 - Church GM, Gilbert W. Genomic sequencing. *Proc Natl Acad Sci U S A.* (1984) **81** (7):1991-5.
- Churchward G, Belin D, Nagamine Y. A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. *Gene*. (1984) **31** (1-3): 165-71.
- 25 - Comstock, L.E., Coyne, M.J., Tzianabos, A.O. and Kasper, D.L. Interstrain variation of the polysaccharide B biosynthesis locus of *Bacteroides fragilis*: characterization of the region from strain 638R. *J. Bacteriol.* (1999). **181** (19) : 6192-6196.
- Cox KL, Baltz RH. Restriction of bacteriophage plaque formation in *Streptomyces* spp. *J Bacteriol.* (1984) **159**(2): 499-504.
- 30 - Dale E.C., Ow D.W., Gene transfer with subsequent removal of the selection gene from the host genome. *Proc. Natl. Acad. Sci. USA* (1991). **88** : 10558-10562.
- Doumith M, Weingarten P, Wehmeier UF, Salah-Bey K, Benhamou B, Capdevila C, Michel JM, Piepersberg W, Raynal MC. Analysis of genes

involved in 6-deoxyhexose biosynthesis and transfer in *Saccharopolyspora erythraea*. *Mol Gen Genet.* (2000) **264**(4): 477-485.

- 5 - Draeger,G., Park,Streptomyces-H.H. et Floss,H.G. Mechanism of the 2-deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin *J. Am. Chem. Soc.* (1999) **121** : 2611-2612.
- Gandecha,A.R., Large,S.L. et Cundliffe,E. Analysis of four tylosin biosynthetic genes from the tylLM region of the *Streptomyces fradiae* genome. *Gene.* (1997). **184** (2) :197-203.
- 10 - Geistlich,M., Losick,R., Turner,J.R. et Rao,R.N. Characterization of a novel regulatory gene governing the expression of a polyketide synthase gene in *Streptomyces ambofaciens*. *Mol. Microbiol.* (1992). **6** (14) : 2019-2029.
- Gourmelen A, Blondelet-Rouault MH, Pernodet JL. Characterization of a glycosyl transferase inactivating macrolides, encoded by *gimA* from *Streptomyces ambofaciens*. *Antimicrob Agents Chemother.* (1998). **42**(10): 2612-9.
- 15 - Huang X. et Miller W. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* (1991). **12** : 337-357.
- Hara O et Hutchinson CR. Cloning of midecamycin(MLS)-resistance genes from *Streptomyces mycarofaciens*, *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J Antibiot (Tokyo).* (1990). **43** (8):977-991.
- 20 - Hara O et Hutchinson CR. A macrolide 3-O-acyltransferase gene from the midecamycin-producing species *Streptomyces mycarofaciens*. *J Bacteriol.* (1992). **174**(15) :5141-5144.
- 25 - Hoffmeister D, Ichinose K, Domann S, Faust B, Trefzer A, Drager G, Kirschning A, Fischer C, Kunzel E, Bearden D, Rohr J et Bechthold A. The NDP-sugar co-substrate concentration and the enzyme expression level influence the substrate specificity of glycosyltransferases: cloning and characterization of deoxysugar biosynthetic genes of the urdamycin biosynthetic gene cluster. *Chem Biol.* (2000). **7** (11):821-831.
- 30 - Hopwood, D.A. Future possibilities for the discovery of new antibiotics by genetic engineering. In Beta-lactam antibiotics. *Edited by M.R.J. Salton and G.D. Shockman. New York: Academic Press.* (1981). 585-598.
- Hopwood D. A., Malpartida F., Kieser H. M., Ikeda H., Duncan J., Fujii I., Rudd A. M., Floss H. G. et Omura S. Production of 'hybrid' antibiotics by genetic engineering. *Nature.* (1985a). **314** (6012): 642-644.
- 35

- Hopwood D. A., Malpartida F., Kieser H. M., Ikeda H. et Omura S. (1985b) *In Microbiology (ed S. Silver). American Society for Microbiology, Washington D. C.*, 409-413.
- 5 - Houben Weyl, 1974, in *Meuthode der Organischen Chemie*, E. Wunsch Ed., Volume 15-I et 15-II,
- Hutchinson, C.R. Prospects for the discovery of new (hybrid) antibiotics by genetic engineering of antibiotic-producing bacteria. *Med Res Rev.* (1988). 8 (4) : 557-567.
- 10 - Hutchinson C. R., Borell C. W., Otten S. L., Stutzman-Engwall K. J. et Wang Y. Drug discovery and development through the genetic engineering of antibiotic-producing microorganisms *J. Med. Chem.* (1989) 32 (5) : 929-937.
- Ishikawa J, Hotta K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol Lett.* (1999) 174(2):251-3.
- 15 - Kieser, T, Bibb, MJ, Buttner MJ, Chater KF, Hopwood DA. *Practical Streptomyces Genetics*. 2000. The John Innes Foundation, Norwich UK.
- Kuhstoss *et al.* Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene.* (1996). 183(1-2): 231-236.
- 20 - Lacalle RA, Pulido D, Vara J, Zalacain M, Jimenez A. Molecular analysis of the *pac* gene encoding a puromycin N-acetyl transferase from *Streptomyces alboniger*. *Gene.* (1989). 79(2):375-380.
- Lakso M, Sauer B, Mosinger B Jr, Lee EJ, Manning RW, Yu SH, Mulder KL, Westphal H. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc Natl Acad Sci U S A.* (1992) 89 (14) :6232-6.
- 25 - Li,T.B., Shang,G.D., Xia,H.Z. and Wang,Y.G. Cloning of the sugar related biosynthesis gene cluster from *Streptomyces tenebrarius* H6. *Sheng Wu Gong Cheng Xue Bao* (2001). 17 (3) : 329-331.
- 30 - Ligon,J., Hill,S., Beck,J., Zirkle,R., Molnar,I., Zawodny,J., Money,S. and Schupp,T. Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. *Gene* (2002). 285 (1-2), 257-267.
- Liu L, Saevels J, Louis P, Nelis H, Rico S, Dierick K, Guyomard S, Roets E, Hoogmartens J. Interlaboratory study comparing the microbiological potency of spiramycins I, II and III. *J Pharm Biomed Anal.* (1999). 20 (1-2):217-24.

- Mazodier P, Petter R, Thompson C. Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J Bacteriol.* (1989). **171** (6):3583-5.
- Merrifield RB, 1965a, *Nature*, 207(996): 522-523.
- Merrifield RB., 1965b, *Science*, 150(693): 178-185.
- 5 - Merson-Davies, L.A. et Cundliffe, E. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome, *Molecular microbiology*. (1994). **13** (2) : 349-355.
- Miller VL, Mekalanos JJ. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. *J Bacteriol.* (1988) **170**(6):2575-83.
- 10 - Nakano, Y., Yoshida, Y., Yamashita, Y. & Koga, T. Construction of a series of pACYC-derived plasmid vectors. *Gene* (1995). **162** : 157-8.
- Nielsen H., Engelbrecht J., Brunak S. et von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering*. (1997). **10**: 1-6.
- 15 - Oh SH, Chater KF. Denaturation of circular or linear DNA facilitates targeted integrative transformation of *Streptomyces coelicolor* A3(2): possible relevance to other organisms. *J Bacteriol.* 1997; **179** (1):122-7.
- 20 - Olano C, Lomovskaya N, Fonstein L, Roll JT, Hutchinson CR. A two-plasmid system for the glycosylation of polyketide antibiotics: bioconversion of epsilon-rhodomyacinone to rhodomyacin D. *Chem Biol.* (1999). **6** (12):845-55.
- Omura S, Kitao C, Hamada H, Ikeda H. Bioconversion and biosynthesis of 16-membered macrolide antibiotics. X. Final steps in the biosynthesis of spiramycin, using enzyme inhibitor: cerulenin. *Chem Pharm Bull (Tokyo)*. (1979a). **27**(1): 176-82.
- 25 - Omura S, Ikeda H, Kitao C. Isolation and properties of spiramycin I 3-hydroxyl acylase from *Streptomyces* and *ambofaciens*. *J Biochem (Tokyo)*. (1979b). **86**(6): 1753-8.
- 30 - Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. U.S.A.* (2001). **98** (21), 12215-12220.

- Pearson W.R. and D. J. Lipman. Improved Tools for Biological Sequence Analysis. *Proc. Natl. Acad. Sci. USA* 1988, **85**: 2444-2448.
- Pearson W. R. Rapid and Sensitive Sequence Comparison with FASTP and FASTA. *Methods in Enzymology*. 1990, **183**: 63-98.
- 5 - Pernodet JL, Simonet JM, Guerineau M. Plasmids in different strains of *Streptomyces ambofaciens*; free and integrated form of plasmid pSAM2. *Mol Gen Genet*. (1984);**198** (1):35-41.
- 10 - Pernodet JL, Alegre MT, Blondelet-Rouault MH, Guerineau M. Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two different mechanisms. *J Gen Microbiol*. (1993) ; **139** (Pt 5):1003-11.
- 15 - Pernodet JL, Fish S, Blondelet-Rouault MH, Cundliffe E. The macrolide-lincosamide-streptogramin B resistance phenotypes characterized by using a specifically deleted, antibiotic-sensitive strain of *Streptomyces lividans*. *Antimicrob Agents Chemother*. (1996), **40** (3): 581-5.
- 20 - Pernodet JL, Gournelen A, Blondelet-Rouault MH, Cundliffe E. Dispensable ribosomal resistance to spiramycin conferred by srmA in the spiramycin producer *Streptomyces ambofaciens*. *Microbiology*. (1999), **145** (Pt 9):2355-64.
- 25 - Pfoestl,A., Hofinger,A., Kosma,P. and Messner,P. Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91T. *J. Biol. Chem*. (2003), **278** (29): 26410-26417.
- 30 - Rao RN, Richardson MA, Kuhstoss S. Cosmid shuttle vectors for cloning and analysis of *Streptomyces* DNA. *Methods Enzymol*. (1987); **153**: 166-98.
- 35 - Raynal A, Tuphile K, Gerbaud C, Luther T, Guerineau M, and Pernodet JL.. Structure of the chromosomal insertion site for pSAM2: functional analysis in *Escherichia coli*. *Molecular Microbiology* (1998) **28** (2) 333-342.
- Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J., Kinashi,H. et Hopwood,D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome *Mol. Microbiol*. (1996). **21** (1) : 77-96.
- Richardson MA, Kuhstoss S, Solenberg P, Schaus NA, Rao RN. A new shuttle cosmid vector, pKC505, for streptomycetes: its use in the cloning of three different spiramycin-resistance genes from a *Streptomyces ambofaciens* library. *Gene*. (1987); **61** (3):231-41.
- Robinson, J.A. Enzymes of Secondary Metabolism in Microorganisms. *Chem Soc Rev*. (1988). **17** :383-452.

- Russell SH, Hoopes JL, Odell JT. Directed excision of a transgene from the plant genome. *Mol Gen Genet.* (1992). **234** (1): 49-59.
- 5 - Sambrook J, Frisch EF, Maniatis T . Molecular Cloning : a laboratory manual 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York USA (1989).
- Schonher B, Geistlich M, Rosteck P Jr, Rao RN, Seno E, Reynolds P, Cox K, Burgett S et Hersherberger C. Sequence similarity between macrolide-resistance determinants and ATP-binding transport proteins. *Gene.* (1992). **115** (1-2), 93-96.
- 10 - Sezonov G, Hagege J, Pernodet JL, Friedmann A and Guerineau M. Characterization of pra, a gene for replication control in pSAM2, the integrating element of *Streptomyces ambofaciens*. *Mol Microbiol.* (1995). **17**(3): 533-44.
- Sezonov G., Blanc V., Bamas-Jacques N., Friedmann A., Pernodet J.L. and M. Guerineau, Complete conversion of antibiotic precursor to pristinamycin IIA by over expression of *Streptomyces pristinaespiralis* biosynthetic genes, *Nature Biotechnology.* (1997) **15** : 349-353.
- 15 - Sezonov G, Possoz C, Friedmann A, Pernodet JL, Guerineau M. KorSA from the *Streptomyces* integrative element pSAM2 is a central transcriptional repressor: target genes and binding sites. *J Bacteriol.* (2000). **182**(5):1243-50.
- 20 - Simon R., Priefer U and Pühler. A broad host range mobilisation system for in vivo genetic engineering : transposon mutagenesis in gram negative bacteria. *Bio/Technology* (1983). **1** : 784-791.
- Simon, *et al.*, Plasmid vectors for the genetic analysis and manipulation of rhizobia and other gram-negative bacteria, p. 640-659. In A. Weissbach, and H. Weissbach (eds.), *Methods in enzymology*, vol 118, Academic Press, Inc., Orlando, 1986
- 25 - Summers,R.G., Donadio,Streptomyces, Staver,M.J., Wendt-Pienkowski,E., Hutchinson,C.R. et Katz,L. Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. *Microbiology* (1997). **143** (Pt 10) : 3251-3262.
- 30 - Van Mellaert L, Mei L, Lammertyn E, Schacht S, Anne J. Site-specific integration of bacteriophage VWB genome into *Streptomyces venezuelae* and construction of a VWB-based integrative vector. *Microbiology.* (1998). **144** (Pt 12): 3351-8.
- 35

- Vara J, Lewandowska-Skarbek M, Wang YG, Donadio S, Hutchinson CR. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J Bacteriol.* (1989). **171**(11): 5872-81.
- 5 - Vara J, Malpartida F, Hopwood DA, Jimenez A. Cloning and expression of a puromycin N-acetyl transferase gene from *Streptomyces alboniger* in *Streptomyces lividans* and *Escherichia coli*. *Gene.* (1985). **33**(2): 197-206.
- 10 - Wahl, G. M., K. A. Lewis, J. C. Ruiz, B. Rothenberg, J. Zhao, and G. A. Evans. Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. *Proc. Natl. Acad. Sci. USA* (1987). **84**: 2160-2164.
- Waldron, C., Matsushima, P., Rosteck, P. R., Broughton, M. C., Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J. et Baltz, R. H. Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*(1) *Chem. Biol.* (2001) **8** (5) : 487-499.
- 15 - Walczak RJ, Hines JV, Strohl WR, Priestley ND. Bioconversion of the anthracycline analogue desacetyladiamycin by recombinant DoxA, a P450-monooxygenase from *Streptomyces* sp. strain C5. *Org Lett.* (2001). **3** (15): 2277-9.
- 20 - Wang, Z. X., Li, S. M. et Heide, L. Identification of the coumermycin A(1) biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. *Antimicrob. Agents Chemother.* (2000) **44** (11), 3040-3048.
- 25 - Ward, J. M., Janssen, G. R., Kieser, T, Bibb, M. J., Buttner, M. J. & Bibb, M. J. Construction and characterization of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase from Tn5 as indicator. *Mol Gen Genet* (1986). **203** : 468-478.
- Wehmeier UF. New multifunctional *Escherichia coli*-*Streptomyces* shuttle vectors allowing blue-white screening on XGal plates. *Gene.* (1995). **165**(1): 149-150.
- 30 - Wilms B, Hauck A, Reuss M, Syltatk C, Mattes R, Siemann M, Altenbuchner J. High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the *Escherichia coli* rhaBAD promoter. *Biotechnol Bioeng.* (2001). **73**(2) : 95-103.
- 35 - Worley K. C., Wiese B. A., and Smith R. F. BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Research* (1995). **5**: 173-184.

- Wu K, Chung L, Revill WP, Katz L et Reeves CD. The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene*. (2000). **251** (1): 81-90.
- 5 - Xue Y, Zhao L, Liu HW et Sherman DH. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc Natl Acad Sci U S A*. (1998). **95** (21): 12111-12116.
- 10 - Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG et Court DL. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A*. (2000). **97**(11): 5978-83.

SEQUENCE LISTING

<110> Aventis Pharma SA

CNRS

<120> POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES SPIRAMYCINES,
SEQUENCES NUCLEOTIDIQUES CODANT CES POLYPEPTIDES ET LEURS
APPLICATIONS

<140> 2,501,445

<141> 2003-10-08

<130> 11246-62

<150> FR 0212489

<151> 2002-10-08

<150> FR 0302439

<151> 2003-02-27

<150> US 60/493,490

<151> 2003-08-07

<160> 161

<170> PatentIn version 3.2

<210> 1

<211> 30943

<212> DNA

<213> Streptomyces ambofaciens

<400> 1

gaattcggca attcccgggg cgccgggaaa ccggcacgcc attccgacca cggcaacggc 60

gtcggaccgg tcgtcactgc gggaaatcgc gagttctcca gacacctgca tccctcatat 120

gttcgccgta ccacgccgtg gcttttctgc cttttcttga tcttcccgag cgtacagcgg 180

gcgaactgcc gggcgacagg aacggccggt gaacgacaat caattgtgtc aagttcggag 240

attgtccacc gattctcgtc accggagacg gcgcgggatc atgcgggacc acacggcatc 300

gcaccaggtc cgcagggtcg gcgctccgac gtcccggccg gtcgcgcaact ccggtgacct 360

gcacgcggag tcctggggcga gcggagaaga tttcagtacc tcgcggcgcg ggacaaccac 420

ttcgcgacga atatgtcagg tctcccagag cggtcctgtc cgtcccgcgc tgacctgtcc 480

gcgcccgtcc cgctccgcgc cggtcgtcat ctcggccgga atttcagtcg gcagctcatt 540

gtgtcaggtt cgccttgccg acattctccg gagattccta agctctgccg gtaaccggga 600

ccggaaccac cgtgccgcgc gttcgggtcca cacaccgctt ttcgaggagt ccgactgatg 660

ggtgaggccg tgacgggacc gatggagctg agcaaggacg cggacgcccg ggggctgctt 720

gagtggttcg cgtacaacag gacgcgtcat ccggtgttct gggacgagac ccgacaggcg 780

tggcaggctt tcggctacga cgactacgtg acggtgtcga acaaccgcga gttcttctcc 840

tcggacttca acatggtgat gccgacgccg ccggaactgg agatgatcat cgggtccgggc 900

acgatcggcg cgctggaccc gcccgcgcac ggaccgatgc gcaagctggt gagccaggcg 960

ttcaccccc gacggatcgc ccggctggag cccaggggtgc gcgcgatcac cgaggagctc 1020

ctggacaagg tggggcagca ggacgtcgtc gacgccgtgg gtgacctgtc ctacgcgctg 1080

ccggtcatcg tgatcgccga actgctgggc ataccgcgcg gcgaccgtga cctgttccgg 1140

gagtgggtcg acaccctgct gacgaacgag ggcctggagt acccgaacct cccggacaac 1200

ttcaccgaga cgatcgcgcc cgcgctcaag gagatgaccg actacctcct gaagcagatc 1260

cacgccaagc gggacgcgcc cgccgacgac ctggtcagcg ggctgggtcca ggcggagcag 1320

gacggccgcc ggctgaccga cgtcgagatc gtcaacatcg tcgcgctgct cctgacggcg 1380

gggcacgtct cctccagcac cctgctcagc aacctgttcc tggtcctgga ggagaacctg 1440

caggcgctgg aggacctgcg ggccgatcgc tccctgggtgc ccggcgcgat cgaggagacg 1500

ctgcgctacc gcagcccctt caacaacatc ttccggttcg tcaaggagga caccaccgtc 1560

ctcggtcgcg tcatggagaa gggccagatg gtgatcgctt ggagccagtc cgccaaccgg 1620

gacccccggc acttcccga cccggacacc ttcgacatcc gccgctcgga cggcaccggg 1680

cacatggcct tcgggcacgg catccaccac tgcctgggtg ccgccctcgc ccgcctggag 1740

ggcaaggtca tgctcgaact cctcctggac cgggtccaag gcttccgcat cgaccacgag 1800

cacaccgtgt tctacgaggc cgaccagctc actccgaagt acctgcccgt ccgggtcgac	1860
tggaactgaa cccgaggggtc tcgtcccgga gtccagggcc gtcccagacc ggccctggac	1920
ctcacgaccg cccgataagg agcgccgcca tcgccgagaa cacagccgag ctccctgccc	1980
ggcgggtcgg caggatcaag ccgtgccggc tgatcaggct cgagcagcac atcgacccgc	2040
gcggcagcct ctccgtgata gagtccggcg tgaccgtgga ctccccgctc cgacgcgtct	2100
actacatgca tggccagacc cagtcctctc ccccgcgcg cctgcacgcg caccgcaccc	2160
tggaacaact cgtcatcgcc gtccacggcg ccttctccat caccctcgac gacggcttcc	2220
agcacgccac ctaccgtctg gacgaaccgc gagccggact ctgcatcggc cccatggtct	2280
ggcgcgtcct gaaggacttc gaccccgaca ccgtggccct ggtcctcgcc tcgcagcact	2340
acgaggagtc cgactactac cgcgactacg acaccttcct gcatgacgca cggagcctca	2400
catgaccatc cccttcctcg acgcggggcg cggtaccgg gagttgcgag ccgagatcga	2460
cgcggccctg cagcgggtgt ccgcctccgg ccgctatctg ctgcacgcgg aactcgcggc	2520
cttcgaggag gagttcgccg cgtactgcga caacgaccac tgtgtggcgg tgggcagtgg	2580
ctgcgacgcg ctggagctgt ccctgcgggc gctggacatc ggtcccgggg acgaggtggg	2640

ggtgcccgcg cacaccttca tcgggacctg gctggccgtg tccgctaccg gggcacggcc 2700

ggtggccgtc gacccgacgc cggacgggct ctccctcgac ccggcgctgg tggaggcggc 2760

gctcaccctt cggaccagag ccctgatgcc ggtgcacctg cacgggcacc cggccgacct 2820

cgacccgcta ctggcgatcg ccggacggca cggcctggcc gtggtcgagg acgccgcgca 2880

ggcccacggc gcccgttacc ggggccgcag gatcggctcg ggccacgtgg tcgcgttcag 2940

cttctacccc ggcaagaacc tcggcgccat gggggacggc ggcgcggtgg tcacgggtga 3000

ctccggtgtg gccgagcgga tccggttgct gcgcaactgc ggctcgcggg agaagtaccg 3060

gcacgaggtg cgctcgaccc actcccggct cgacgagttc caggcggccg tgctgcgggc 3120

caaactgccg cggctcgacg cgtggaacgc ccgccggggc ggcacggccg aacggtacgg 3180

gcggggccctg ggtccggtac cgcagatcgc cgtcccggtc accgctccct gggccgaccc 3240

ggtgtggcac ctgtacgtga tccgctgcgc ggagcgcgac gagctgcgcc gccggctgga 3300

acgagccggg gtccagaccc tgatccacta ccccgtgccc ccgcaccggt ccccggccta 3360

cgccgacgac ccggccggcg caccggcggg gaccaccccg ctcaagtgagc gcctggcggc 3420

gcagagcctc agccttcccc tgggaccgca cctcggggag gacgaggccc gcgccgtcgt 3480

ggcggcggtc cgggcggcgt ccgcagggct ggcggcgtag ccgacgccgg acggccagcg 3540

ttttcctcta	gtgacggaga	aacgatgacc	gaggtcatgt	cagggcgtcc	cggaatgaaa	3600
gggatcatcc	tgcagggcgg	cggagggacc	cgcctacgcc	ccttgaccgg	cacgctgtcc	3660
aagcaactgc	tgcccgtcta	cgacaagccg	atgatctact	accgctgtc	cgtcctgatg	3720
ctgggcggca	tccgcgagat	cctcgtcgtc	tcctccaccc	agcacatcga	gctgttccag	3780
cggctgctgg	gcgacggctc	ccgcctcggc	ctcgacatca	cctacgccga	acaggccgag	3840
cccgagggca	tagcgcaggc	catcaccatc	ggcaccgacc	acatcggcga	ctcaccggtc	3900
gcgctcatcc	tgggcgacaa	catcttccac	ggccccggct	tctcggccgt	gctccagggc	3960
agcatccgcc	acctcgacgg	ctgtgtgctg	ttcggctacc	cggtcagcga	cccgaagcgc	4020
tacggcgctc	gcgagatcga	cgaccagggc	gtactgctgt	ccctggagga	gaaaccggcc	4080
cggccccgct	ccaacctcgc	cgtcaccggc	ctctacctct	acgacaacga	cgtgggtcgac	4140
atcgccaaga	acatccggcc	ctcggcgcgc	ggcgaactcg	agatcacgga	cgtcaacagg	4200
acctacctgg	agcagaaacg	cgcgggctc	atcgaactgg	gccacggctt	cgcctggctc	4260
gacatgggca	cccacgactc	cctcctccag	ggcggccagt	acgtccagct	catcgagcag	4320
gcccagggag	tgcggatcgc	ctgcatcgag	gagatcgccc	tgcgcatggg	cttcatcgac	4380
gccgacaccc	tccaccggct	cggccgcgaa	ctgggcacct	ccggatacgg	cgcgtacctg	4440

atggaggtgg ccacccgtgc aggcaccgaa tgagacgccg cgccggcccc cccgctccgc 4500

cgcccgacgg ccgcccggccc ggatcctcgt caccgggggc gccggcttca tcggctcgcg 4560

cttcgtgaac gcgctgctgg acggctccct gccggagttc ggcaaaccgc aggtgagggc 4620

gctcgacgcg ctcacctacg cgggcaacct ggccaatctg gccccggtgg gcgactgtcc 4680

ccggctgcgg atcttcccgg gggacatccg cgaccgcggc gcggtcacc caggcatggc 4740

gggggtcgac ctggtggtgc acttcgcggc cgagtcgcac gtggaccgct cgatcgacga 4800

cgccgacgcc ttcgtgcgca ccaacgtgct gggcaccag gtcctcctcc aggaggcact 4860

ggccgtacgc cccgggctgt tcgtgcacgt ctcgacggac gaggtgtacg gctccatcga 4920

ggaggggtcc tggcccgagg agcaccgcgt gaaccccaac tcgccctacg ccgcctcgaa 4980

ggcgtcctcc gacctgctgg cgctggccca ccaccgcacg cacggactgc cgggtgtgcgt 5040

caccgctgc tccaacaact acgggcccta ccagtaccgc gagaagatca tcccgtgtt 5100

caccagcagc ctctcgacg gcgggaccgt cccgctctac ggggacggcg gcaaccggcg 5160

cgactggctg cacgtggacg accactgccg gggcatcgcc ctggtggccc ggggcggccg 5220

gccccgcgag gtctacaaca tcggcggcgg caccgagctg agcaacgtcg agctcacgga 5280

gcgtctgctg aaactgtgcg gagccgactg gtcggcgggtg cggcgggtgc ccgaccgcaa 5340

gggccacgac	cggcgctact	ccgtcgacta	caccaagatc	gcggacgagc	tgggttacgc	5400
gccgcggatc	accatcgacg	aagggctgga	gcggaccgtg	cactgggtacc	gggagaaccg	5460
cgcgtggtgg	gcgcccgcga	agagggggcg	atgacggtga	cgaccgcata	cgtggacccg	5520
ctcgacctgt	ggctccgccc	gtaccagccg	tccgcgtcac	ccgccgtccg	gctgggtgtgc	5580
ttcccgcacg	cgggcggctc	ggcgagttcg	ttcctgccgt	tcacccggca	gctgccggac	5640
cggatcgagg	tcgtggccgt	ccagtacccc	gggcgccagg	accgcaggag	cgaaccgctg	5700
gtcgacacca	tcgagggact	ggccgagccc	ctggccggcc	tgctggaggc	gcaggccggc	5760
cccccggtgg	tgctgttcgg	gcacagcatg	ggcgcgctgg	tggcctacga	ggtcgccccg	5820
gcgctccagc	ggcggggagc	ggctccggtg	cgcctggtgg	tctccggggc	ccgggcccc	5880
gccgtcgacc	ggccgatgac	cgtgcacctc	tacgacgacg	accggctggt	cgaggaactc	5940
cgcaagctcg	acggcaccca	cagccaggtg	ttcgccgatc	cggagctgct	ccggctggtg	6000
ctgcccgtga	tccgcaacga	ctaccggggc	gtggcggcct	acgccaccg	cccggggggc	6060
ccgctggact	gccccctcac	cgtgttcacc	ggcgcgcgac	accccaccgt	gaccgcggcc	6120
gaggcggcgg	cctggcacga	ggcggcggcg	tccgacgtcg	agacgcgcac	cttccccggt	6180

ggccacttct tcccgtacca gcggaccgcg gaggtgtgcg gggccctggt cgacacgctc 6240

gagccgctgc tgtcggcccg gacgcgcggt gtccggcggg tccgcccggg gtgacgtggg 6300

cacggtcgag tacgccgtcc accggcgtag cgcggaacgg gtgaggggtct ccgccgacac 6360

cctggacagc ccggtcaccg cgctggcggg ggtgccccgc tggctggagg aataccaccg 6420

ggcgcaccgc ttccacgtcg agccgatccc cttcgaccgg ctccggcgggt ggtccttcga 6480

gccgggcacc ggcgacctgc ggcacgagac gggccgcttc ttctccgtgg aggggctgcg 6540

caccagctcg gacgccgatc cggtcgcccc tgtccagccg atcatcgtgc agcccagagt 6600

ggggctgctc ggcatcctgg ccggggagtt cgacgggggtg ctgcacttcc tgatgcaggc 6660

caaacccgag ccgggcaacg tcaacggggt gcagatctcc cccacgggtgc aggccacgcg 6720

cagcaacttc gacgaggtgc accacggcgg gtccaccccg ttctcgacc acttcatcca 6780

ccgccccggc cgccgggtcc tgatcgacag catccagtcc gaacagggcg actggttcct 6840

gcacaagcgc aaccgcaaca tggtcgtcga gatcgacacc gacatcgagg ccgacgcgcg 6900

gttccgctgg ctgaccctcg ggcagatccg ccggctgatg ctccaggacg acctcgtcaa 6960

catggacacc cgcagtgtgc tggcctgtct gccacccgcg cacggcacgc ccgacgacgg 7020

tgacgactcc ttcccggcgg cgctgcgcgg ctccctctac ggggagaccg cgccgttgca 7080

cgatctgcac gccatcacca gctgcctcac cgacgtccgg gcgctgcggg tgctgcgcca 7140

gcagagcgtg ccgctcgacg acgcccggcg ggacggctgg gagcggaccg ggagcgcgat 7200

ccggcatcgc agcggcaggc atttcgagat catggcgggtg gaggtgaccg cggagcgccg 7260

tgaagtggcc tcgtggaccc agccgttgct gcgcccgtgc tcgcagggac tggcggccct 7320

gatcaccg cgatcaacg gggtgctgca cgccctgggtg gcggcgcgggt cggaggtcgg 7380

cacgtcaac gtcgccgagt tcggaccgac cgtccagtgc cggcccgcg aggcggacgg 7440

ccagtcgccc ccgtacctgg accgggtgct gacggccgga gccgaccgcg tccgctaoga 7500

cgtgggtgcag tcggaggagg gcgggcgctt ctaccacgcg cgcaaccgct atctggtggt 7560

cgaggcgggg ccggagctcg acacgggctg cccgcccggc ttctgctggg ctaccttcgg 7620

ccagctcacc gaactgctcg cgcacggcaa ctatctcaac gtcgaactcc gcaccctcat 7680

ggcgtgcgca cacgcctcct actgaatggt cacgaaagct gcaccgcgcg ggagaatcgg 7740

cagcgcgcca ccggccggcc ggcaccgga aggtaaagcg ccgttctccc gcatcggcgc 7800

cctgcgggaa acggcggaac ggccggccc gaccgcgcgc aattcccggc gggacacggt 7860

gggagcccgc acgaggaacc gctttccccg ccttcgggtgc gcccggcgc gggaccaccc 7920

ccgcctcccg gccgggccc ggaatacgac gggggcggcc gaggacattc ctttcccgc 7980

tccggaaaag cgcgccccga gggcccccca atgccggggc ggacggacgg cgactgcgcg 8040

cggacggcgg cccggcgctc aacgcacctg cccgagtccg gacgagacag cgcgacgcga 8100

gaggcgaaaa tgatcaatct cttccagccc cagatggggg ccgaggaact ggcggcggtg 8160

tccgaggtct tcgacgacca atggctcggc cacggacccc ggaccgcggc gttcgagtcc 8220

gcgttcgcgc agcacctcgg ggtcggcccc gagcacgtcg tcttcctcaa ctcgggcacc 8280

gccggcctct tcctggccct ggagtcgctc ggccctgcgg ccggcgacga ggtcgtgctc 8340

ccctcgccca gcttcctcgc cgcggcgaac gccgtacagc tctcgggagc gcgcccgggtg 8400

ttctgcgaca ccgaccgcgc gacgctgaac cccgccctgg agcacatcga ggcggccgctc 8460

accccgcgca ccagggccgt catcgcgctc cactacggcg gccaccccg gcacatcgtg 8520

cgcacgcgcg agcgtgcgc ggagcggggc atcacctga tcgaggacgc cgcgtgctcc 8580

gtggcctccc gcgtcgacgg ccgaccggtc ggcaccttcg gcgacctcgc catgtggagc 8640

ttcgacgcca tgaaggtcct ggtcaccggc gacggaggga tgatctacgt caaggacccc 8700

ggggcggccg cccggatccg gcgcctcgcc taccacggcc tcacgcggtc cagcggcctg 8760

ggatacgcca gggctctcgg gcgctgggtg gagatggacg tccccgaacc gggccgcgcg 8820

gtcatcgga acgacctcac cgcggccatc ggcgcggtcc agttgcgcgc gcttcccggc 8880

ttcgtggccc gccgcaggga gatcgtcgcc ctgtacgaca gcgaactgag ctcgctggag 8940

ggcgtgctga caccgcccgc gccacccgcg gggcacgagt ccacgcacta cttctactgg 9000

atccagctgg ccccggcgt ccgggaccgg gtggcacgcg acctgctcac cgacggcatc 9060

tacaccacct tccgctacgc acctctgcac aaggtgcccg cctacggcca caccggaggc 9120

gaactgcccg gcgtggagcg ggcgtccgaa cggaccctgt gcctgcccct gcaccccggc 9180

ctgtcggacg ccgacgtccg caccgtcgtg tectccctgc gcagagccct gagcgccgcg 9240

gatccggccc ccgcctgacg cgcgcacgcg ccacggcccc tgtcccggcg gtgaccgccg 9300

ggacaggggc cgtggcgcgt cctgcggacg gtccggggcca ccccgccgtc ccgtcgggtgc 9360

gtcagcgacg cgtgccgagg aagaagcccc gtgagccctc gcccgtcacg acgtactcga 9420

cgtccaggcc ggccttgccg tacgcggcct cgtactccgc acgcgagaac agcgtgaggt 9480

agtccacctc gctgcgggtgc cggatgccct ccgcggtgtg cgcgatcagg tagtggatct 9540

ccatccgggt ccgcccgcgc tcccgggtcg agtgggagac acgggcgacg ccctgggtcgc 9600

ccggtgccgt ccgcagggcg tgggtggaga cgtggccgtc caggaagggt tcggggaagt 9660

accacggttc gacggccagg acaccgtccg cggtcagggt ccgcgccatg gcggcgacgg 9720

cgtcctccag gtcggccgtg gtctccaggt acccgatcga gctgaacatg cagaccacgg 9780

cgtcgaacgt ctgcgcgagg tcgaacgccc gcatgtcacc gcggtggagg gtgacgccgg 9840

ggagccgctc ctcggcgcgg gccgccatcc actcggacag ctccaggccg ctcacgcggt 9900

cgtagagctt ggcgaaggcc tccaggtggg taccggtgcc gcacgcgatg tcgagcaggc 9960

tggccgcgtc cggcgtagct tcccggatca ggtcggtagc ccgggcggcc tcaccgcggt 10020

agtccttgcg gtcctggtag agcaggtcgt agacctcggc ggcactgtcg ttctcgtaca 10080

tgggaatcct ccgggccgtg aagggggaac cgtgggtctg tcgaagagga gtgcgccgtg 10140

cgctcggggc gggggcggcc tgccgccggc ccctcgcgat gatctccgta cggacacca 10200

acagttcacg ccgagccggg gtcaaggaac gacggggtgg tcagtcaagt cgggcgctcc 10260

gccccggggg cggggcgcgc cggcgacgcg cacggattcg gccaacgggt tgctcgttcc 10320

gccggaaatc acggtgtggc ccccgggcca ccgggtagct tatgcctcgt tcaccgcagc 10380

ggttgaagag gcagccttca accccggccc ggcctttatg gaattcattt ccaccgtgcc 10440

gcaacacca ctgaaggacg gccggatatc ggccatgaag ccccggcctt tcagccaggc 10500

accctctctt gtcgaataga gtatgtcctc cgctgaagcc gccgaagacg gacgaagggg 10560

acgaacggtc acctcggtag atctagacgg aatccttgaa agcgtaatag cctgtcaatg 10620

cttttggtaaa gcacagggat ggggggtgcct gcgggatgag tgacctgggt tctggtgaag 10680

aagggtccga agaagacgag tcggacgacg cactcgcctt cctcgagtgc atcgcccggg 10740

cggcaccacg gagcgaatac gaccgggtca tggcccgcgc cgaacgctcg ggcgccgacg 10800

aggaccggat gcgccgactg gagcgcttca accggctcgc cctcaccgcg cagtcgatga 10860

tcgagtaccg ccgcgaccgg gaggcggagc tcgcggccct ggtcgacgcc gcgcacgagt 10920

tcgtcgccgc ccggcggggc aaggacctgc tggagtccat cggccgcaga gcacggctgc 10980

tgctgaagct ggacgtctcc tacgtcggcc tgcacgagga ggaccggccc ggcacggtgg 11040

tgctgagcgc cgacggcaac gcggtcaagg tcgccgagag ctaccggctg ccggccgacg 11100

gcggactggg cgccatggtg cgcacctgcc gcgctccctt ctggaccccg gactacctcg 11160

gggacaacag cttcacgcac gtcgaggccg tcgacgacat cgtccgcgcc gaaggcctgc 11220

gcgcggtcct ggccgtcccg ctgtgcgcgc ggggcgaacc gatgggggtc ctctacgtcg 11280

ccgacctca ggtgcggcat ctgaccccca acgaggtcac cctgctgtgc tcgctcgccg 11340

atctggccgc ggtggcgatc gagcgcaacc ggctggtcga ggagctccac gacaccatcg 11400

ggcaactgcg ccaggacatc ggcgaggccc gcaccgcctt cgcgcgcacc cgcaggtccg 11460

ccgacctcca gtcgcacctg gtcacgcagg tgatggacag gcgcggcgcc gactcgttac 11520

tcgcgacggc cgccgaggcg ctcggcggcg gagccggcct gtgcagcccg ctcgggcgcc 11580

cgctcgccga gtacgggacc ctgcgccccg tcgccccac ggaactgcgc gcggcgtgcc 11640

gccggggcgc cgagaccggc cggcccacct ccgtggcccc gggggctctg acggtgcccc 11700

tgcttcccgg gggcaacgcc ggcttcctgc tgaccgacct cggtcggac gcggaccaca 11760

ccgccgtccc cctgctcccg atggtcgccc gcacctcgc gctgcacctg cgcgtccagc 11820

acgacgactc cccaaggcg cagagccacc aggagtctt cgacgacctg atcggggcgc 11880

cccgtcacc cacgtcctc agggaacgcg ccctgatgtt ctccctcagc ttccgccgcc 11940

cgcacgtggt gctggtggcg gacggacccc gcgggacctc gccgcggctg gaggcctccg 12000

gcgccgacta cgcgaaggag ctcggcgggc tgtgcagcgt gcgggacggc gccgtcgtcc 12060

tgctgctgcc cggcgacgac cccgtcgccg tggcgacgac cgccgccccg gagctgaccg 12120

accgcgcgg gcaccccgtc accgtggggg tcgcgggccc cgcctcgacc gtcgacggca 12180

tcgccgacgc gcaccgtgag gccgcgaagt gtctggagac cctccgcgcg ctcggcggcg 12240

acggcggcac cgcgtgcgc tccgacctgg gtttcctcgg catgctcctc gccgaggaga 12300

acgacgtccc cggttacatc aggacgacga tcggccccgt ggtcgactac gacaccacc 12360

gcttcacgga tctggttccc actctgaggg tgtacctgga gtcgggcagg agccccacgc 12420

gtgccgcaga gacactgcgc gtgcacccga acaccgtctc acggcggctg gagcgcatcg 12480

gcgtactgct gggagaggac tggcagtcac cggagcgggt gctggacata caactggccc 12540

tgcggtctta tcaggtgcgc tcggcgtctt cctcgcaacc ggcgctccgag acccgggccc 12600

tgctcggatc gctgcgcgag tgacctctcc ggacggggcg gcgccccgcc cgtcggcgcg 12660

gatcgcgccg acaggcgggc gttgaccacc ggccaccggt ccccgtacg acgaggtccc 12720

gggcccgcgt ggttcagccg gcgctgaact cggcgatgcg cccgtctccc atggtcagcc 12780

gggcgcgcgt gaaccgggac cgcattccgac ggtcgtgggt gacgaccacg acggcgccgc 12840

ggtagtccgc gagtgcctgc tccaactcct ccaccagcac cggggtgagg tggttggtgg 12900

gctcgtccag cagcagcagg tccatcgggt cgctcaccag ccgggcgatc tcgatccggc 12960

ggcgctgccc gtaggacaga tccttcacgc gtcgcccag gtcggacggg ctgaacaggc 13020

cgagcgacag cagtttctcc gcgtggctct ccaggtagcc ctcccggccc tgggcgaagg 13080

cccgcagcac ggtcagtccg ggcgcccagg gcgtctcgtc ctgccgcaga tgaccgacct 13140

ggcagccgac gcgcaccgag ccgccgtccg gctccagttc cccggacagc acccgcaaca 13200

aggtggactt gcccgcaccc ttgggacctg tgacgagcag ccgttcgccg ggccggatcg 13260

tcagggagtc cacggcgagc cgacccgcga cgcgcacgtc ggtgagttcg gccaccgcct 13320

cctccgcctc cgggcccgcg gtgtcgatgc gggcggcgaa ggacaacggg tcggcgggag 13380

cgtggaccgg gttctcggtc agctgcgcca cgcgttgctt cgcgttgcg atccgcacca 13440

tcgcgccgtg gtcgcgcctt cgctgcggt aggcgccgtg gccgaacacg gagagggaca 13500

tcttgcgcg gatgccgtcc atccgcgcca cggtggaggt gatcagccc cggttgcggt 13560

cgagttcggc acgccactcc tcgtactccc gcagccgccg ctcgcgttcc acggccttgg 13620

ccgtcaggta gccctcgtag ccgttgccgt agcgggtgac gctgccggag tcgacctcca 13680

ggatcgtggt ggtgagccgg tcgaggaaga cccggtcgtg ggtgaccgcg atcaccgtgc 13740

cgcggtggcc ggccagggtg tcctccagcc attccatcgc ccggtcgtcg aggtcgttgg 13800

tcggttcgtc caggagcagc agctccggcg acgaggcgag ggtcgcggcg agcgcgaggc 13860

gggagcgttc gccaccggag agggttccga gcttgcggtc gcggtccagg ctcggcagtc 13920

cgaggccgtg cagcgcgacc tccacgcgca cgtcggcctc gtagccgcca cgcgcctggt 13980

actgctcgac cagagcggcg tagcgctgga ggccggcgga cagctcgcgc tcggagccgt 14040

tctcgtcgct ctcgcccagc tccgcctcgg cctcgcgcat cgccgcttcg agctcgcgca 14100

ggtcggacag ggccagggtc acggcgctct ggacgggtggc gtcgaggggc agttccagtg 14160

tctgcgccag gtagccgacg ccgccgggag cgaccacggt gagcgcgccg ttgtcgggct 14220

ccacgcggcc ggcgaggatc ttgagcagcg tggacttgcc ggaaccgttg tcgccgatca 14280

cgccgacctt ctcgcccggc ttgatgctga aaccgacccg gtcgagcacg acacagtcgt 14340

ggtagcgctt cgtgatgtcg tgtagggcgt attgcgcaat cgacacgcgt aagtctcctg 14400

tttccacgat gaggatgagt ggatgcgtga gcgcgctcgc agaaagaacg gaaagcagaa 14460

gggacgccac cactgcggac atggccggtc aggggggtgc acgagcacgc tgctggatgc 14520

ggcaggcgga gtcagctcaa cgccggggcat cctcatctat cacagagatc ccatgcgaag 14580

aactatagcc gtgctacccg gtgccgcgca acggtatgcg tgtcgcaccg gccgacgtga 14640

tgcagtcgga ccgggatgat cgcttgtccg gcggccggat gcctagcctc gggagcaacc 14700

acagcgggtct ttcacgagag gggtcgacca tgggcgatct caggaaccgc atcaccgagc 14760

tggtcgcgcg gtaccaccgg gaacaggcgc ccgggggctt cgttcccggg acgacgcacg 14820

taccggtctc cggcgcggtg ctgagcgagg aggaccggct ggcgctggtg gagacggcgc 14880

tggagatgcg gatcgcggcc ggcccggcct cccggggctt cgagcggcag ttcgcccggt 14940

acctcgggct ccggaaggcg cacctgacca actccggttc ctccgccaac ctctcgcgcc 15000

tcggcgcgct cacctcgccg cagctggagg agagacggct gcgtccgggg gacgaggtcg 15060

tcacggtcgc cgccgggttc cccacgacgg tcaacccgat cttccacaac gggctgggtgc 15120

ccgtcttcgt ggacgtcgag ctcggcacgt acaacacgac gcccgagcgc atcgagcggg 15180

ccatcggccc ccggaccagg gcgatcatga tcgcgacgc cctgggcaac cccttcgagg 15240

ccgaagaggt ggcccgcctc gcggacgagc gggagctgtt cctcatcgag gacaactgcg 15300

acgcggtggg gtcccgtac cggggcaggc tcaccggctc ctccggcgac ctgtcgaccg 15360

tcagcttcta tcccgcgcac cacatcgca tgggtgaggg gggctgcgtg ctcaccgaca 15420

acctggccct ggcgcggatc gtggaatcac tgcgcgactg ggggcgcgac tgctgggtgcg 15480

agccgggtga ggacaaccgc tgccctaagc ggttcgacca gaagatgggt gacctgccgc 15540

ccgggtacga ccacaagtac atcttctcgc acgtcggtta caacctgaag tcgaccgacc 15600

tgcaggcggc cctcgggctg tcccagctga cccggatcga ggagttcacc gaggccaggc 15660

gcccgaactg gcggcatctg cgcgccgcgt tggacgggct gcccggtctg ctgctgcctc 15720

atgccacacc gggcagcgat ccgagctggg tcgggttcct catcacctg gaccgggacg 15780

ccgcgtacag cagggcggcc ctggtcgacc acctggaatc gcgccggatc agcaccgcc 15840

gcctgttcgg gggcaacctc gtgcggcacc ccgcctacac cgaccgtcgg taccgggtgt 15900

ccggtccct ggagaacagc gacctgatca ccgaccagac gttctggatc ggggtcttcc 15960

ccggcatcac cccggagatg atcgccctacg tcggcgacac gatccgggag tgcgtgctca 16020

agcactcctg agggggcggcg gtacgggggt gcgcgtggag gggaggaccg gggcggtcct 16080

cctccggggc gtcacggccg cggaagcgtc cgaagctccg aggcgaggtc ggggtggacc 16140

cgcgcggtga gcagggtgcc ctccggcgtg tgctcgggtc tgagcacctc gccctcgtcg 16200

tgcaccgcg ccaccaggct cccctcgtcg taggggatca cgacctccac ctccacctcg 16260

gggtgcggca gcaggcggtc gatcagttcc cgcagctcgt cgatgccccg acccgaacgg 16320

gcggacacga cgatcgcgtc cggtcctgc tccagcagac gggcgaggac gtccgggtcc 16380

gcgacatcgg ccttggtgac gaccacgacc tcggtggact cgccggcgcc cacgtcccgc 16440

agcacctcgc gcaccgaggc cagctgcgcg ccgggggtccg ggtgcgaacc gtcgaccacg 16500

tgcagcacca gatgcgcgtc cgcgacctct tcgatcgtgg aacggaacgc ctccaccagg 16560

tggtgcggga ggtggcgtag gaagcccacg gtgtcggcga tggtgtaggg gcgcccgctc 16620

ggcgtcgtcg cccgccgcac ggtcgtgtcc aggggtggcga acagggcggt ctccaccagc 16680

acgccggctc cggtgaggcg gttgagcagc gatgacttgc cggcgttggt gtagccggcg 16740

agggcgaccg acagcacctt gttgcgccgt cgctcctccc ggttcacgtc ccggccggtc 16800

ttcagctgct ccagctcccc gcggagcctg gccatcttgt cgttgatccg ccgccggtcc 16860

gtctcgatct tcgtctcacc gggaccgcgc gtggccatgc caccgccgcc accgccgcc 16920

atctgccggg acagcgactg gccccagccg cgcagccgcg gcagcatgta ctgcatctgc 16980

gccagcgcca cctgcgcctt gccctcccgg gactgggcgt gctgcgcgaa gatgtccagg 17040

atcagggccg tgcggtccac gaccttgacc ccgacgacct cctccagggtg catcagctga 17100

ctggggctca gttccccgtc gcacaccacg gtgtcggcgc cgtctcctc gacgatgtcg 17160

cgcagctgcg acgccttgcc cgagccgatg tacgtcgccg ggtccggctt ctgccggcgc 17220

tgcacgacgc cgtccagcac gagggcgccc gcggtctcgg cgagcgccgc cagctcggcg 17280

agcgaactgt cggcctccgc ggccgttccc gaggtccaga tgccgacgag caccacccgc 17340

tcgagtcgca gctggcggtg ctcgacctcg gagacgtcgg tcagttcggg ggagagcccc 17400

gcgacgcggc ggagggaggc ccggtcctcg cgggtctaact gatcgccgtc ccattcctcg 17460

gcgtcggcgc cggccatcgt gtcgccatc agtgcgtcgg ctcgctgcgt gccggcgaag 17520

tcctcgggat gcgtcaaagt acttccaatc tgggggggttc gagagcggtc cggggcgggg 17580

cgcgtccggc ctgcggggcc gggcgcgggg ccgggggggtc ggggccgacg tgtccgcacc 17640

tcgccgtcgg gggccggtgc cgtggaggtc gctgccggaa ctcacatgcc gtccacccta 17700

gccatgtcgc cggggtgtgc aaggacgtat cggcgatcac ctccccagct gacagagtgt 17760

gcaccgggta ctgcgccgc gggcgtgaag cgggcgagtg tggatgtcat gcgagtactc 17820

attatcgggg gttcacagtt cgtgggccgg gccttcgccg ccgaggcact ggccgcgggg 17880

caccgggtca ccacgttcaa ccggggtgtc agcggcaccg acctgcccgg cgtcgaggcg 17940

gtcagggggc accgcgaggt ggccggcgac ctggagcggc tgggtgtccgg aaggcactgg 18000

gacgcggtcg tggacacctg cggttacgtg ccccgcacgg tgggggcctc ggccgcggcg 18060

ctgtccgggc acgcggacac ctacctctac gtctccagca tcgcctgcct gcccgactgg 18120

gcgcaggcgg tccgtccggt ggacgacgac tcacctgcc acgactgcc gccggacgcc 18180

ggaccggacc acgccgacgg tgactacggc gtcctgaagg ccggctgcga gcgcgccgtg 18240

gaccggcact tcgcggggccg gaccctgcac ctgcggggccg gtgtcatcct cgggccgcac 18300

gacaccatgc gcatgctcga cgctgggctg tggcgcatgc gcgtcgccga gggggagcac 18360

cgccgggtgc tcgccccggg caaccccgag gtgggcatgc gcctgatcga cgtacgcgat 18420

gtcgccgtct tcggcctcga ctgcctcgcg gacggccgta ccggcgccctt catcgtcaac 18480

ccgccggaga agaacaccac cttcggggag ttgctcacgg agtgcgtaa ggccaccggt 18540

tcggccgcgg agccggtgtg ggtcgacgag gggttcctcg ccgagcacgg cgtgagtccg 18600

tggacggacc tgccgatgtg ggtgcccgcac accgcgcggg acaccctcgt gtgggcggcc 18660

ggagcaccgc gcgcccgggc cgcgggtctg gcctgccggc ccttctccga caccgtgcgg 18720

gacgcctggg aggtcgtccg ggaccggccc gtcccggaac tgccgctcgc ggccggctgc 18780

ggcctgtccc tgagccggga gaaggagctg ctgcgccct gggacgctcg cggcgggtgcg 18840

gcggcgggct gacgcggccc ccgaggagtgc gcacgccgtt tcccgctcgc ggggagcggc 18900

gctcgcgtgt ccgccccgcg gcccgggcgc gcgtcctgtg tcccggcgcc cggggcgcg 18960

ctacacggtg aaggtctcgg cgcggggccac gatctcgtcg accagccccg cctggcgag 19020

cgcggcgtag gactcggcg agacgtcctc ggcccggatc accgcacggc ggaagaccga 19080

caggatgttc acgaagtgcc ggtccacggg caggacgcgc tcctcccggg ggtcctggcg 19140

ggacagccgc agtacggggg ggtggctgtc cggggtcgtg aagacgtggc tcagggcgag 19200

cgagccgggtg ctgccgtgca gctcgtacgc cgagcggtag ccgtgttcca tgccgaaggc 19260

gaggtgggcg gccacgccct ggggggcggc gagcaggacg ctgccggaga ccacgacgcc 19320

ccggcggcg tcacggcgca gcacggcgcc ggtcaggcgc agttcgggtc cgaggaagtg 19380

cagcgcggcc cgcagcgggt agacgccgtt gtcgagcagt gcgccgccgc cgatgtcggg 19440

tcggtagcgc atgtcgtcgt cggcgcgcgg cgggatggtg aaggcggcg agaaggtgcg 19500

cagctctccg atggcaccgg cctccaacag ggctttgacg gtggcgtggt gagagtgggtg 19560

gaggaacatg aagttctcca tcagtaccag tccgcgctcc cgggccatgg cgaacaggcg 19620

ggccgcgtcg gcgtgggttg cggcggcggg cttctccacg aggacgtggt tgcccggccc 19680

cagcgccgcc gcggcccatt cggcgtgcag catgctgggc accgcgatgt agacggcgctc 19740

cacgtcgggg cgttccagca gggcctcgta gggcgcgacc gcctcgcagt cgaagtgcct 19800

gccaggggcc ttggcccggc ccgcgtcgcg gctgccgacg caggtcagca cggtgccggg 19860

ggtggagagc agggcgggca gggcgcgacg gccgcgatg tcgccgcagc cgatcgctcc 19920

gaagcgcagc accggggacg cttccggccg cggcggggcc ggtgcggcga acggttcagc 19980

agaagtaccc agggcgggtc tggatggttc cgtcaatcgg tccggatggt cacggcacag 20040

cgtggtgaac ggccggcgca caccacagag tcgcttccgg caatgggcgg ccctgaagct 20100

accgggcggc ccgacgcgac ggcaaggccg atctgacaca gtgccgcgcc gagggaccgt 20160

cggtgccccg cgtaccgct gacctgacac tacgccgacc aatctgtgga cggggcgggtg 20220

gccgcccggg ttaggtgagc gcggcgcccc cggtgaccgc gacgccggtc cgaagcggcc 20280

ctcggggcac gccgttgag agaggagacg cgggtgacgg acgcgatcac gaccgagctg 20340

gccgaccgcg aactggggcg cagactgcac cggatacgcg gcgtccactg gtatttcggc 20400

aaccacggtg acccgtagcg cctcatcctg cgcggtcaga ccgacgaccc gtcggtgtac 20460

gaggagcggg tccgcgaggg cgggccgctg ttccgcagcc gtaccgggac ctgggtgacc 20520

gcggacccgg aggtggccgc ggccgtgctg ggcgactcgc gcttcggtgc gctggaccgc 20580

gccggacggc gcccgagga gtacctcag ccgtcgcccg ccacgtacct ggggctggac 20640

cgcgcgcgt acgcgcgtct gcggcgggtg gccgagcccg tgctgggcgc ggacgccgcc 20700

gccgcgtggc gccggctcgg cgaggacgtc gggcgccggc tgctcgccgg ccgcggttcc 20760

ggcctcgacc tgacggcgga cttcgcccgc cggctgcggc cattggtcct ggccgcgtgg 20820

ctcgggggtgc cgggcgaacg gtgcgacgag tgggaggagt cgctgcgggc ggcggggccg 20880

ctgctggacg gtctgctgtg tccgcagacg ctggcgggca cccgtgcggc ggactcggcc 20940

gccgaggggc tgcgcgcgt gttggacgag gtggtcgccg cgcgccccgg cgggtccggc 21000

gaggggtcgg tggcccgcac ggtcggcgcc ggagccgccc ccgacgacgc ggtggccgcc 21060

gccgtgtgcc tggcgctctc ggccgtcgaa ccgacgacga ccctggtgtg cgaagcggtc 21120

cggctgctgc tcgaccgacc cgagtgggtg cggcggttgt gcgactcccc cgctctggcg 21180

ccggccgcgg tccggcacac cctgcggcac gcgcccccg tgcggtgga gagccgggtg 21240

gcccacgagg acgtgacggt ggcggatcgt ccgctgcccg ccgggagcca cgtggtggtg 21300

ctcgtgggcg cggcacggcg cgcgggcgcc ccggccgcgg agccggcgga cctggcgggc 21360

gcaccggcgg cggagctgcc ggacgacctg tggttcgcgc tgtccgggga gttcgtcggc 21420

cgtgccgccg agaccgcgct gggcgtgctg gccgaggccg ccccgggact gcggcgggac 21480

ggcgacatcg tccggcgggc ccgttccccg gtcctcggca ggtacgcgcg gttccccgtc 21540

gcgtactcct gacgggcccg cggccggcgt cccctcagtc ccccacgacg tttcatgaaa 21600

ggagtgccgt gcgcgtcctg gtgacctcca tcccgcacca cacgcactac taccacctgg 21660

taccgctgat ctgggctctg cgtgcctcgg ggcacgaggt ggtggcgggc ggccagccgt 21720

cgtggtcga cgccatcacc gccagcggca tcccggcggt cgccctggcc gaggaggagt 21780

cgtgggcgca gatcttcgag gaggtcgagg gcgatctcca gccgtatcag cacggcatcg 21840

acgagttcga cttcttcggc accctgaagg acgagctgga ctgggagaag ctgctcgccc 21900

agcaggtgat cctgtccggc ctgtggctgg aaccgctcaa cggcgccacg accctcgaca 21960

gcatcgtcga cttcgcccgg gcctggaagc ccgacctggt gctgtgggag ccgttcacct 22020

atgcggggcc ggtggcgggc cgggcgtgcg gggccgcgca cgcccgcgtc ctgtgggggc 22080

cggacacgat cgggctgctg cggacgaagt tccttcaggc ccaggcgcggt cagccccagg 22140

agcaccggga cgaccgggtc gcggagtgga tgacctgggc cctggcgcg c tacgggtgcg 22200

acttccggga ggaggacgtg ctcggtcagt ggagcgtgga cccgatggcg gagggcgtca 22260

gtctgggcct cgacctgccg accgtcccga tgcgctacac cccgtacaac gggtcggcg 22320

tgatccccga ctggctgacc gaggaaccga aacggcctcg ggtctgcctg accctggggg 22380

tgtcctcgcg ggagcacagt gaggacgagg tcccggtgca gaggtttatc gaggcgctgg 22440

ccgatctcga catcgagctg gtggcgaccc tggacgacgc ccagcgggac ctgctgccga 22500

ggatccccga caacacgcgc atcgctgact tcgtgcccac ggacgcgttg ctgccgacgt 22560

gctcggcgat catcaaccac agcggttcgg gcacgtgcaa caccgccgcg ctgcacgggg 22620

tgccgcagat catcctcggc ggcatacctgg acgccgccgt acggcagcac atgttcgcgc 22680

agaactccgc cgccctcacc ttcgctccgg aggaggtgac cggcgcgctc ctgaggagcg 22740

cgctgggtgcg cctgctcgag gagccgcggt tccgcgacgg cgcgcggcg ctgaaggagc 22800

ggatgcgggc catgcccagc ccggccggga tcgtgccgac cctggagcgc ctcacggccc 22860

agcaccgccg ggcgtgttga accggcgcg gcggccgtgc cggcgggtgac cgcccgaccc 22920

gactctcgcg tgtgatcgat ctcgtcgact cagccgtgga ccggttcgcc tgtccgcgcc 22980

cgacactgga gtgctgatgc gggccctctt cacgaccgcg ccgctcgcg gccacctgct 23040

tccgctggtg cccatcgcg tggccctgcg ggcggccggc caccaggtac tggtaggcgac 23100

ccgggaggac ttcgtgccg tgcacctgcg gtcggggctg ccgtccgcct cgtgcggggc 23160

gcccgcgcg gacctggcg gcgcggccga ggcggggcg ctcgcgcggc cccgcggagc 23220

ggcggaggct cggggggtcc tgagcggggc gctggcgcg ctcgcccggg gcagtctggc 23280

gggggtgcg cggctggcg acgcctggcg gccgatctg atcgtcagcg aacgggccga 23340

gttcgcccgg ccgctggtcg cggcggccct cggggtcccg tgggtccgct accactggtc 23400

ggtctcgtcc ctggaggagt accggcgagc ggccgaggcc gagttcgcg ccgagctggc 23460

ggcgctcggc ctcgaccggt tcccggaggc ggcgcgcgtg ctcgatccgt ggccggtgtc 23520

gctgcgccg ccggacgcg tgcgccacga cggggtcccg cactaccgg cccacgggga 23580

cgcccccgtc cccgactggg cgttcacgcg cggtcgcggg ccgcggatct gcgtgacgct 23640

cggcaccatg ctgccccggt acggcgccgc cgggatggcc gacttcctga cggagctggt 23700

ggcggagacc cgcggagggg actgcgaact gctcgtggcg gtcgacgacg acgtcgtcgc 23760

gcggtggccg tcgctgccct ccgcggtgcg gtacgccggc cggctgccgc tggcggaggc 23820

gctgcccgcg tgcgacgcg tggtagacca cggcgggcag ggcacgtccc tgaccgcgct 23880

ggccgcgggt cggccgcagg tcgtcatggc gcggctcgac gaccagttcg acaacgcgcg 23940

ggcactggcg gcggcggggg cgccctgct cgtaccgcg tcccgggcca ctccgcggc 24000

cgtggccgcg gggcgccgc aagtgcgga gaacgcctg tatgccaagg cggcagccg 24060

gctcgccgag gagatggcg tgctgccgc gccgtcggcg gcggtcggac tcctggaaca 24120

cccggggccc gggccggaca tgccgcggag ttaccggaac gaggatgcg tgtgacgtga 24180

atctggaagt actcaaccgt tcgaacgatc cgcgcgggcc ggtgatcac gtggtcggcg 24240

cgtccggctt catcgggtcc gccctggtcg ccgagctggc gcgcatgccg gtgcggctgc 24300

gggcggtggc ccggcgcgag acccccgctc ccgcgggggc acgggcccgc gtcgaggtcc 24360

gccgggcgga cctcgcccgc ccggacgagg tcggggccgc cgtcgagggg gcggacgccg 24420

tcgtgcacct cgccgcccac atcggcgggc gcgggtcgtg gcgcgcggcc gacgagcgg 24480

cgctgcgggt gaacgtcggc ctgctgcgc acgtggccga cgcgttcggg gaccgctcgg 24540

ggcccgcccc ggcgtggtc ctggccagta ccctccaggc cggcgtcgag ctgtcccggc 24600

agggcccgta cgcccggcag aagtcggcgg ccgaggaggc cctgctgcgg gccgcctccg 24660

aggaggtggc ccgcggcgtc gtgctgcggc tgccgaccgt ctacggggcg agcccgctga 24720

ccgggtggac gggccgcggg gtggtcgcgt cgggtggcac gcaggccgtc tcgggcgagc 24780

cggtcacgat gtggcacgac ggcacggtcg ggcgcgatct gctccacgtg gaggacgcgg 24840
cccgcgccctt cgcggcggcg ctcggtcacg tggagcggct ggacggcggc acgtgggtccg 24900
tcggtacggg ccggctggag cccttgggag aggtgttctc ggccctcgcc gggctggtgg 24960
ccgagcggac ggggaggccc ccgtaccgg tggctctccac ggagccgccc gaccatgccg 25020
aggcgggcga cttcgagagc gcggtctgtg acccctccgc gttccgcgcg gtgaccgggt 25080
ggtctccctt cgttccgttg cgggcggggc tcggcgccgt ggtggagacg atggtggccg 25140
acggagcgcg ggggtgggatc cgaacgtgag cacggaccgg gagcaggccg cgcacacgcg 25200
gctcggtcgc agcgcgaccc tggtgagccg gctctggctg ggcaccgtga acttcagcgg 25260
ccgggtcgag gacggtgacg cgatgcagct gatggaggcg gcggtcgacc gcggcatcaa 25320
ctgcatcgac accgcggaca tctacggctg gcggatccac aagggccaca ccgaggaact 25380
ggtgggcccgg tggctggcca agagcgccgc gcggcgggag gacgtcctgc tggccaccaa 25440
ggtcggcggg gacatgagcg aacggctcaa cgacggcggc ctgtcggcgc ggcacatcgt 25500
cacggcctgc gagcagtcgc tgcggcgccct gggcgtggac cacatcgacc tgtaccagat 25560
gcaccgcgtc gaccacgccg cgccgtggga cgagatctgg caggcgatgg accgtctggt 25620
ggcgagcggc aaggtgacct acgtggggtc gtcgaacttc gccggctgga acgtcgccgc 25680

cgcgccaggac gcggcccggc ggcgccagtc cctcggtctg gtgtccgagc agtgccctgta 25740

caacctggcg gtgcgccacg ccgagctgga actgctgccg gccgcccagg cgtacggact 25800

gggcgtgttc gcctggctgc cgctgcacgg cgggctgctc agcgggggtgc tgcgcaagct 25860

cgcgggcggc gtcgcggtga agtcggcaca ggggcgggcc cagctgctgc tgcccagact 25920

gcgcgcgacg atcgaggcgt acgaggggtt ctgcggccgg atcggcgcg atccggccga 25980

ggtcggtctg gcctgggtgc tgtcccggcc ggggatcagc ggcgcggtga tcggtccgcg 26040

cacggtggac cagctggact cggcgctgcg gtccctggac ctggtcctcg gggaggccga 26100

actggccgag ctggacgcca tcttcccgcc cctgggcaag ggcggccggg cgccggacgc 26160

gtggatcagc tgaagggggt gcatcggccg acgtcacgcc ggccgatgcc gccggtcaca 26220

cgacgtcgag cgcgggcagc gggaagacca gtcggccgcc gccgtcgagg aactcccgt 26280

cccgttcgac gaaccggtcc cggtagatcc agggcaggac cagcaactgg tccggcttct 26340

gcgccttcgc gtctcctcg gacacgatgg ggatgcccg cccgggggtg aaacgccccg 26400

ccttctcctc gctcacctcg ccgatgcacg gcaggctccc ttcggtgatc ccgcagtact 26460

ggaggatgac gttgcccttg gtggaggcgc cgtacccgag ggtcagcagg ccctcttggc 26520

gggagcggtc caggaagccg cgcagggcgt cccgctggtc ggcgacgcgg cgggcgaagg 26580

cctcgaacgg tgccatgccg tccagtcccc cggcggcctc ggccggcccgg atgcggggcca 26640

ggcccgcctc gtccctcggg tgccgggaac cggtcctggc gagcgtgacg cacaggctgc 26700

cgccgtacac ctcggtgagc tcggcccggg tgacggtgag gccgacgcgt tccgccatcc 26760

actcgatctg gcgcagcgcg tagtactcca ggtgctcgtg gcagacgatg tcgtacgcgt 26820

cggcttcgag catggcgggc aggtagctct gctccatcat ccagacgccg tcctcggcga 26880

ggacgtcgcg gacgtcgctc atgaagcgca gcgggtccgg caggctcgtag aacatcgcg 26940

tggaggtgac ggccttcgcc cgccgcgccc cgaagcgggt ctcgaaggtc gcgcgggtga 27000

agtagtccac gaccaactcg gcgttcggcg ggtacaggtc gcggaacttg ccgccggtcg 27060

ggtcgatgcc gaccagtcgc gggccgtcgg cggggtagcc cttgaggagc gtggcgtcgt 27120

tgctgccgat gtcgaggacc aggtcgtccg ggccgaggtc caccagccgg cggacggcgg 27180

cgaccttgcc atgcagggtg tcgaccatga agggccggat gcccgagcgg tagccgtagc 27240

cctcgccgta catgagggtc gggtcggggg tgtggcgagc ttgcacgagg ccgcatccgg 27300

ccggggaaca cgcgacgagt tccagcggga ccgacggcac gacctcgtcg cggtcggccg 27360

ggaacacccc ggtgagcgcc tgttcgcccga ggtcgagtac ggagagcagc tccttggtgc 27420

cgcagacgcg gcacgcggtg gcaatcatgg ggtcctttcg ggatcctggc cggggcgccg 27480

ccggcccgtg gccaggtcgg gagcttccag acggaggggtg tcggggggcg gccgggggga 27540

cggcaccagg tgcaggcggc cgatcgcgcc ggtggcggcg gaggcgtcct ggcacgggca 27600

gggatcgtcg gtgaatccgg cgaagcggta ggccacttcc atcatccggt tgcgttcggt 27660

ggcccggaa tgggtccga ggtgcacccc ggcccgggtgc gcctgggtccg tcagccagcg 27720

caggatcacc gtgccggccc ccagcgacac caccggcac gaggtggcca gcagcttgat 27780

ccgccaggcc tcggggccgc gccgcagcag cacgacgccg accgcgccgt agggaccgaa 27840

gcgatcgggtg acggtgggtga ccagcacctc gtggtcgggg tcgtcgatga gggcgcgcag 27900

ctcgtcctcg gagtagtgca ctccggtggc gttcatctgg ctggtgcgca gggtcagttc 27960

ctcgacccgc gacagctcca ggggcgtggc gcggctgatc cgcattccgga tgtccagcga 28020

gcgcaggaag tccgcgtcgg gtccggtgaa gtcggaccgc tcggcgtccc ggcggaagga 28080

cgcctggtac atggagcggc ggcgggtcga gtcgacgggtg acggtgcccc ggctgaactc 28140

cgggaggccg gtcagccggg tcgcctgctc ggcggcgtag gtgcggacct ccggcagttc 28200

gtgggtgacc tcggcccgtt cgaagggctg gtcgtcgatg aaggcgaggg tgctcggcgc 28260

gaagttcagc cggtcggcga tctcgcgga cgaactcgac ttccggcccc agccgatcct 28320

cgaggagcacg aagtactcgg cgacgccgag ctgttcgagc ttcgcccagg cgtgatcgtg 28380

gtcgttcttg ctggccaccg cctgcaggat gccgcgcgcg tccagctcgg tgatgggtccg 28440

caggacgtcc ggggcgagcc ggacctcgtc ctctcagagg aggggtgccct gccagaggggt 28500

gttggtccagg tcccagacca ggcatttgac caacggctcg gcggcgttgt cccgggtgcac 28560

gtcgttcctt tctctcgata cggcggttgc gaccgcgact cctcgacagg gcgtacgcgg 28620

cccgcggccg gcggggacac gggccggttc acccggccgc ggacatcacg tgccggggcca 28680

ggaccagttc gcagatctcg ttgctgccct cgatgatctc catgagcttg gcgtcgcggg 28740

gcgcccgtgc gacgacgtgt ccctcccgcg ccccggccga cgccagcacc tgcacggccc 28800

gttcggcccc gcgtgcggcg ccggtggccg cgacgtgctt ggccaggacg gcggcgacca 28860

ccatgtcggg gctgccctcg tcccactggg cgctggcgtg ctcgcacgcc cgggcggcgt 28920

gctgttcggc gacgaacagt tcggcgagggt gccgggacgac gagctgggtg tccgagagcc 28980

gcgtgccgaa ctgctcccgt ccgccggcgt gacgtgcggc ggccggccagg caggcgcgca 29040

ggatgcccag ggagccccag gccaccgaca tccggccgta gctcagcgcg gtggtgacca 29100

gcagggcggg ggtgcggtcg tggccctgga gcagggcgtc tgccggcagc cggacccccgt 29160

ccaggtggat gtcggcgtgt ccggccgcgc ggcagccgtg cgcgtcgggtg atgcgctcga 29220

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __3__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __3__

NOTE: For additional volumes please contact the Canadian Patent Office.

REVENDICATIONS

1. Polynucléotide codant un polypeptide impliqué dans la biosynthèse des spiramycines caractérisé en ce que la séquence dudit polynucléotide est:

- (a) l'une des séquences SEQ ID N° 111 ou 141,
- (b) l'une des séquences dérivées de la séquence en (a) en raison de la dégénérescence du code génétique, ou
- (c) un polynucléotide présentant au moins 80%, 90%, 95% ou 98% d'identité en nucléotides avec un polynucléotide de la séquence en (a).

2. Polynucléotide selon la revendication 1, caractérisé en ce qu'il est isolé à partir d'une bactérie du genre *Streptomyces*.

3. Polypeptide résultant de l'expression d'un polynucléotide selon la revendication 1 ou 2.

4. Polypeptide selon la revendication 3, ledit polypeptide étant impliqué dans la biosynthèse des spiramycines.

5. Polypeptide selon la revendication 3 ou 4, ledit polypeptide ayant une séquence comprenant la séquence SEQ ID N° 142.

6. Polypeptide selon la revendication 3 ou 4, ledit polypeptide ayant une séquence comprenant la séquence SEQ ID N° 112.

7. Vecteur d'expression caractérisé en ce qu'il comprend au moins une séquence d'acide nucléique codant un polypeptide selon l'une quelconque des revendications 3 à 6.

8. Cellule hôte comprenant un vecteur d'expression approprié permettant l'expression d'un ou plusieurs polypeptides selon l'une quelconque des revendications 3 à 6.

9. Procédé de production d'un polypeptide selon la revendication 3, 4, 5 ou 6 caractérisé en ce que ledit procédé comprend les étapes suivantes:

- a) insérer au moins un acide nucléique codant ledit polypeptide dans un vecteur approprié;
- b) cultiver, dans un milieu de culture approprié, une cellule hôte préalablement transformée ou transfectée avec le vecteur de l'étape a);
- c) récupérer le milieu de culture conditionné ou un extrait cellulaire;

- d) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape c), ledit polypeptide ; et
- e) le cas échéant, caractériser le polypeptide recombinant produit.

10. Bactérie du genre *Streptomyces* productrice de spiramycine caractérisée en ce qu'elle possède une suppression, une substitution, une délétion et/ou une addition d'une ou plusieurs bases dans au moins un gène comprenant une des séquences SEQ ID N° 111 ou 141 ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique et/ou qu'elle surexprime au moins un gène comprenant une des séquences SEQ ID N° 111 ou 141 ou l'une des séquences dérivées de celles-ci en raison de la dégénérescence du code génétique, ledit au moins un gène possédant une suppression, une substitution, une délétion et/ou une addition d'une ou plusieurs bases et/ou étant surexprimé codant pour un polypeptide impliqué dans la biosynthèse des spiramycines.

11. Bactérie du genre *Streptomyces* selon la revendication 10 caractérisée en ce que la surexpression du gène considéré est obtenue en augmentant le nombre de copie de ce gène et/ou en plaçant un promoteur plus actif que le promoteur sauvage.

12. Procédé de production de spiramycine(s), caractérisé en ce qu'il comprend les étapes suivantes:

- (a) cultiver, dans un milieu de culture approprié, une bactérie selon l'une quelconque des revendications 10 à 11;
- (b) récupérer le milieu de culture conditionné ou un extrait cellulaire; et
- (c) séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), la ou lesdites spiramycines produites.

13. Utilisation d'une séquence selon l'une quelconque des revendications 1, et 2, d'un vecteur selon la revendication 7 ou d'une cellule hôte selon la revendication 8 pour la préparation d'antibiotiques hybrides.

14. Souche de *Streptomyces* selon l'une quelconque des revendications 10 à 11 caractérisée en ce qu'il s'agit de la souche OSC2/pSPM75(1) ou de la souche OSC2/pSPM75(2) déposée auprès de la Collection Nationale de Cultures de Microorganismes (CNCM) Institut Pasteur, 25, rue du Docteur Roux 75724 Paris Cedex 15, France, le 6 octobre 2003 sous le numéro d'enregistrement 1-3101.

15. ADN recombinant caractérisé en ce qu'il comprend :

(a) un polynucléotide codant pour un polypeptide impliqué dans la biosynthèse des spiramycines susceptible d'être obtenu par amplification en chaîne par polymérase en utilisant le Couple d'amorces de séquence suivante :

5' AAGCTTGTGTGCCCCGGTGTACCTGGGGAGC 3' (SEQ ID N° 138) et

5' GGATCCCGCGACGGACACGACCGCCGCGCA 3' (SEQ ID N° 139) et comme matrice le cosmide pSPM36 ou l'ADN total de *Streptomyces ambofaciens*, ; ou

(b) un fragment d'au moins 1200, 1400, 1450, ou 1500 nucléotides consécutifs de ce polynucléotide codant pour un polypeptide impliqué dans la biosynthèse des spiramycines.

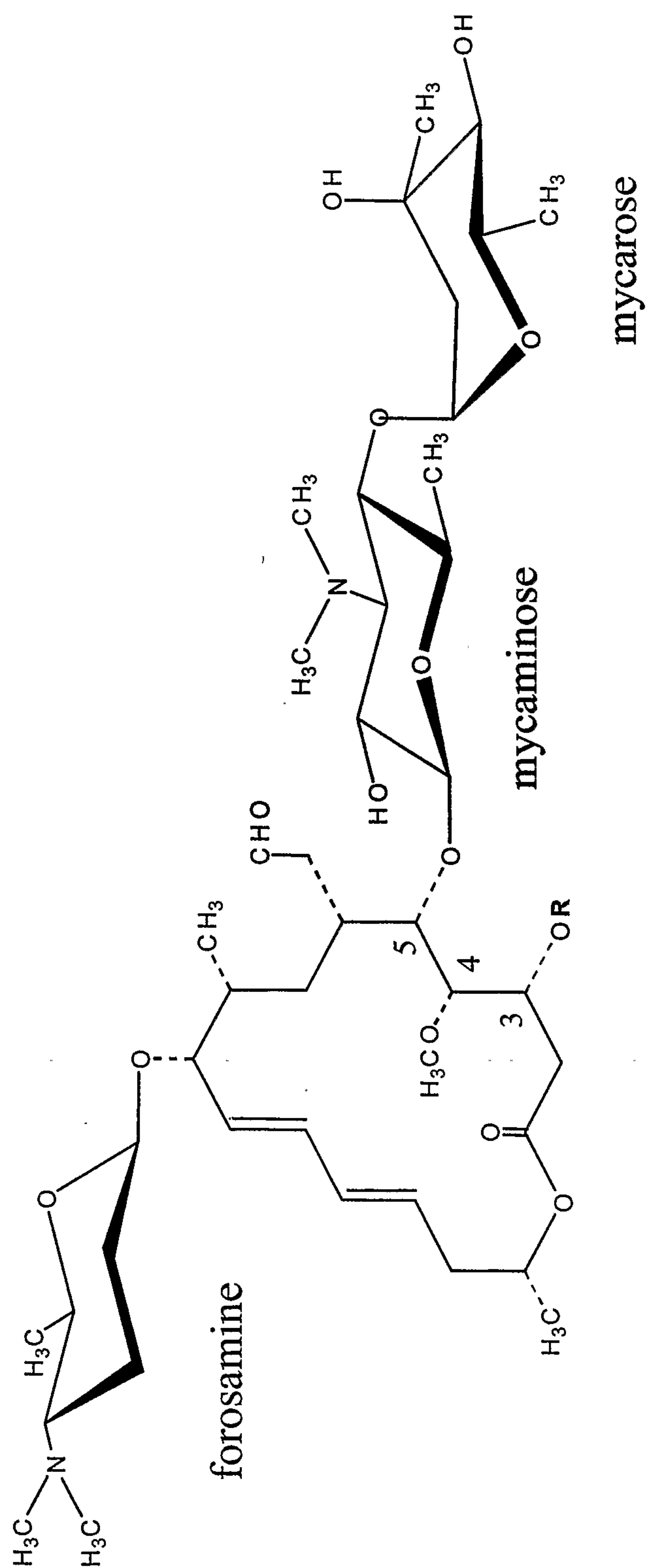
16. Fragment selon la revendication 15(b), ledit fragment étant efficace pour amplifier les séquences définies dans la revendication 1 ou 2.

17. Cellule hôte dans laquelle a été introduit au moins un ADN recombinant tel que défini dans la revendication 15 ou la revendication 16.

18. Procédé de production de spiramycine(s), caractérisé en ce qu'il comprend les étapes suivantes:

- a. cultiver, dans un milieu de culture approprié, une cellule hôte selon la revendication 17;
- b. récupérer le milieu de culture conditionné ou un extrait cellulaire; et
- c. séparer et purifier à partir dudit milieu de culture ou encore à partir de l'extrait cellulaire obtenu à l'étape b), les spiramycines.

1/41



R=H spiramycine I

R=COCH₃ spiramycine II

R=COCH₂CH₃ spiramycine III

FIG. 1

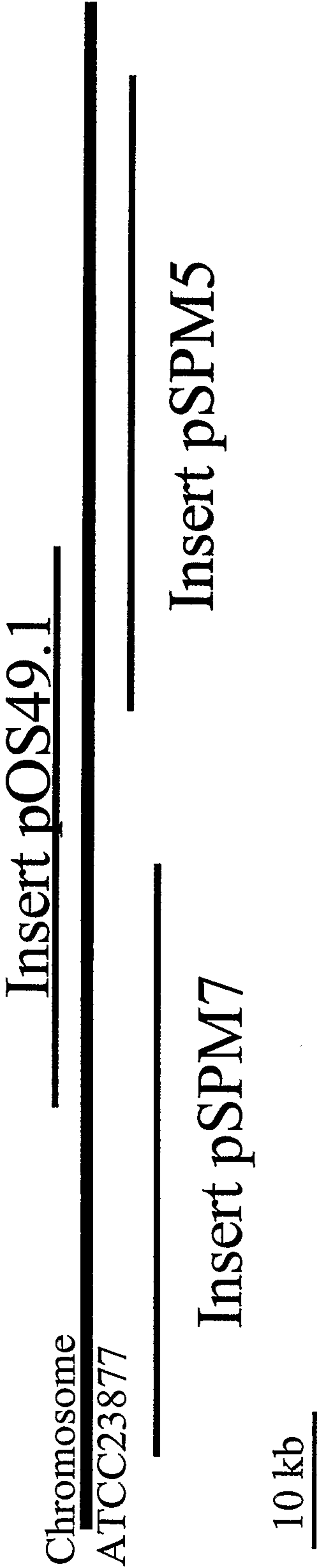


FIG. 2

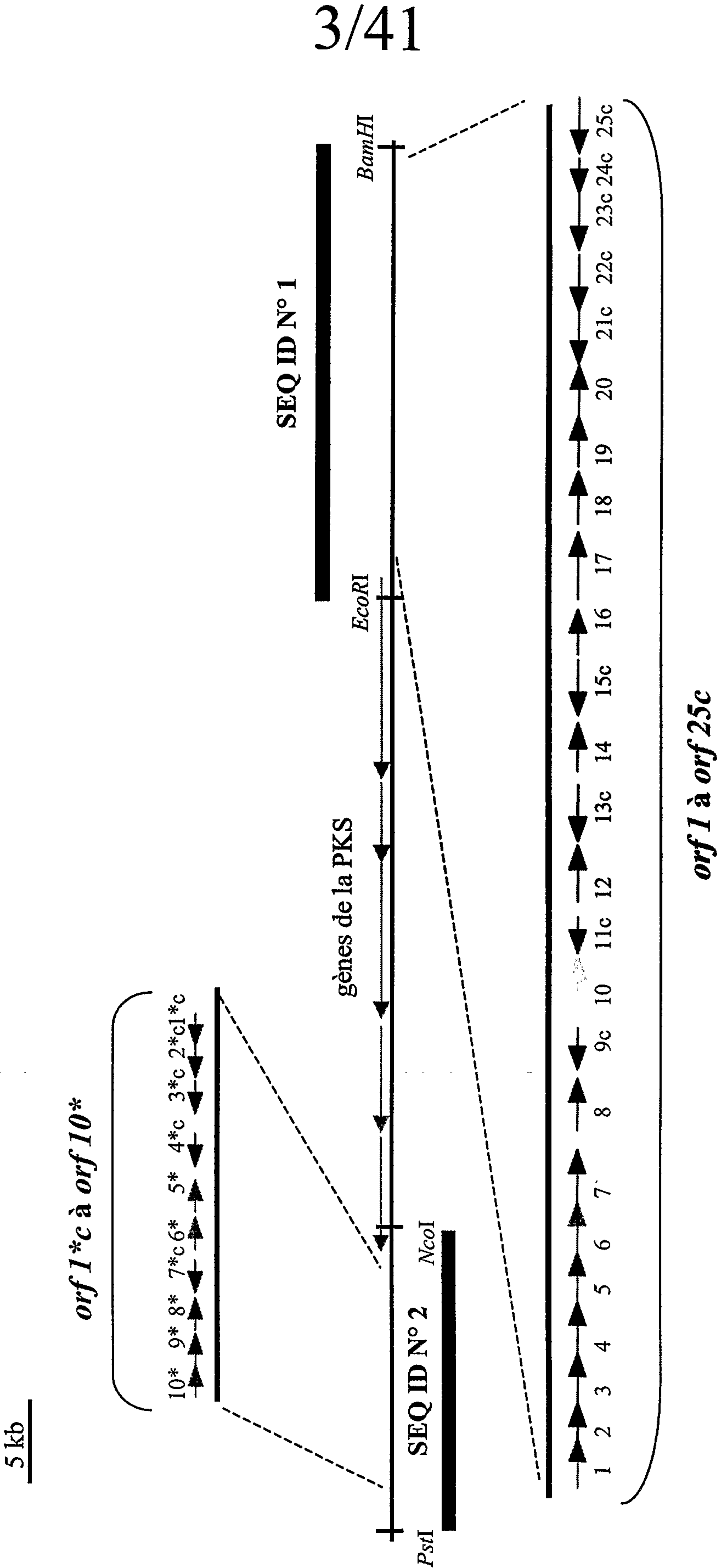


FIG. 3

4/41

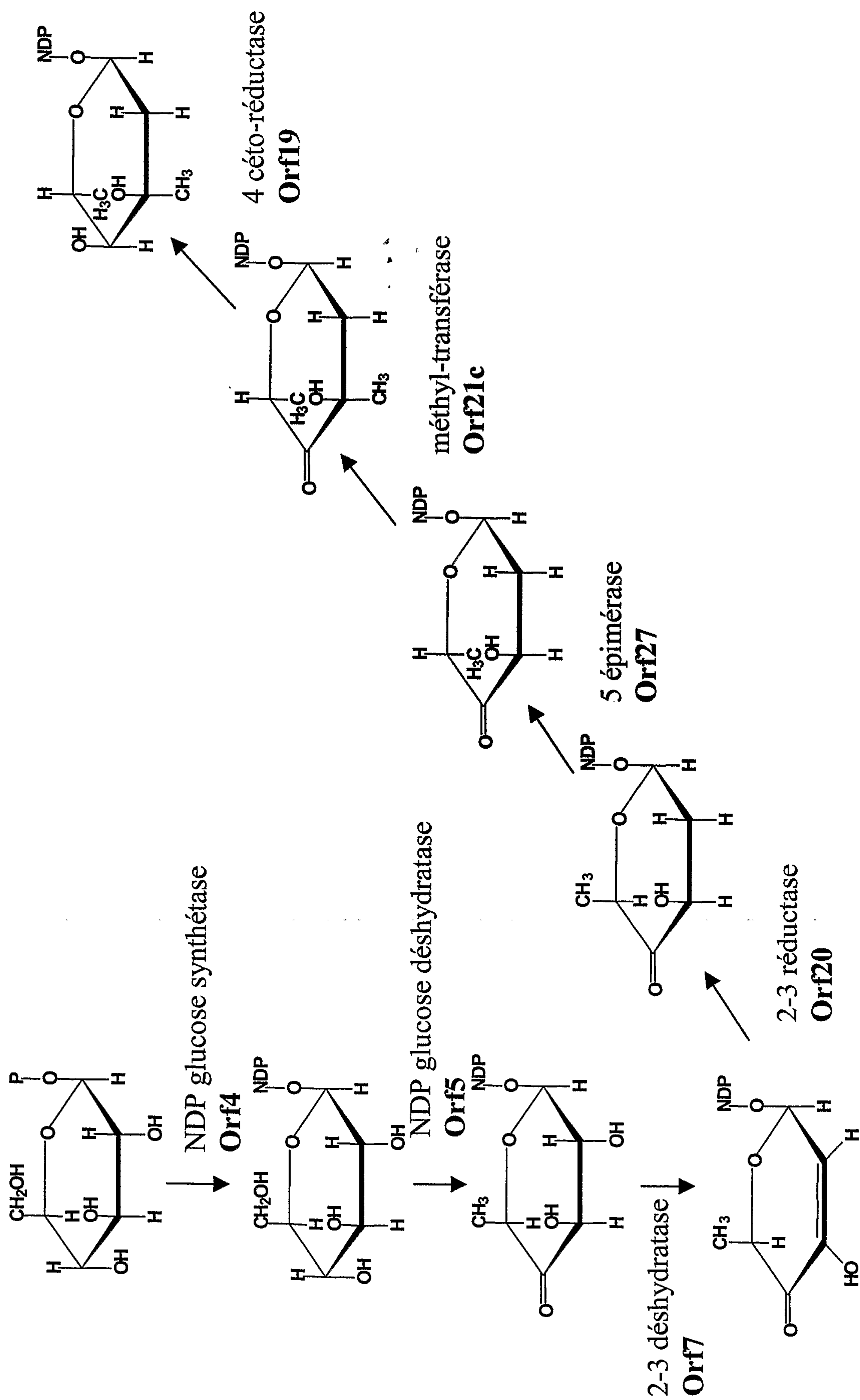


FIG. 4

5/41

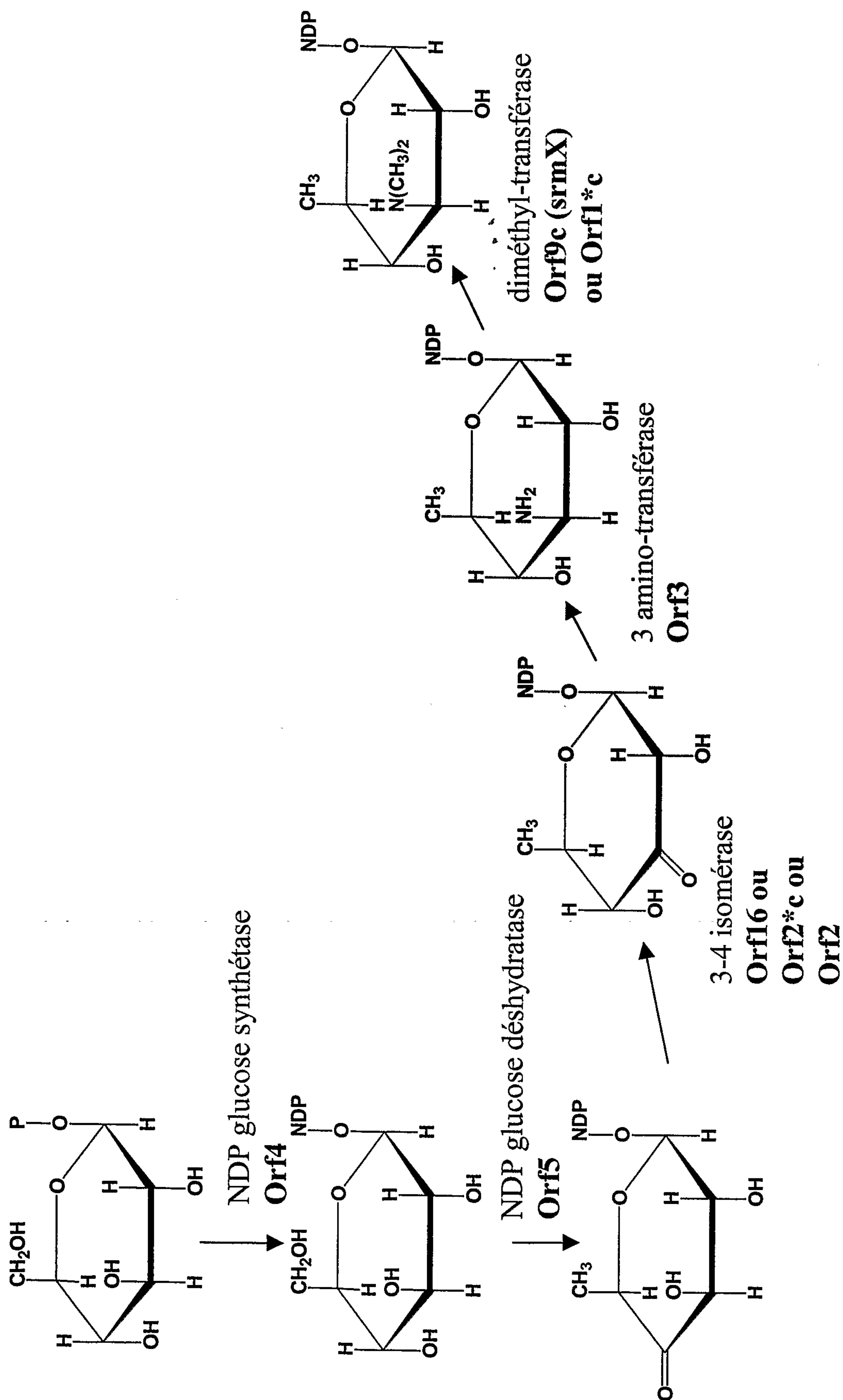


FIG. 5

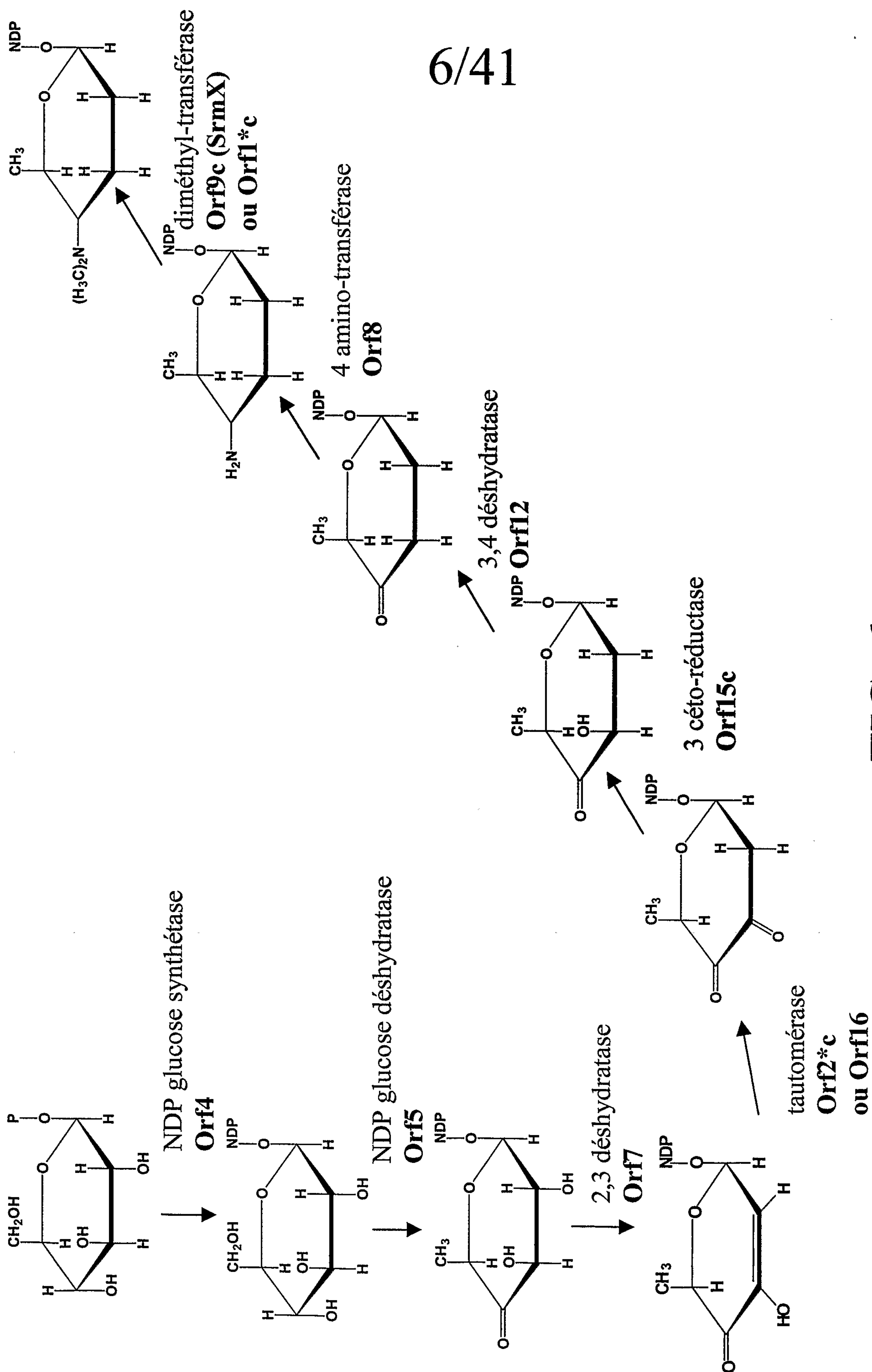


FIG. 6

7/41

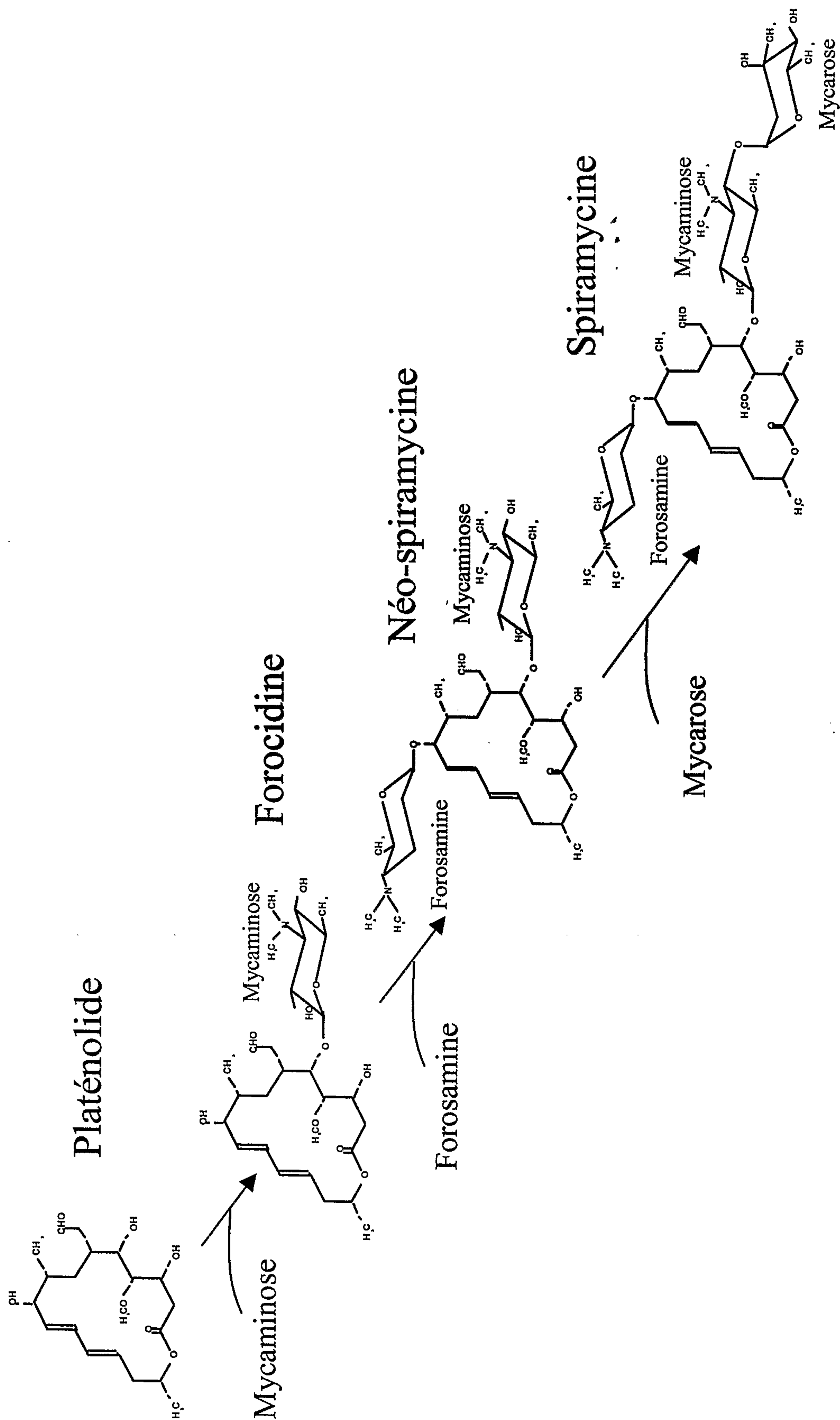


FIG. 7

8/41

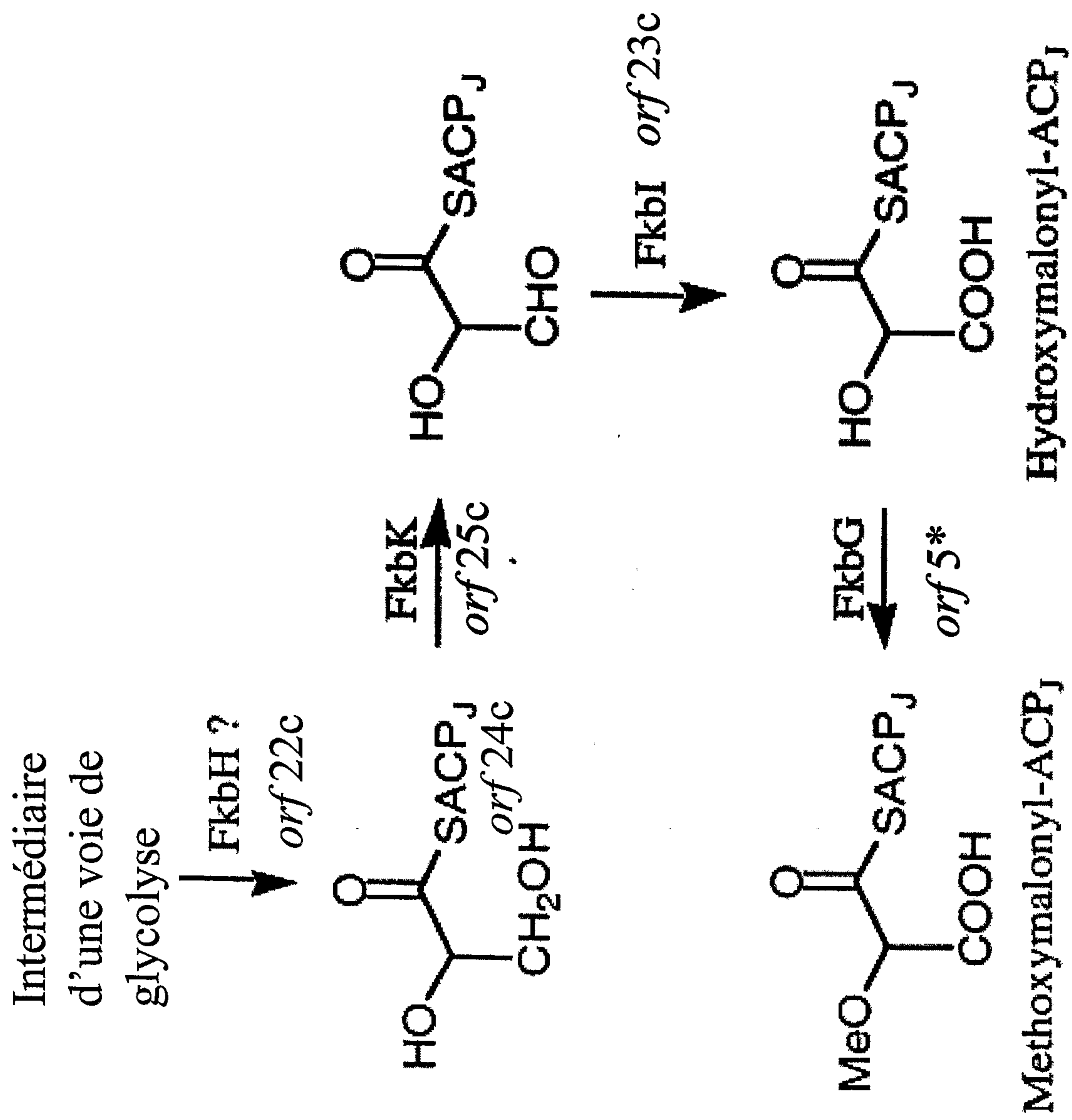


FIG. 8

9/41

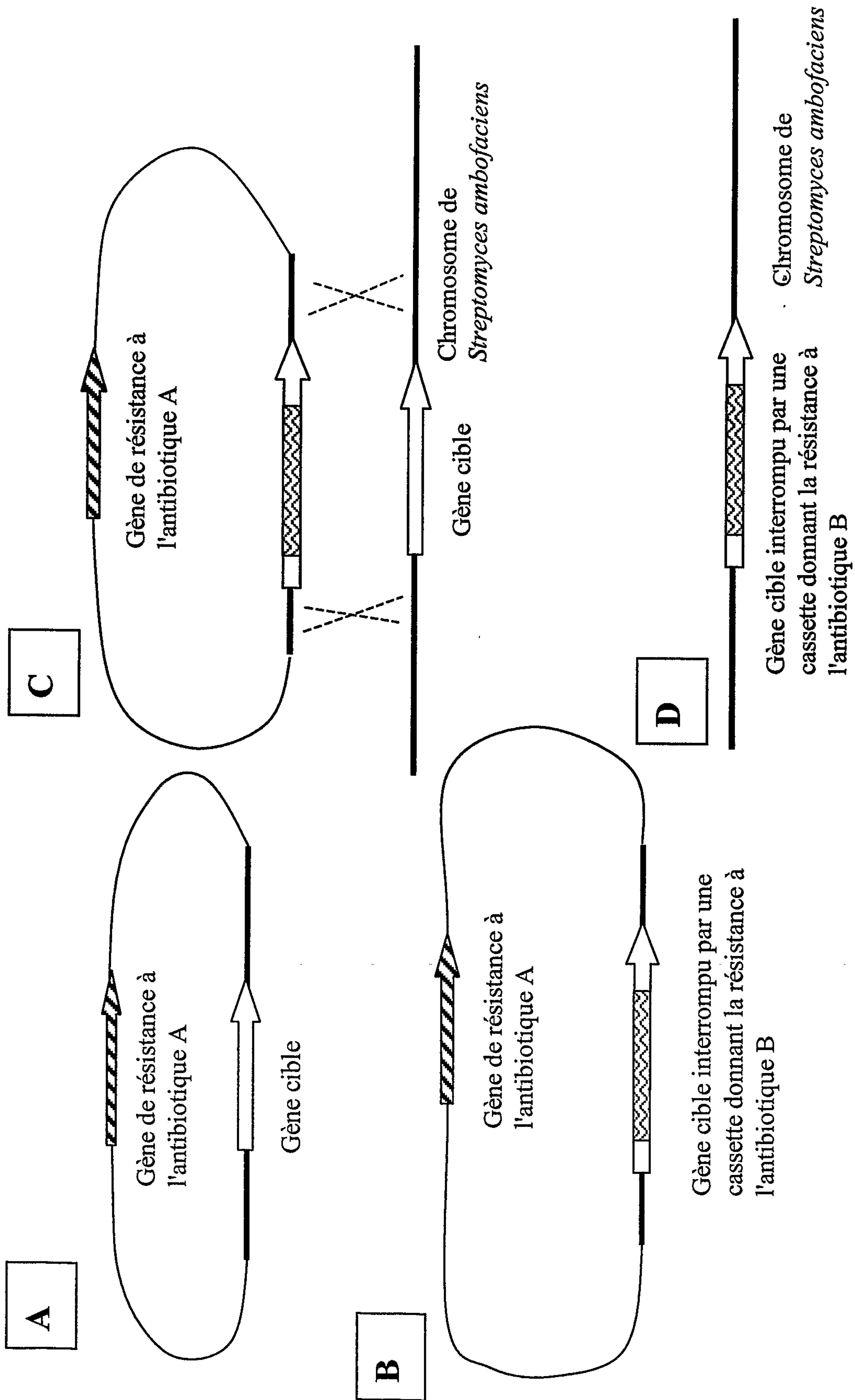


FIG. 9

10/41

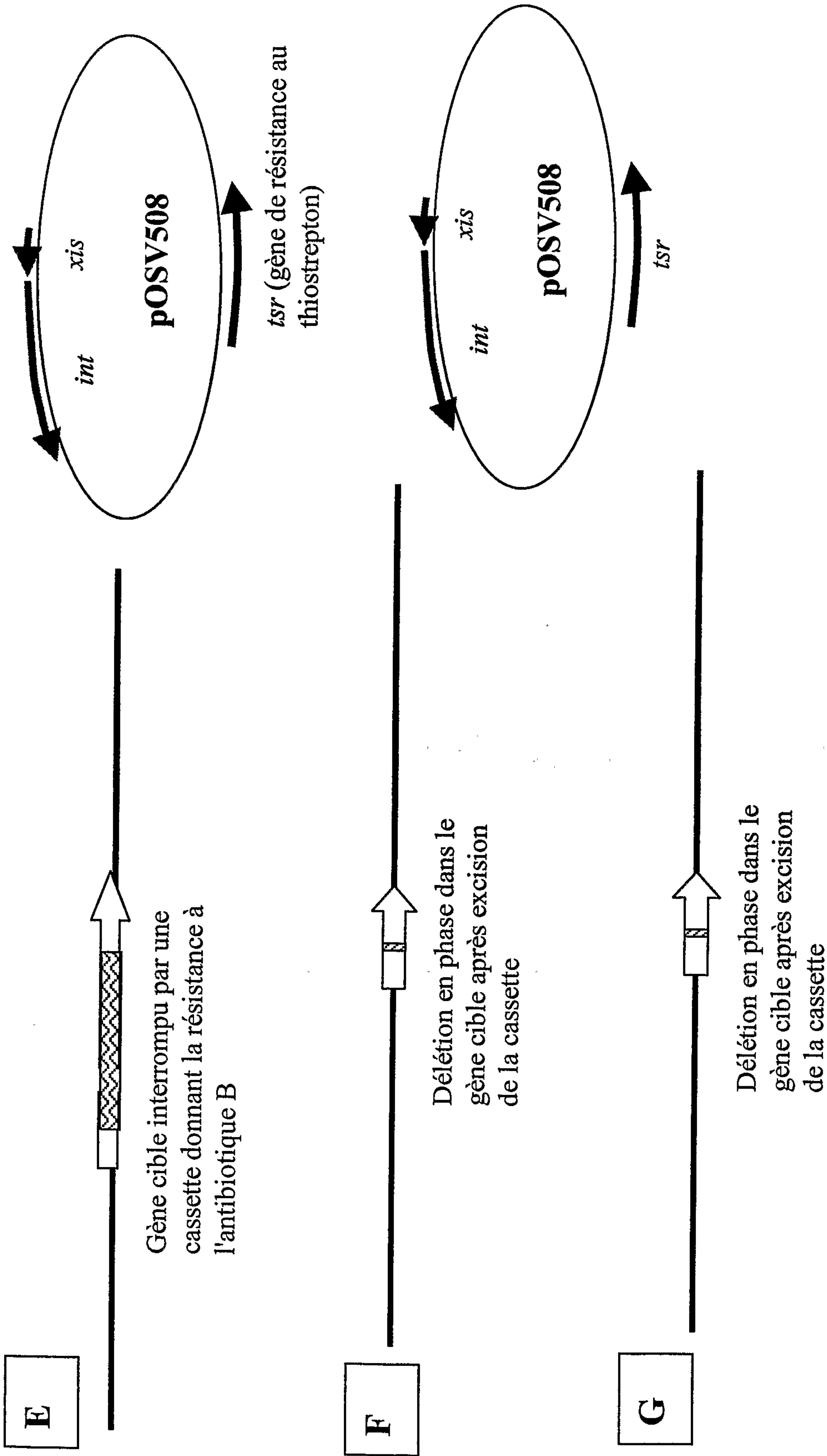


FIG. 10

11/41

Amplification de la cassette excisable

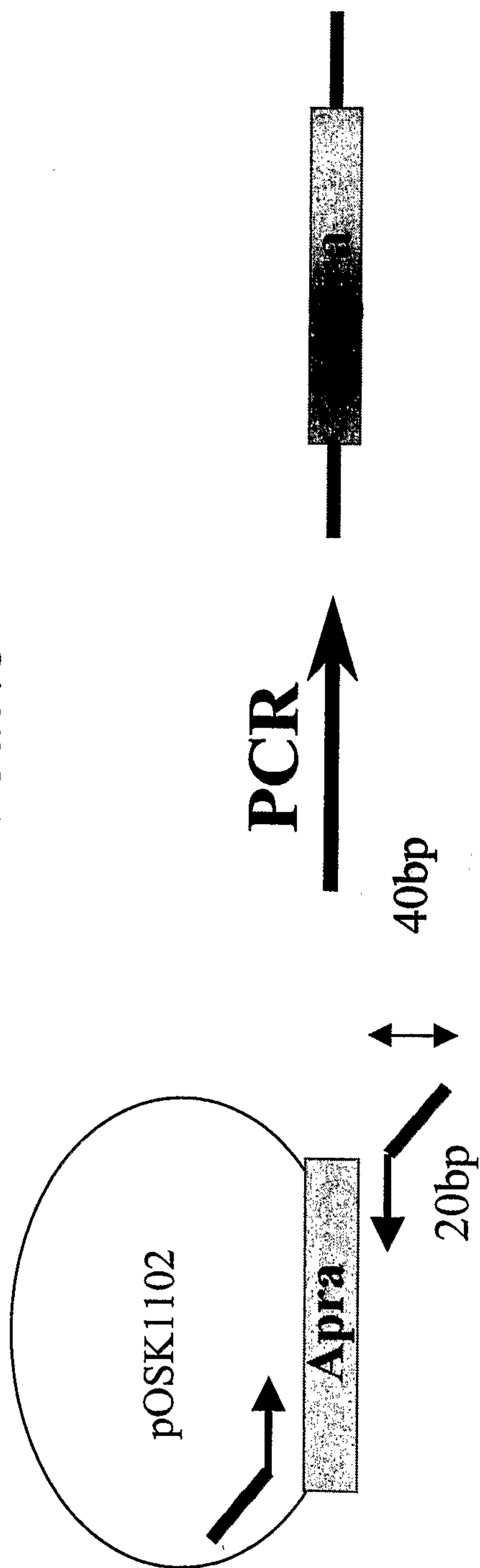


FIG. 11

12/41

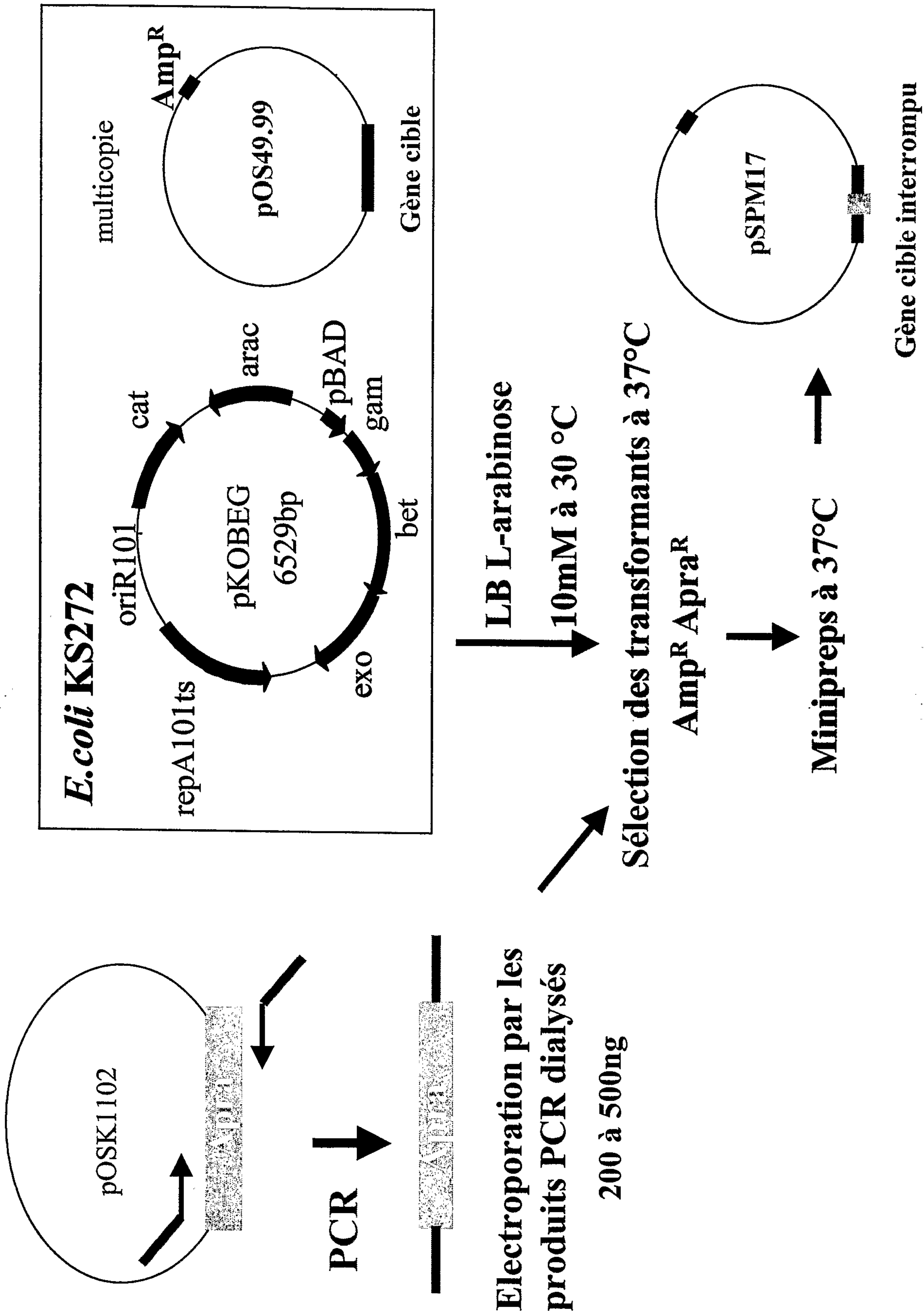


FIG. 12

13/41

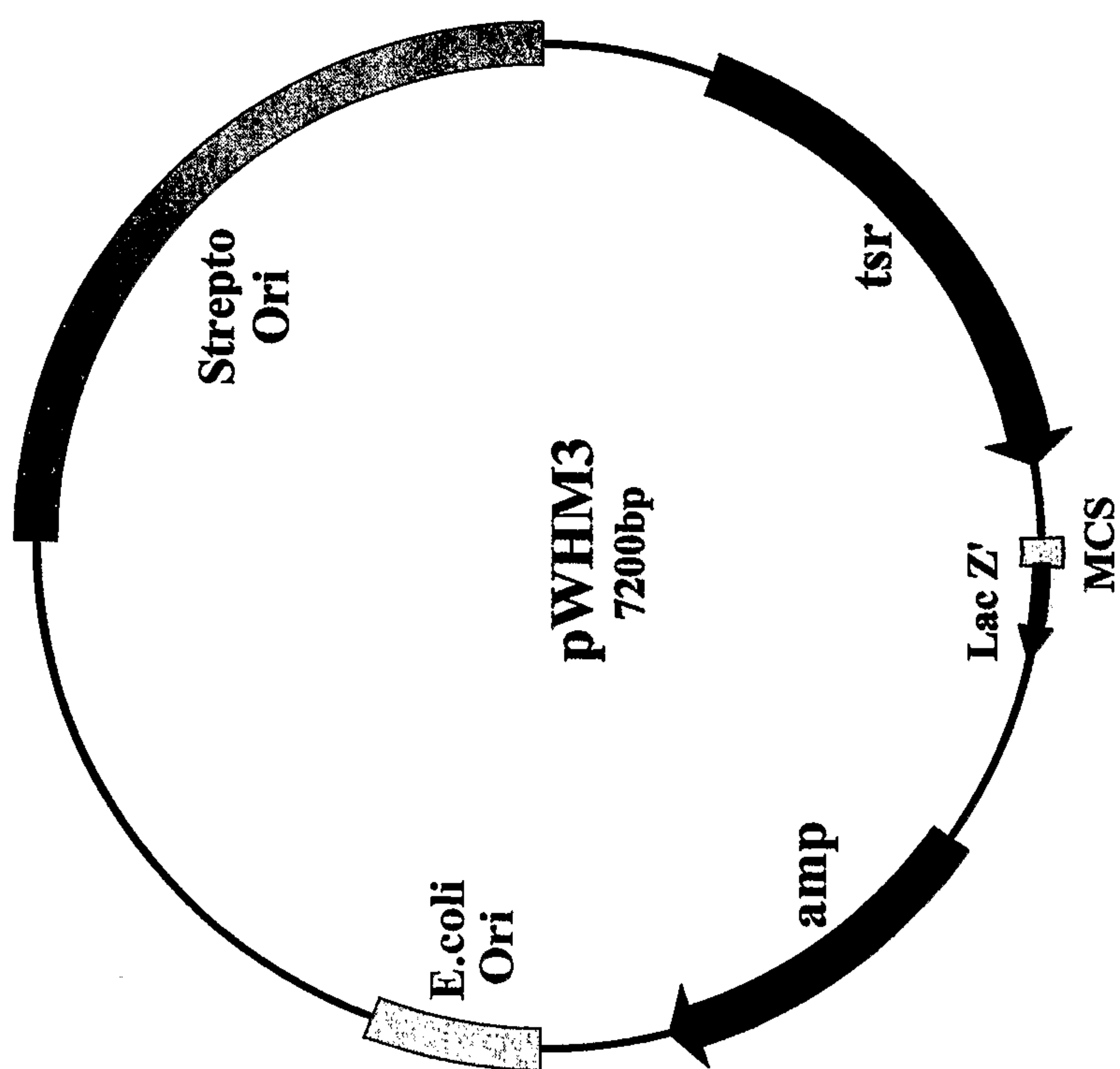


FIG. 13

14/41

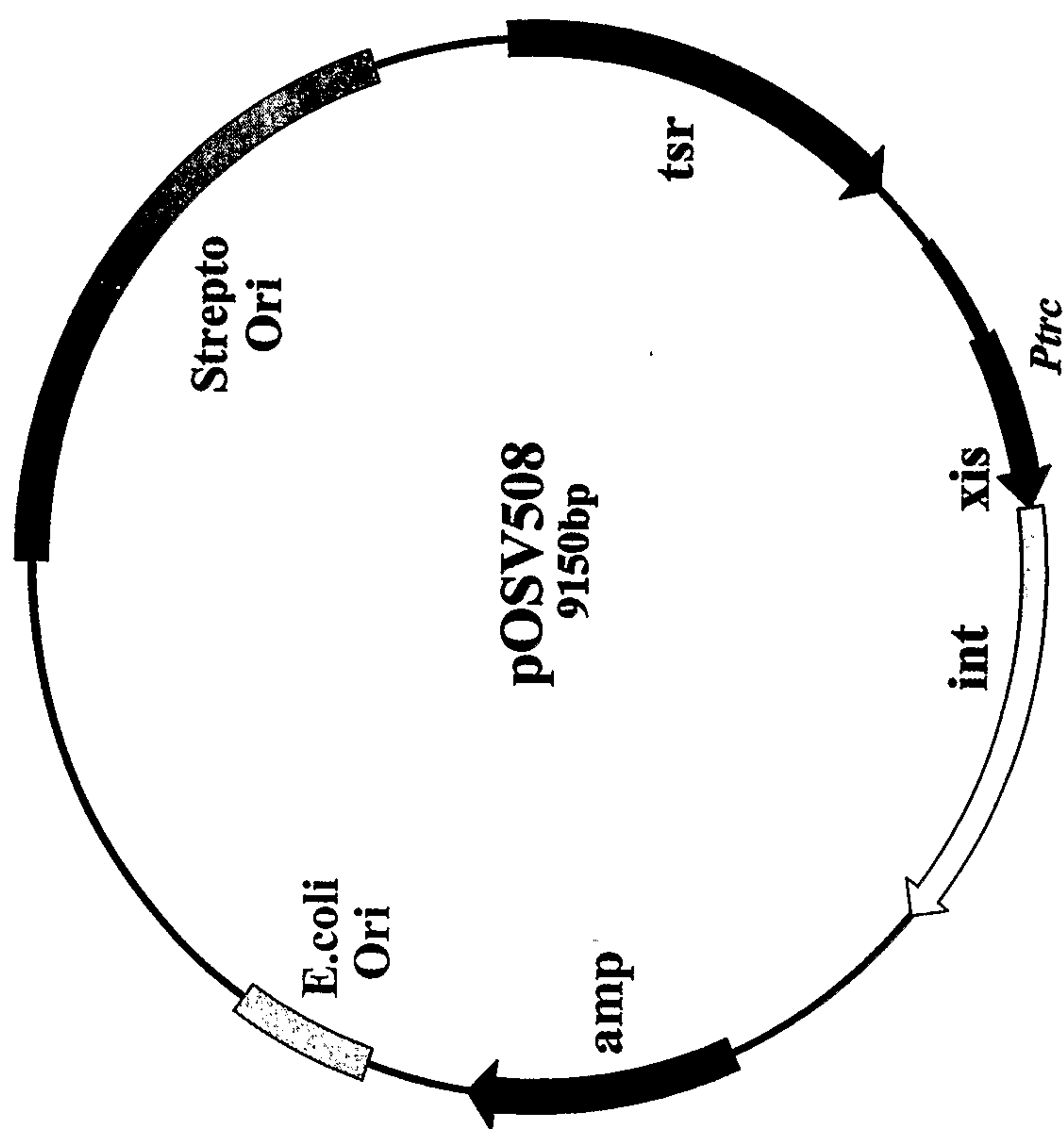


FIG. 14

15/41

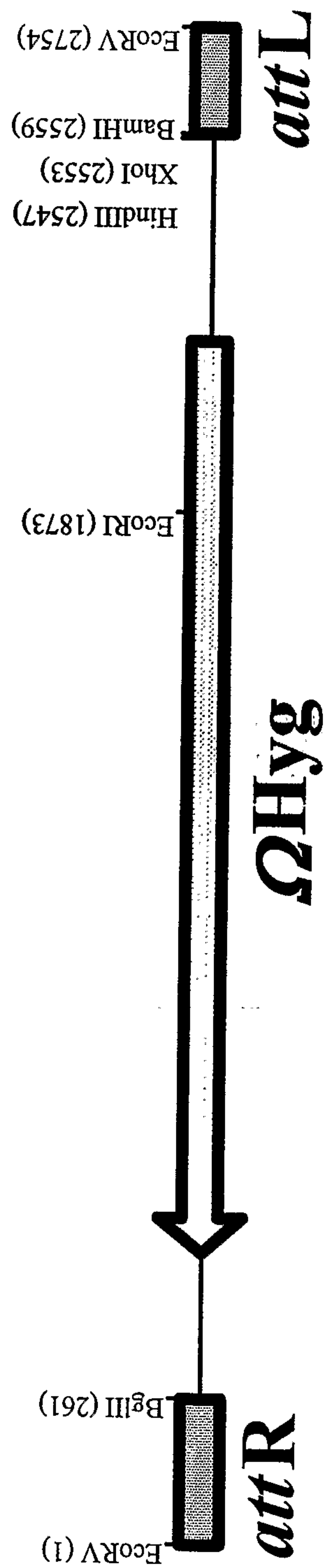


FIG. 15

16/41

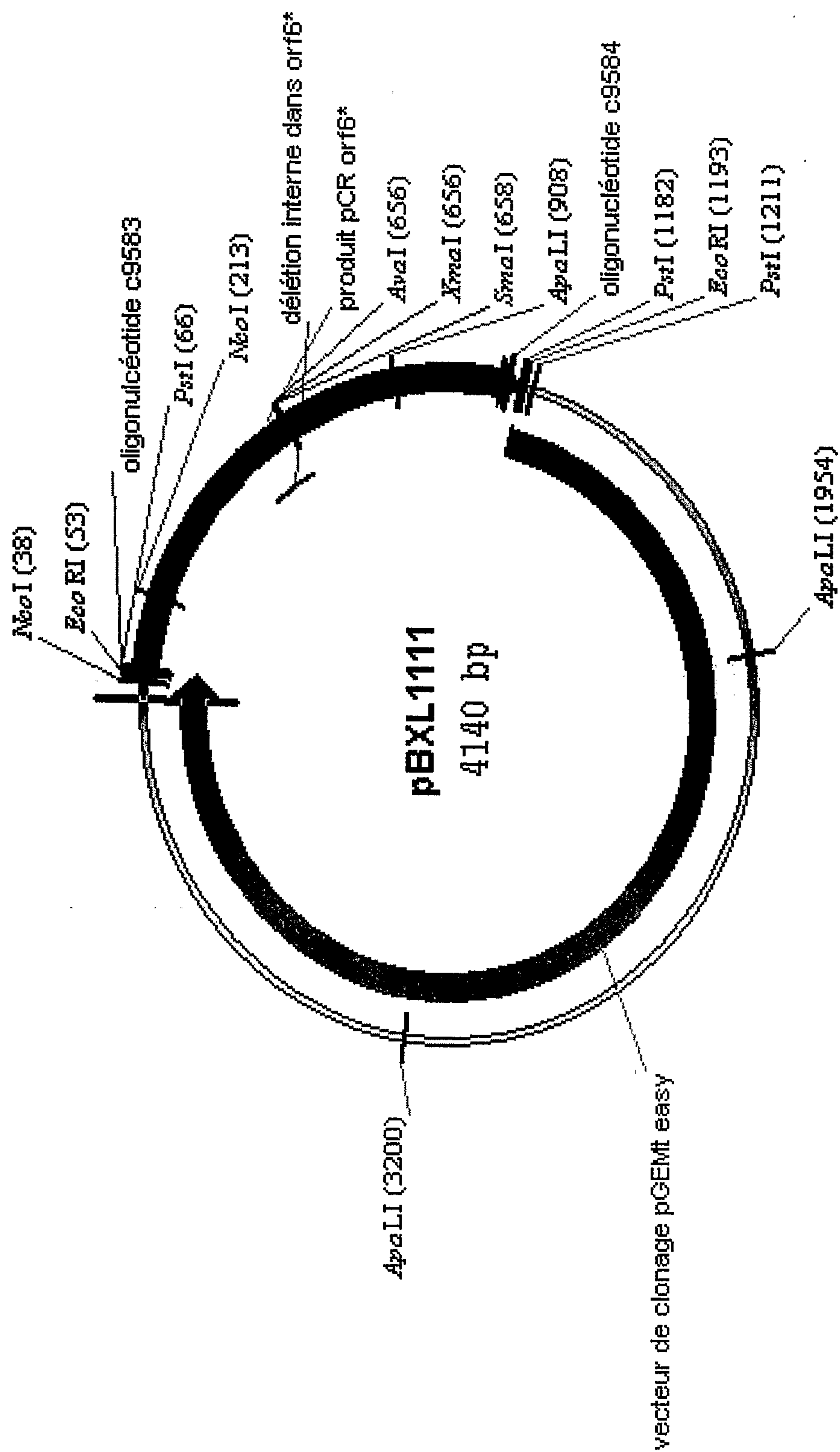


FIG. 16

17/41

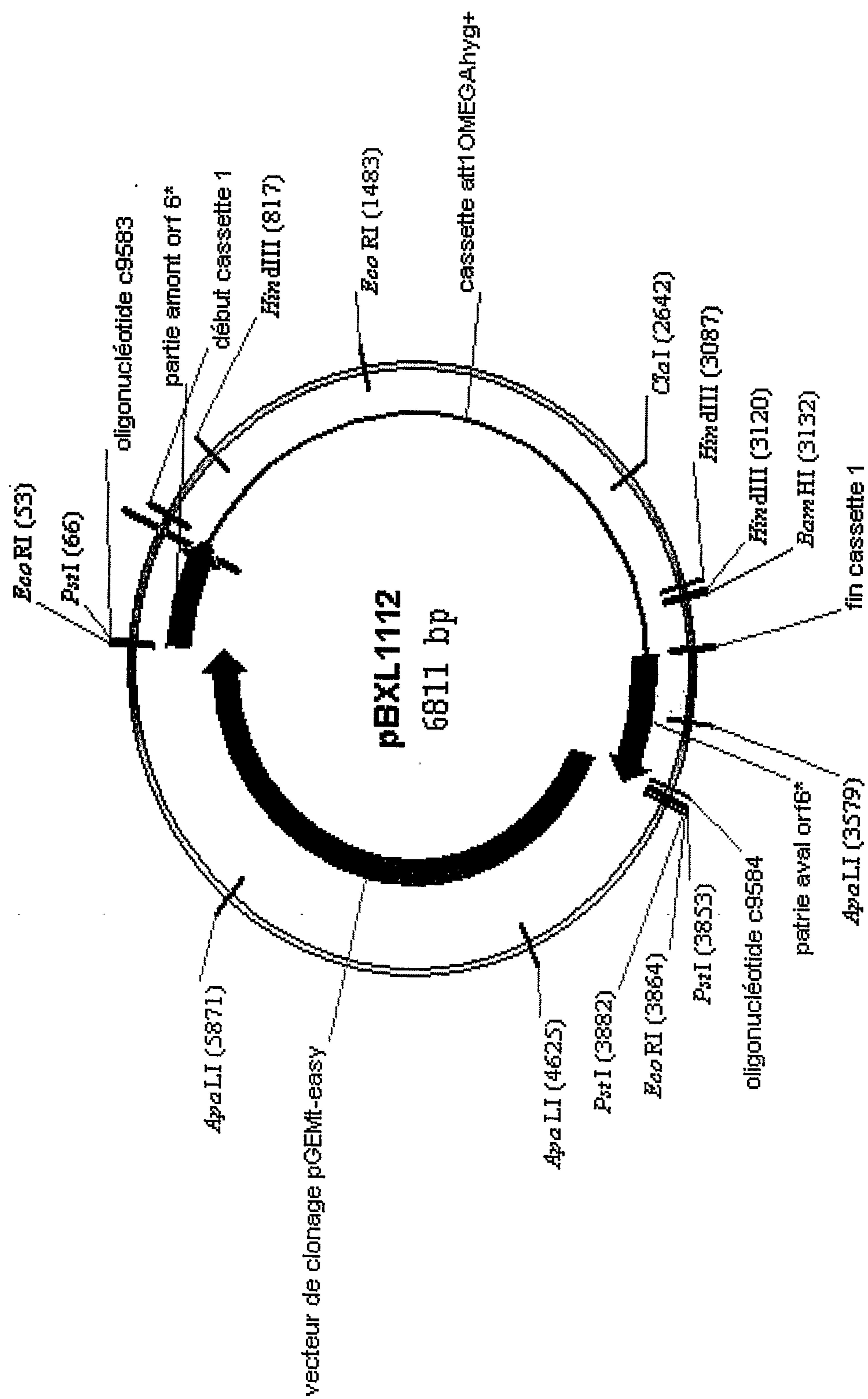


FIG. 17

18/41

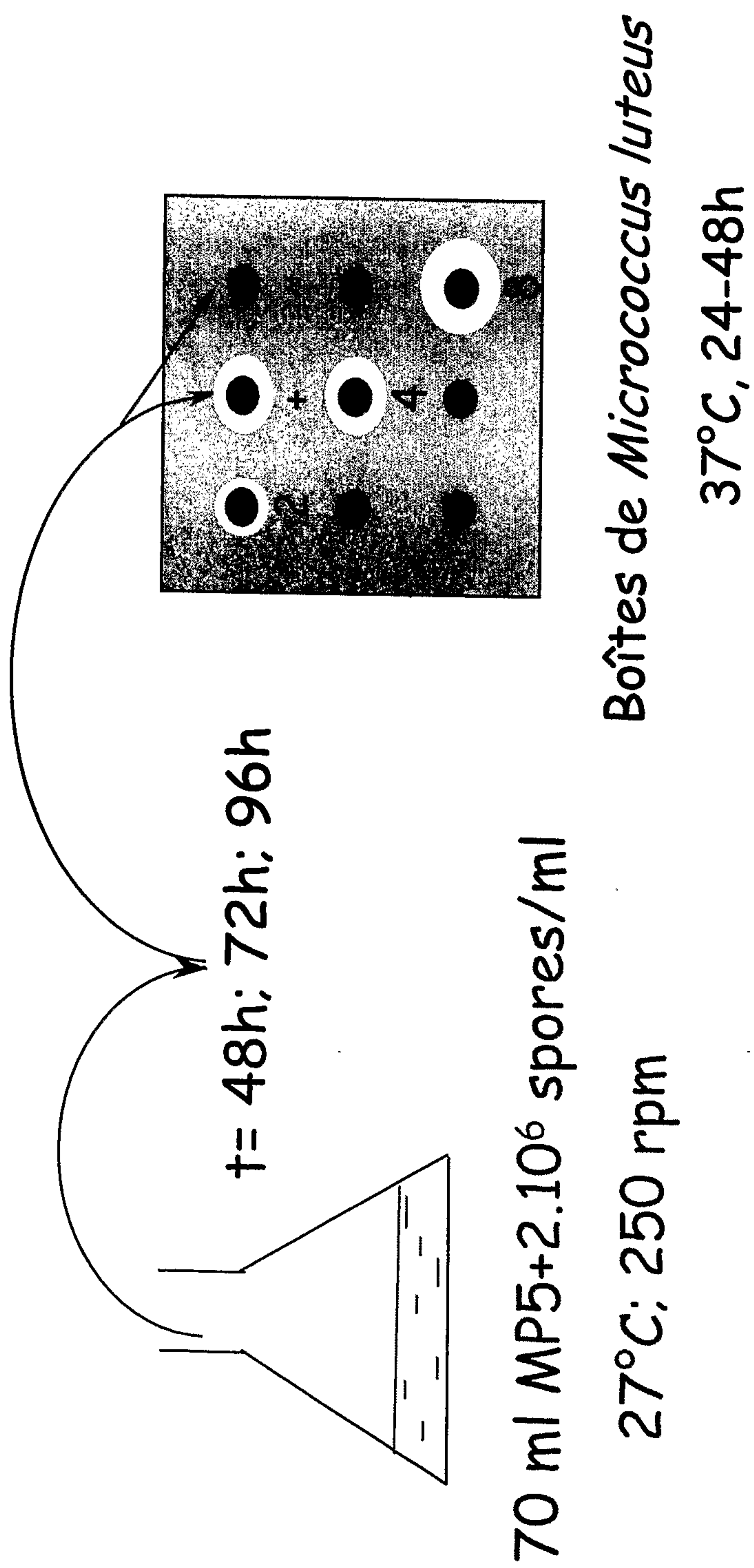


FIG. 18

19/41

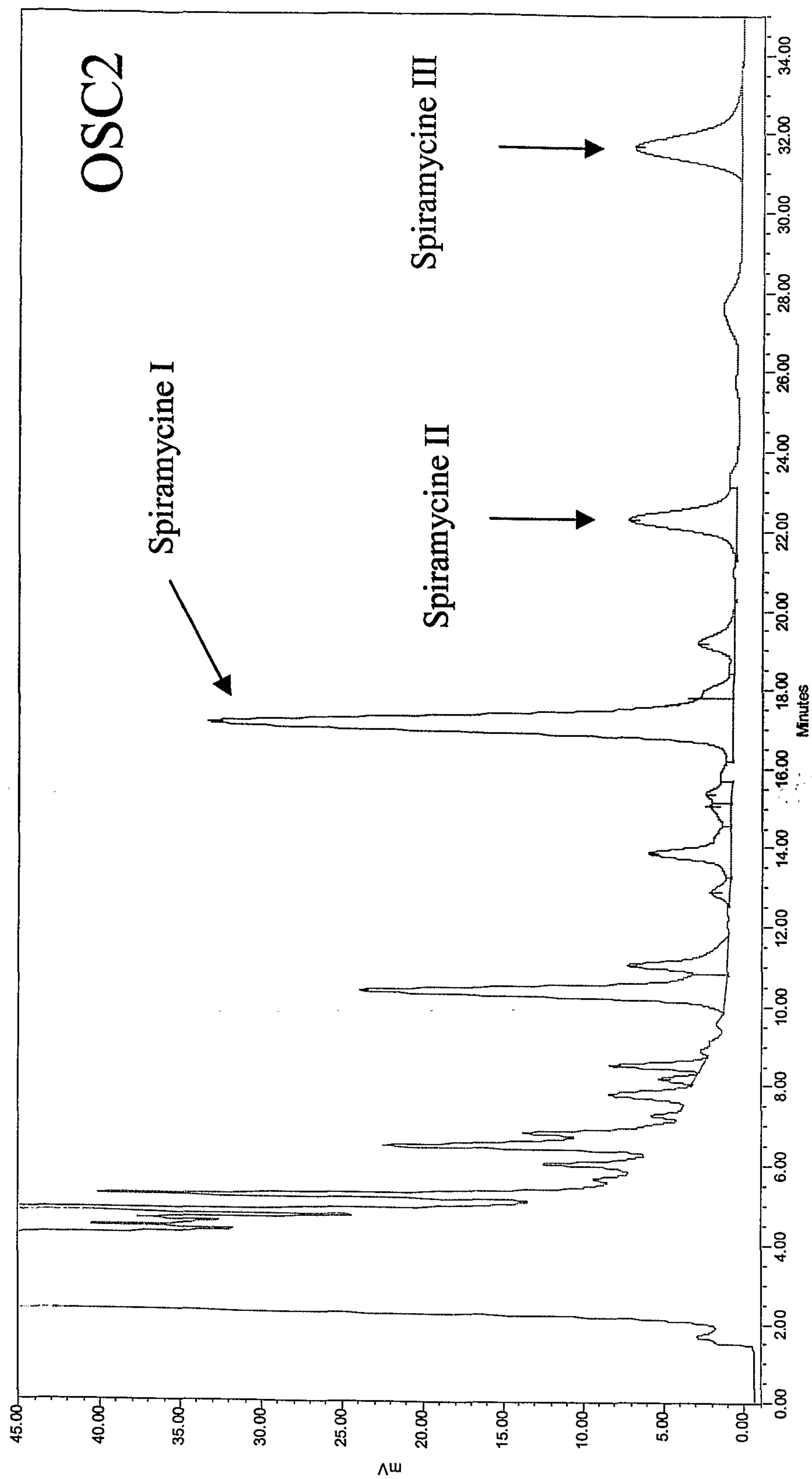


FIG. 19

20/41

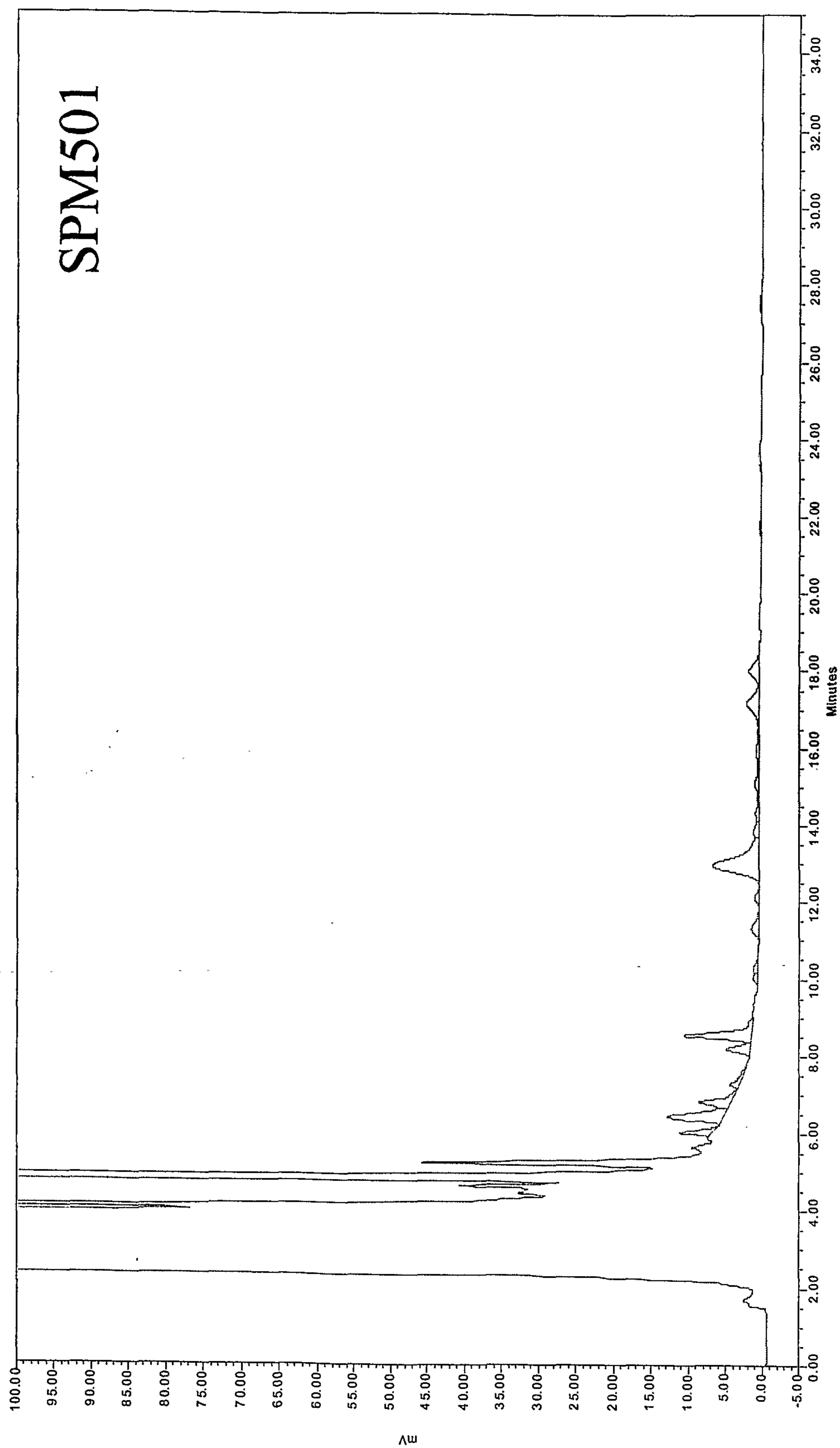


FIG. 20

21/41

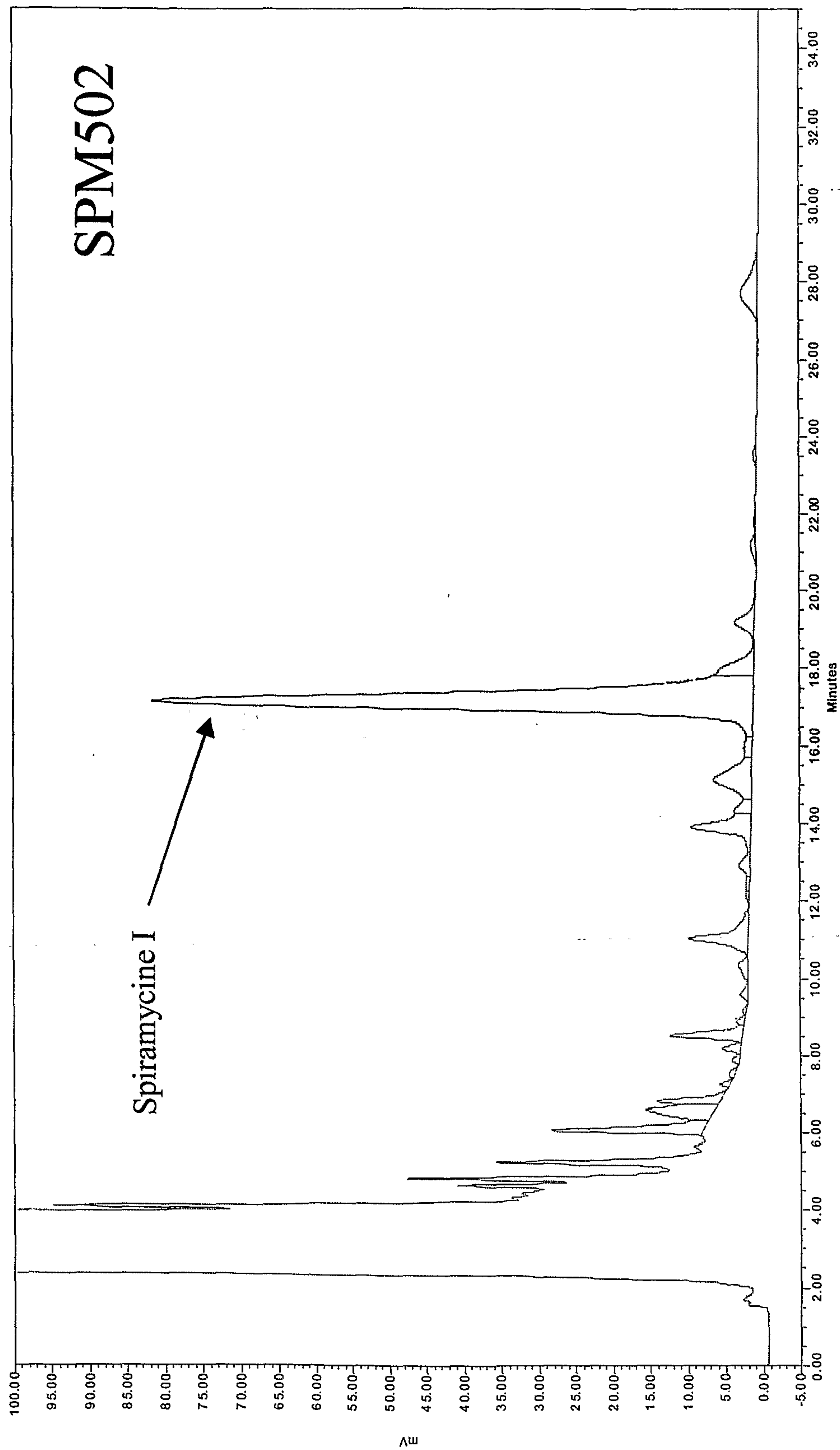


FIG. 21

22/41

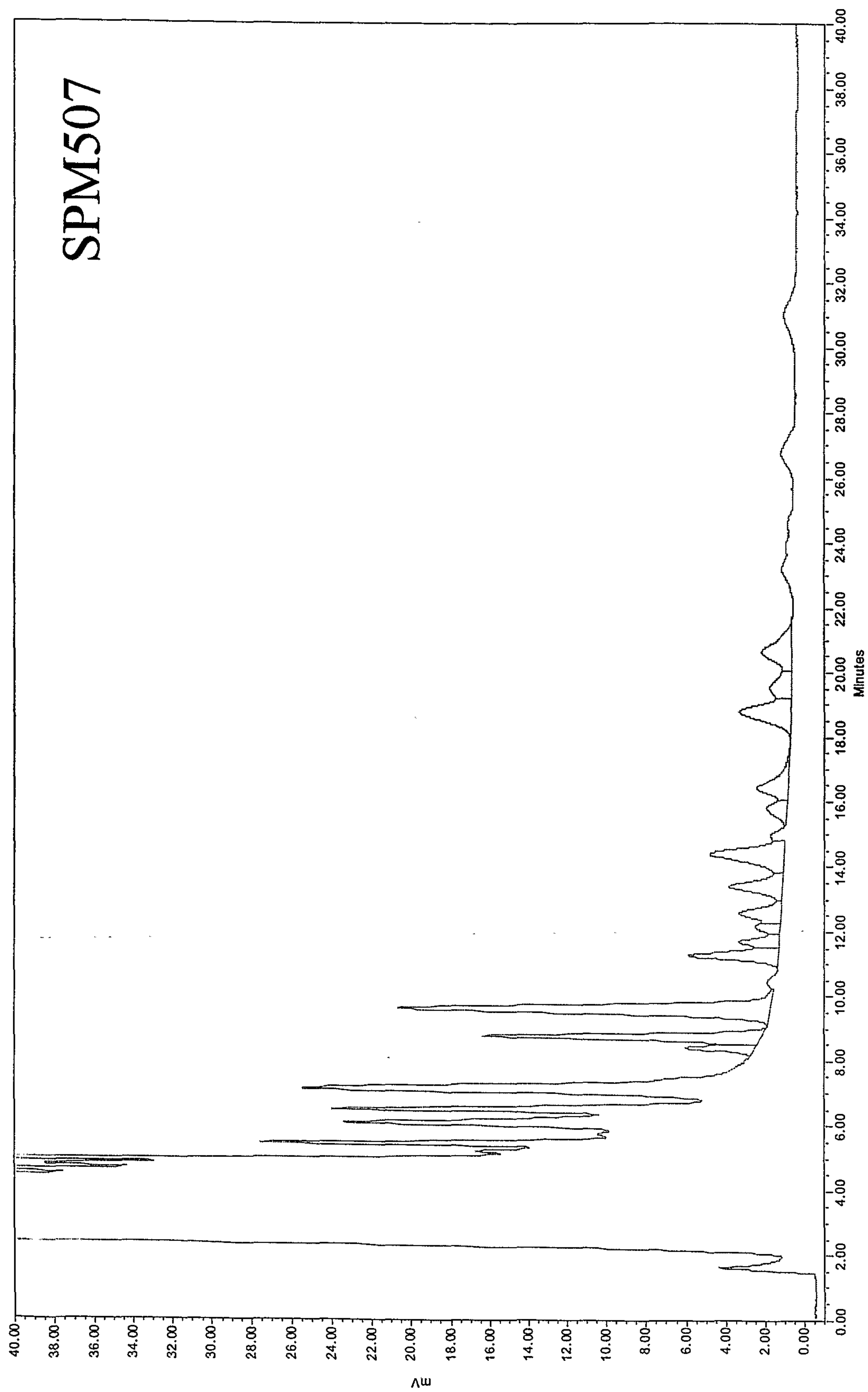


FIG. 22

23/41

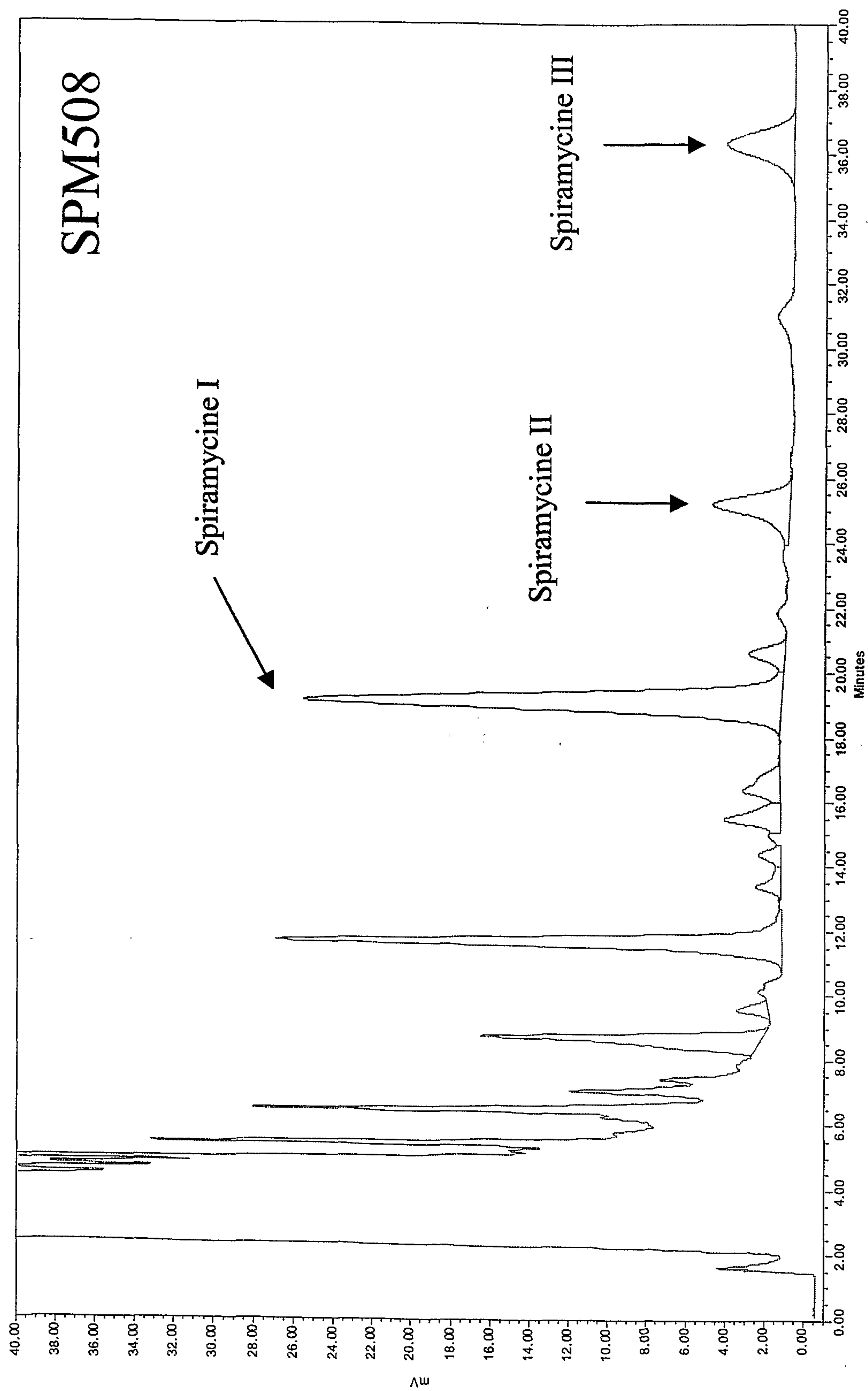


FIG. 23

24/41

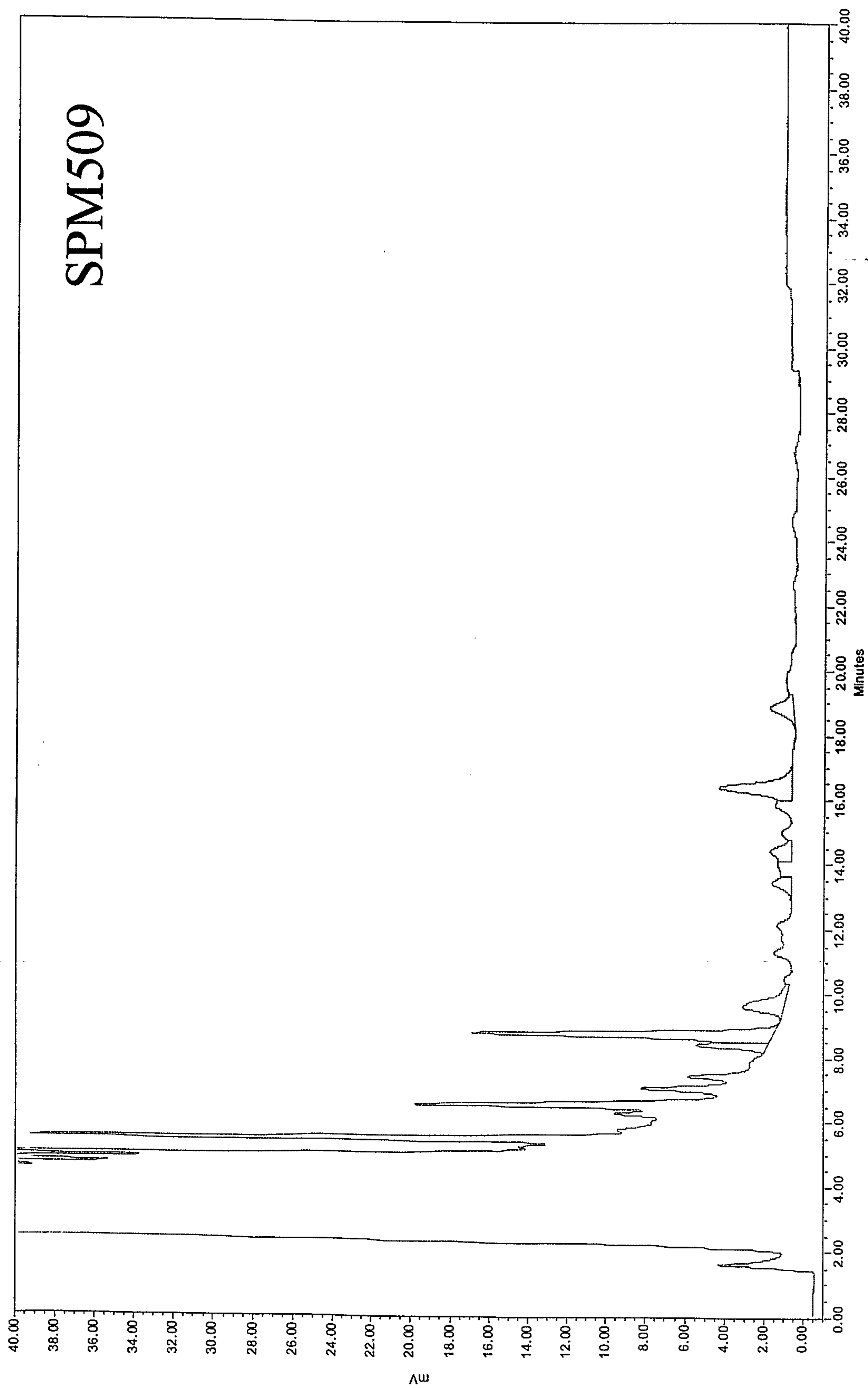


FIG. 24

25/41

FASTA (3.42 Sept 2001) function [optimized, BL50 matrix (15:-5)]
ktup: 2 join: 37, opt: 25, open/ext: -10/-2, width: 16

Smith-Waterman score: 1467; 63.130% identity (66.480% ungapped) in
377 aa overlap (3-367:11-380)

orf3	10	20	30	40
	MTIPFLD----	AGAGYRELRAEIDAALQ	RVSASGRYLLDAELAA	FEFEFAA
tylB	10	20	30	40
	MTGLPRPAVRVPFHDLRD	VHAATG--VESEIGGALL	RVAARGRYLLGAELAA	FEERFAE
orf3	50	60	70	80
	YCDNDHCVAVGSGCDALE	LSLRALDIGPGDEVVPA	HTFIGTWLAVSATGAR	PVAV----
tylB	50	60	70	80
	YCGNAHCVAVGSGGLDDA	RLALWALGVGEGDEVIV	PSHTFIASWLAVSATG	ATPVPVEPGD
orf3	110	120	130	140
	--DPTPDGLSLDPALVEA	ALTPTRALMPVHLHGH	PADLDPDLLAIAGRH	GLAVVEDAAQA
tylB	110	120	130	140
	PGEPPGGAFLLDPDRLEA	ALTPTRAVMPVHLYGH	PVDLDPVGAFAPHGL	AVVEDAAQA
orf3	170	180	190	200
	HGARYRGRRIGSGHVVA	FSFYPGKNLGAMGDGG	AVVTGDSGVAERIRLL	RNCGSREKYRH
tylB	170	180	190	200
	T-ARYRGRRIGSGHRTAF	SFYPGKNLGALGDGG	AVVTSDPELADRLRL	LLRNYGAREKYRH
orf3	230	240	250	260
	EVRSTHSRLDEFQAAVL	RRAKLPRLDARNARRAG	TAERYGRALGPVPQIA	VPVTAPWADPV
tylB	230	240	250	260
	EERGTNSRLDELQAAVL	SVKLPYLDWNTRRREIA	ARYGEALAGLPGVTP	-EGRVAEPV
orf3	290	300	310	320
	WHLYVIRCAERDELRRRL	ERAGVQTLIHYPVPPH	RSPAYADDPAGA-PAG	THPLSERLAA
tylB	290	300	310	320
	WHQYVLRSPYRDRLRRRL	AEAGVETLVHYPVAVH	ASGAYAG--AGPCPAG	GLPRAERLAG
orf3	350	360	370	380
	QSLSLPLGPHLGEDEAR	AVVAAVRAASAGLAAY	PTPDGQRFPLVTEKR	
tylB	350	360	370	380
	EVLSLPIGPHLPDEAVE	VVIAAVQSAALDSWEE	GP	

FIG. 25

Comparison of:
(A) SRMD=Query
(B) MDMA=Sbjct

>gi|320536|pir||A60725 midcamycin-resistance protein mdmA - Streptomyces mycarofaciens
Length = 271

Score = 380 bits (975), Expect = e-104
Identities = 211/269 (78%), Positives = 232/269 (85%)

(A) Query: 7 LPPQSVSAPADSRSTARREWGQNFFRTAAACRFSQLDGSDTTPPDSPNDLMTVEIGAG 66
+ P S SAPA SRSTARRE GQNFFR+AAACRFS QLD P S D++TVEIGAG
(B) Sbjct: 1 MSPISASAPAAASRSTARRELQNFFRSAAACRFSQDLDAFCADLPGLADVLTVVEIGAG 60

(A) Query: 67 SGRVTKVLASPGTPLLAVEIDPRWARRLAAESLPDVTVVNEDEFLTTLQLPGQPVRLIGNLP 126
SGRVTK LAS G LLAVEID WARRL AESLPDVTVVNEDEFL LQLP QP+RLIGNLP
(B) Sbjct: 61 SGRVTKALASAGRSLLAVEIDAYWARRLTAESLPDVTVVNEDEFLNLQLPRQPIRLIGNLP 120

(A) Query: 127 FVTGTRMLRRCLDMGPARMRQGVFLLQREYVVGKRTGAWGGNLFNAQWEPWYSFDRGLAFS 186
FV+GT++LRRCL++GP RM Q VFLLQREYVVGKRTGAWGGNLFNAQWEPWY+F+ GLAFS
(B) Sbjct: 121 FVSGTKILRRCLLELGNRMCAVFLQREYVVGKRTGAWGGNLFNAQWEPWYTFEGGLAFS 180

(A) Query: 187 RQDFTPVPRADTQTLMVAPHRRPSVPWREKAAYQRFVQRVFDTGQMTVGDAARKVLRRGH 246
R +F+PVPRADTQTL+V P RRPSVPWRE+ YQRF Q++FDTGQMT+G+AARKVLRRGH
(B) Sbjct: 181 RNEFSPVPRADTQTLVVMPPRRRPSVPWRETDYQRFQIFDTGQMTIGEAAARKVLRRGH 240

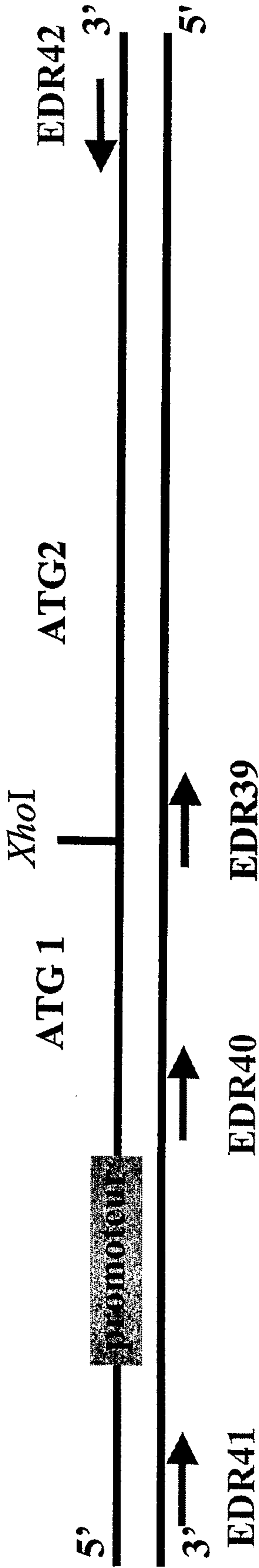
(A) Query: 247 AQFVRGAGVRPADRVKDLTVPEWTALFRA 275
AQFVR AGVRPADRVKDLTV +W ALFRA
(B) Sbjct: 241 AQFVRSAGVRPADRVKDLTVRDWAALFRA 269

FIG. 26

27/41

phase 1 (att1):	atcgcgcgcgcttcggttcgggacgaaagaggtagat
phase 2 (att2):	atcgcgcgcgcttcggttcgggacgaaagaggtagat
phase 3 (att3):	atcgcgcgcgcttcggttcgggacgaaagaggtagat

FIG. 27



ATG 1: ATG situé le plus en 5'
ATG 2: ATG le plus en 3'

EDR39-EDR42 → "orf10 court" pSPM523
EDR40-EDR42 → "orf10 long" pSPM524
EDR41-EDR42 → "orf10 pro" pSPM525

FIG. 28

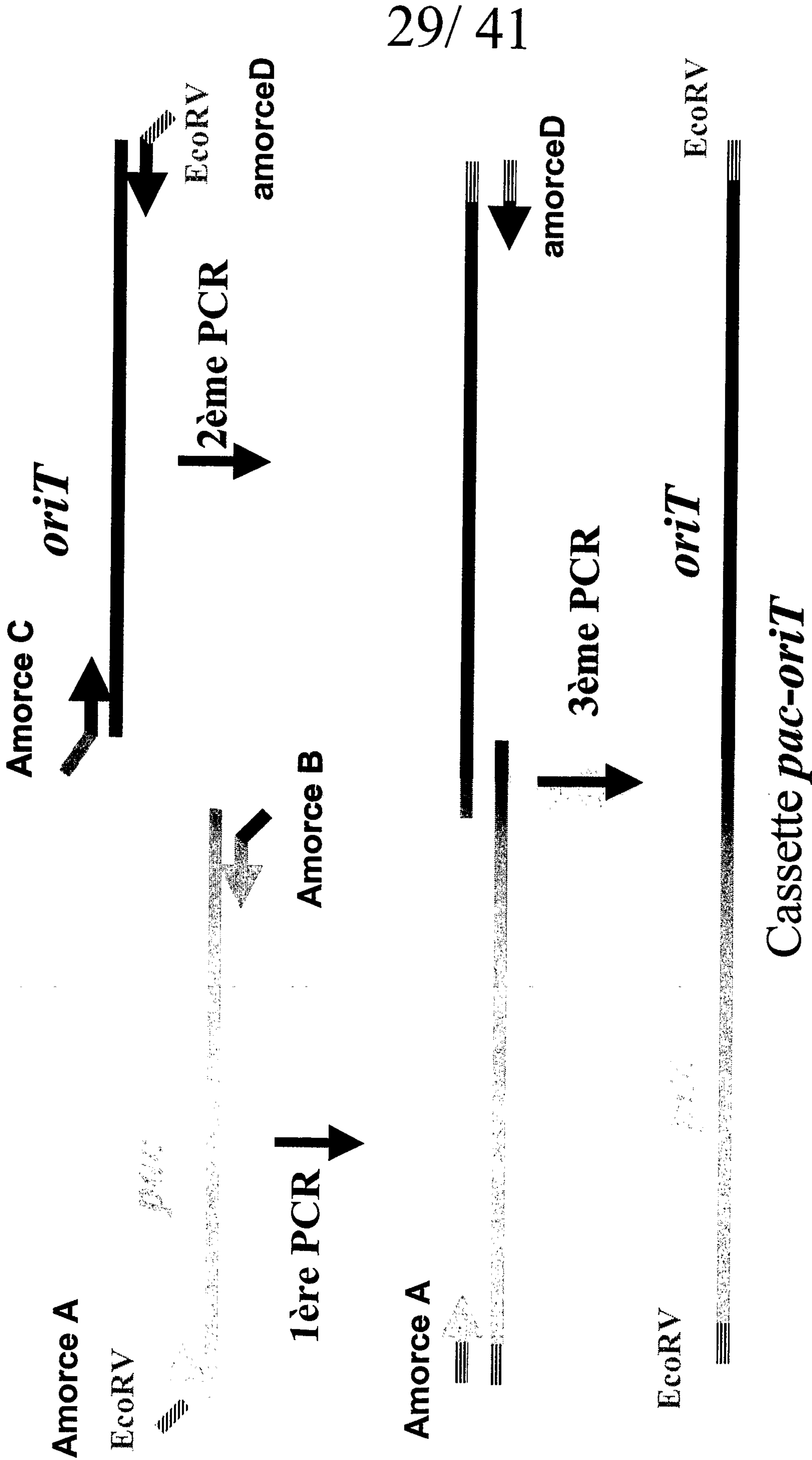


FIG. 29

30/41

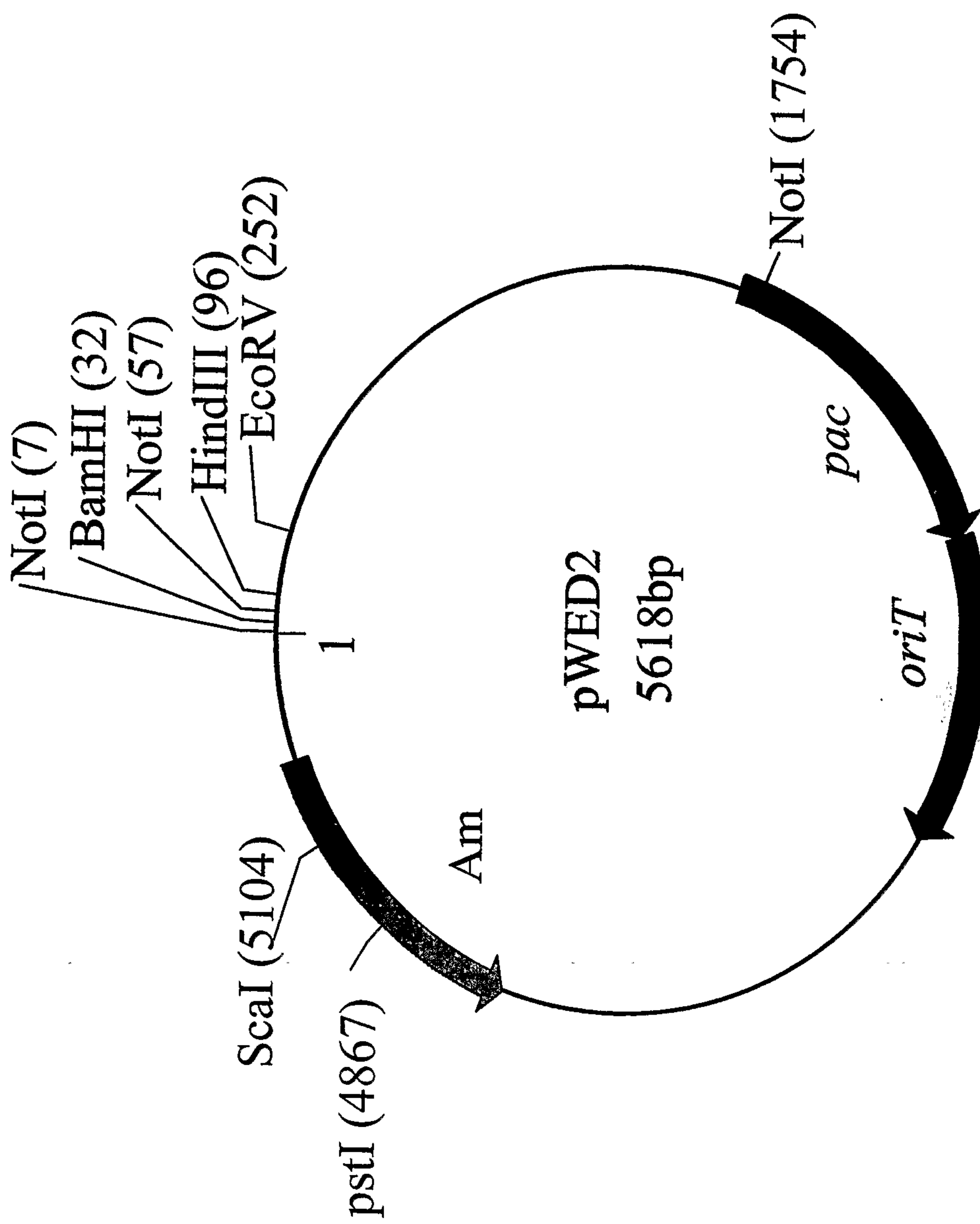
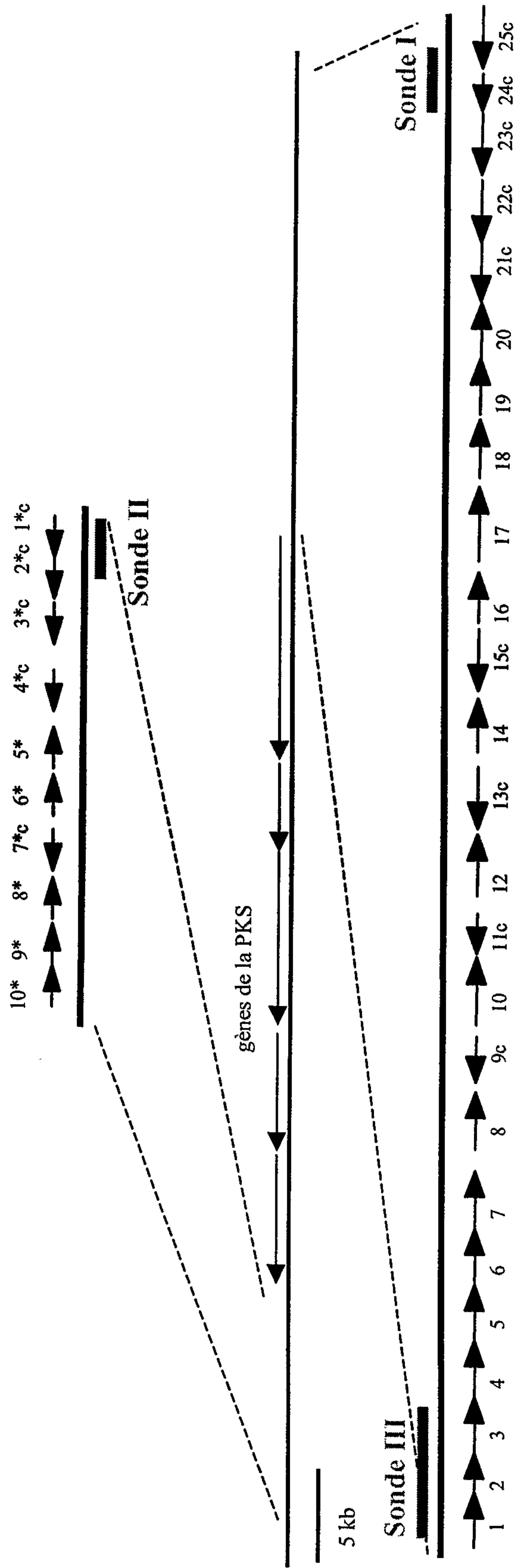


FIG. 30



Sonde I: produit PCR ORF23c-ORF25c

Sonde II: produit PCR ORF1*c-ORF2*c

Sonde III: fragment d'ADN couvrant *orf1*, *orf2* et *orf3*

FIG. 31

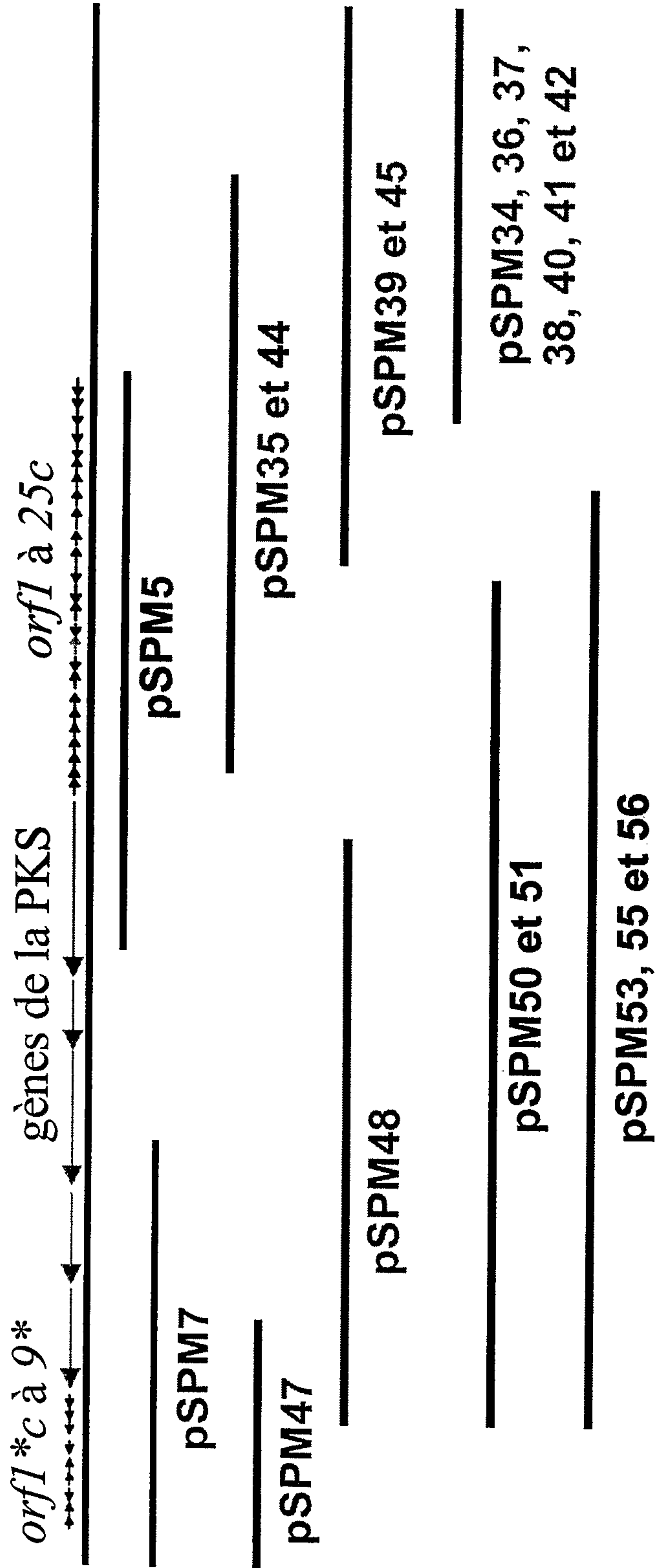


FIG. 32

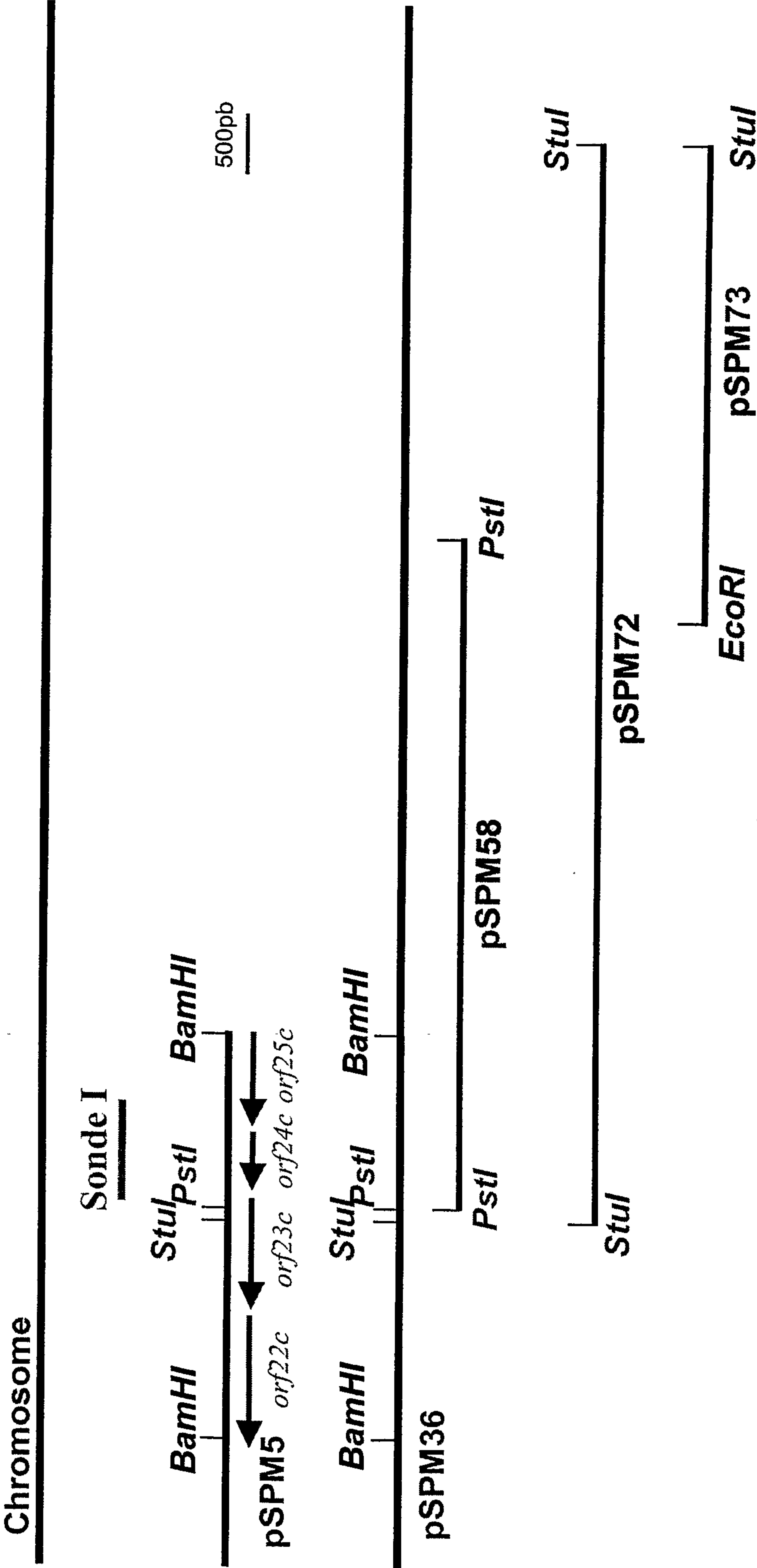


FIG. 33

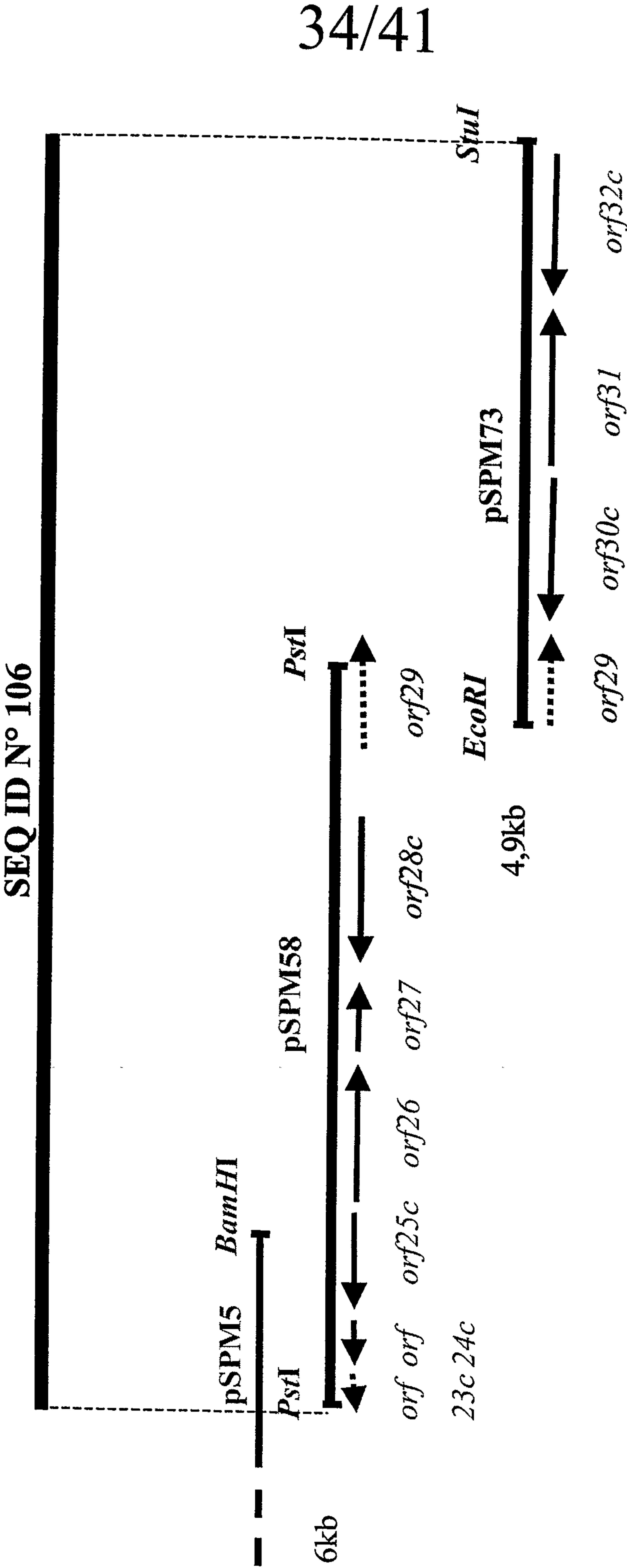


FIG. 34

35/41

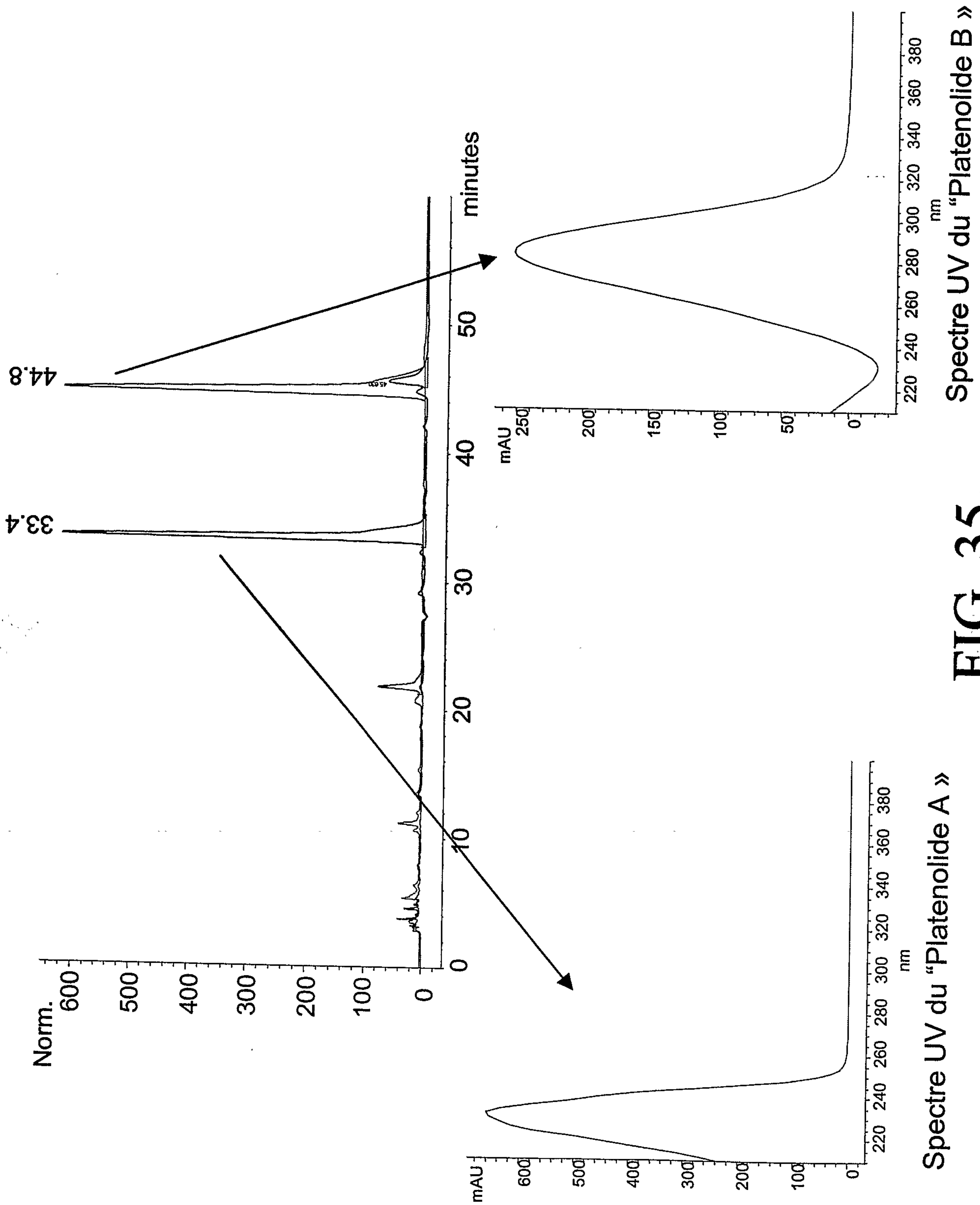
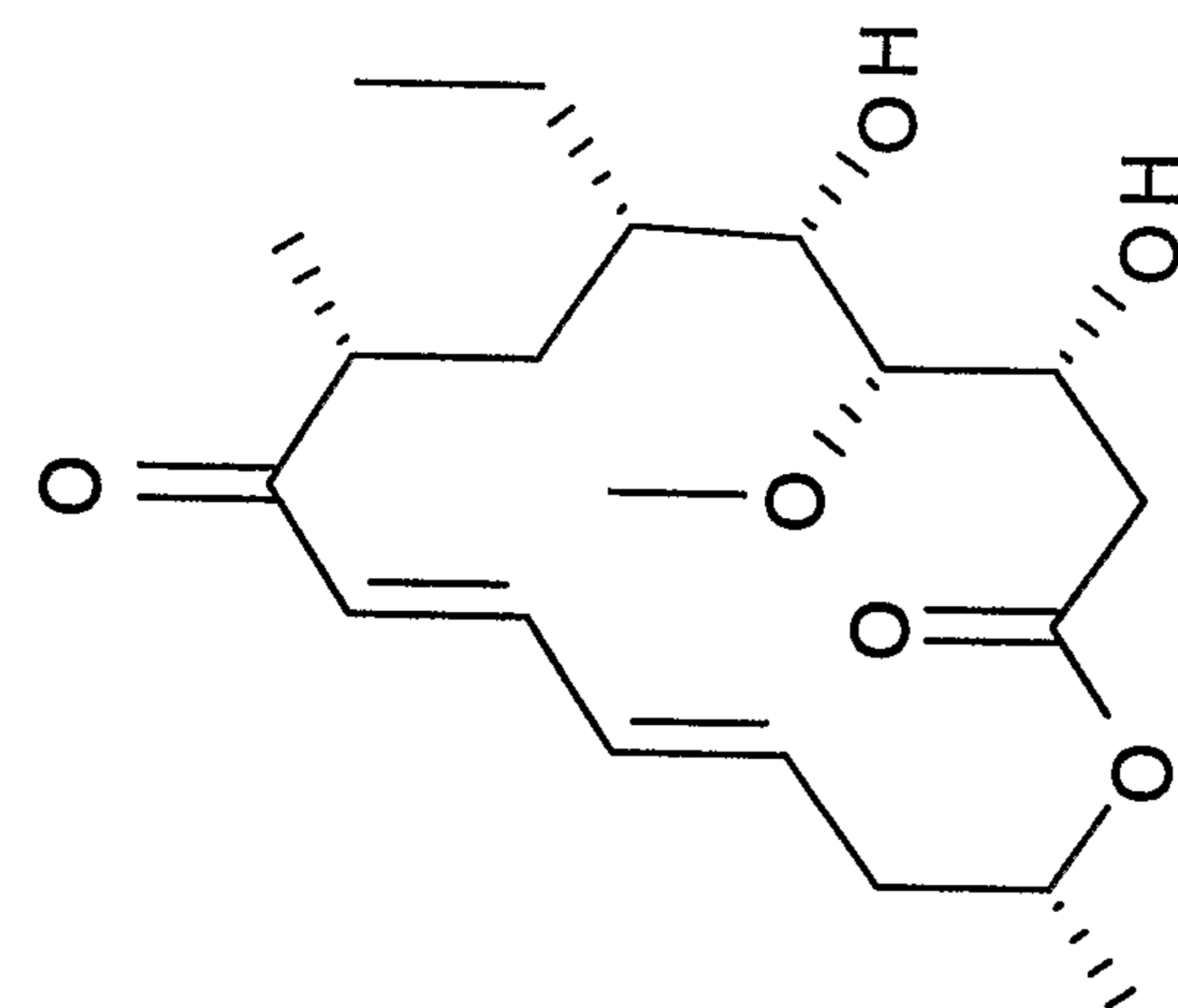
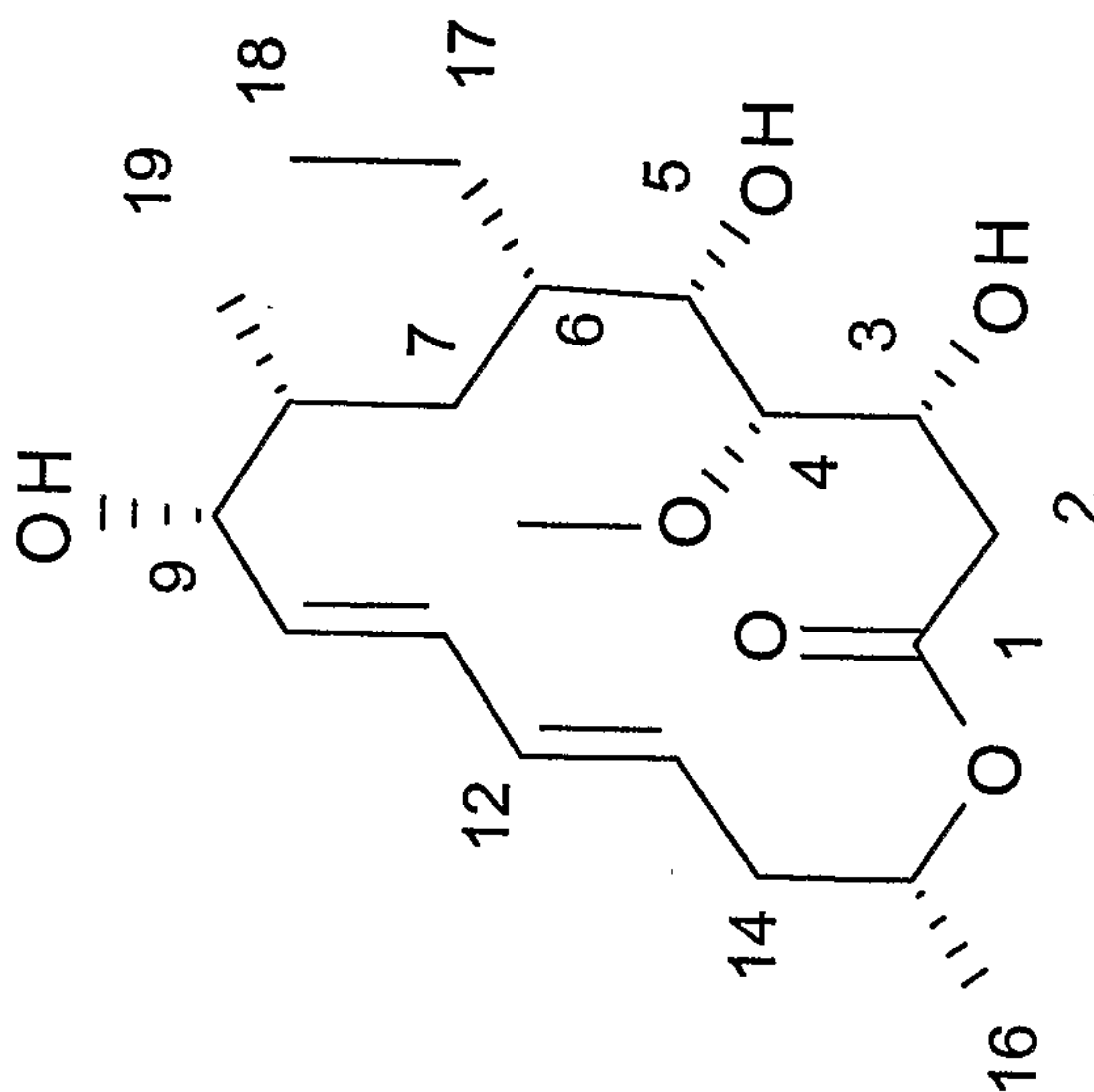


FIG. 35

« Platenolide A » MM=370



Masse moléculaire =368
Formule moléculaire =C20H32O6



Formule moléculaire=C20H34O6
Masse moléculaire =370

FIG. 36

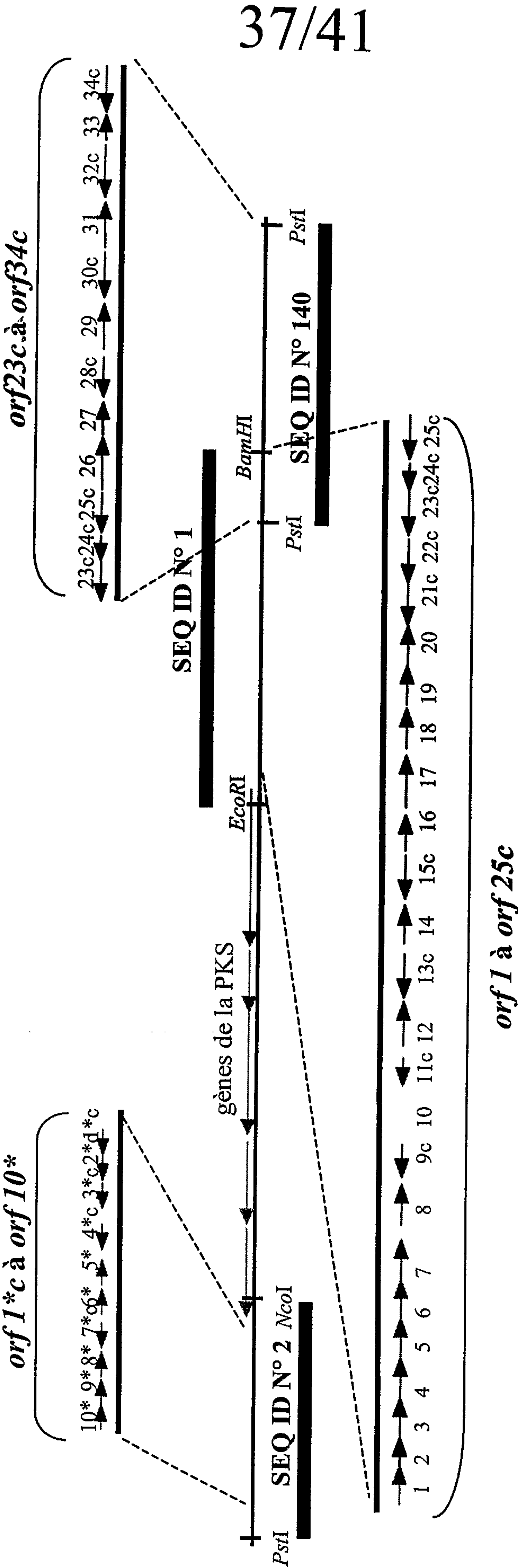


FIG. 37

38/41

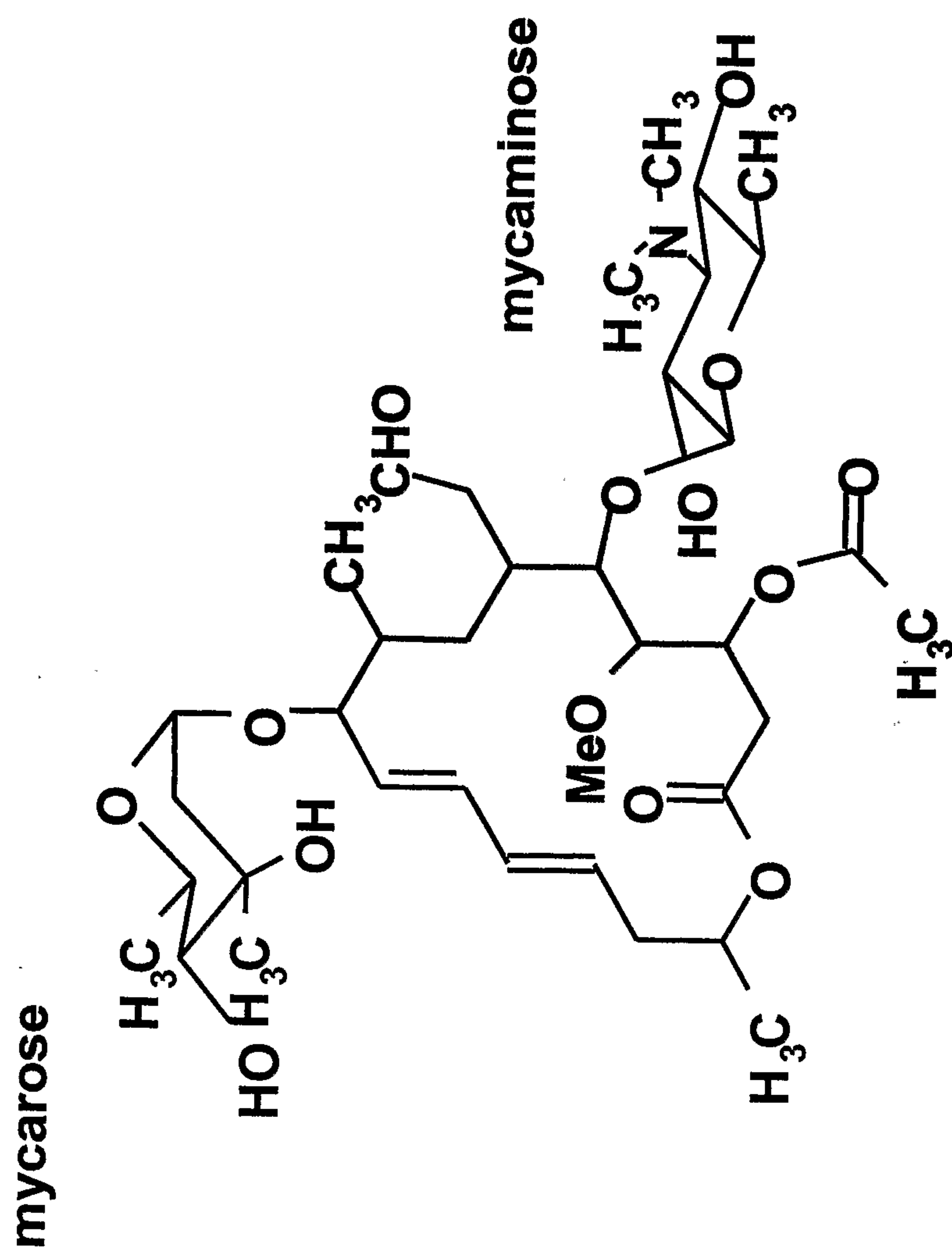


FIG. 38

39/41

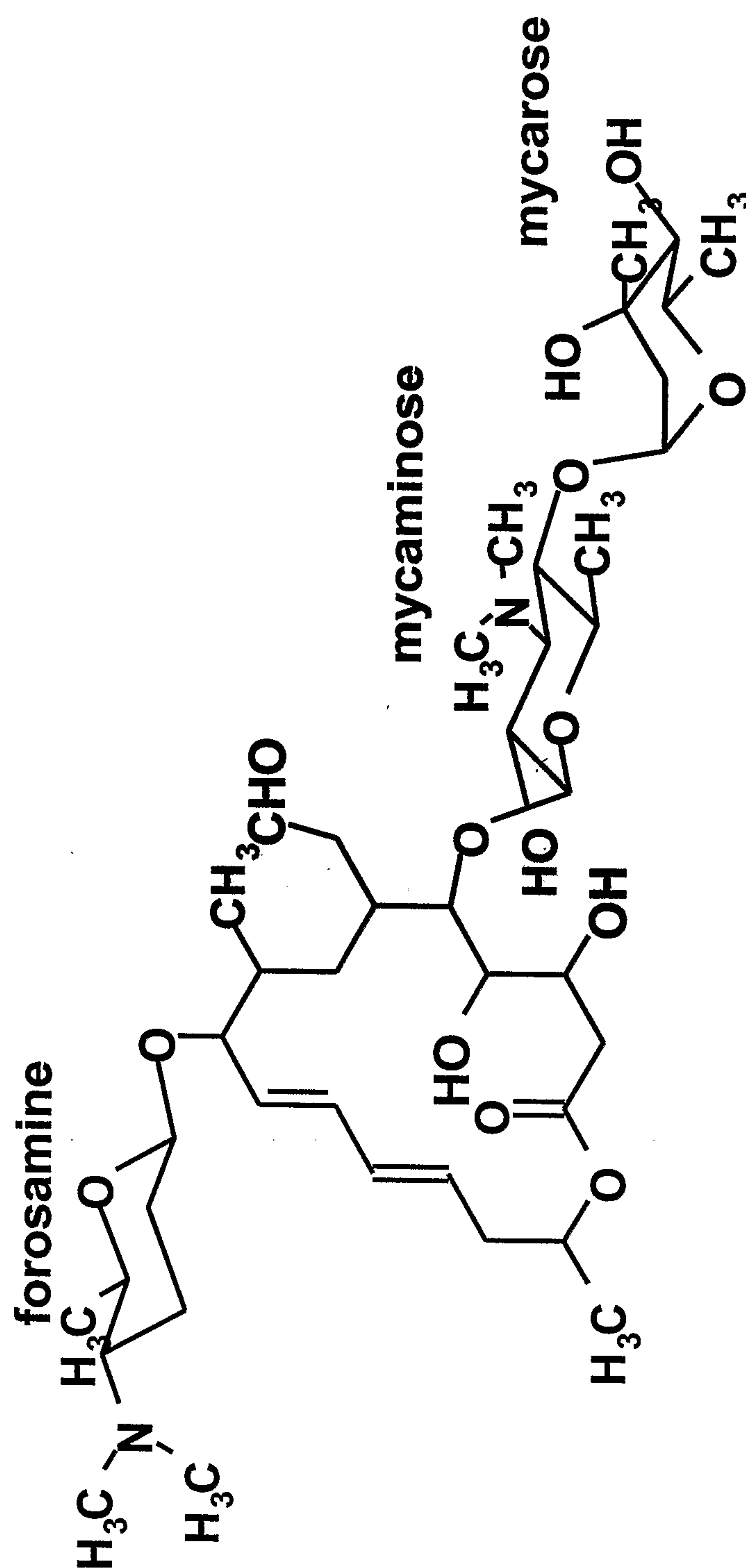
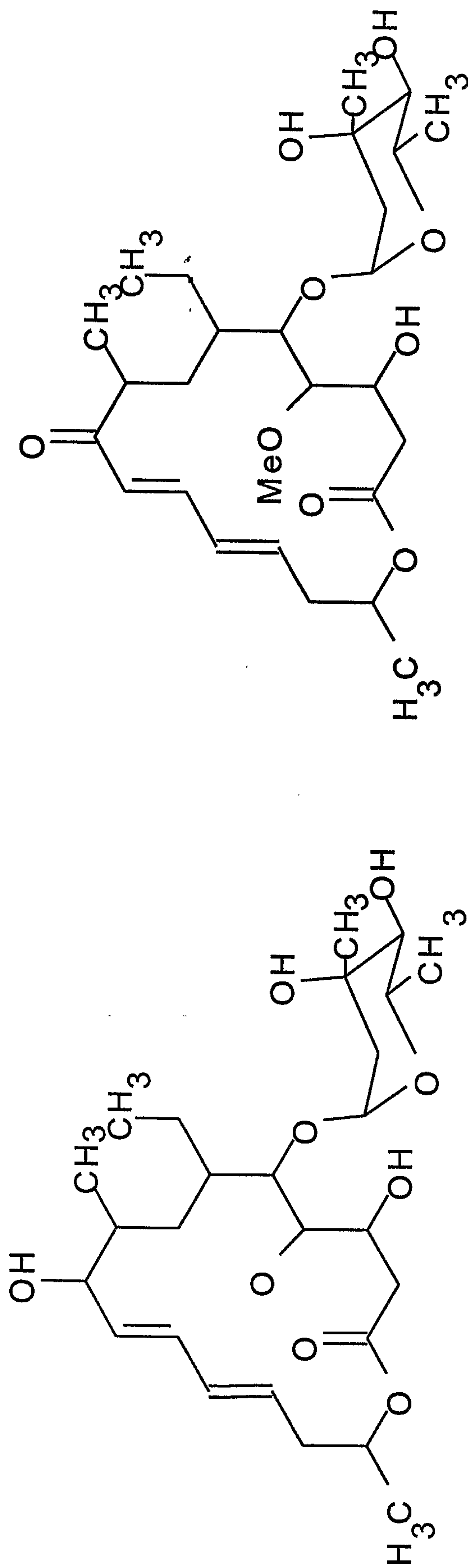


FIG. 39

40/41



Platénolide B + mycarose

Platénolide A + mycarose

FIG. 40

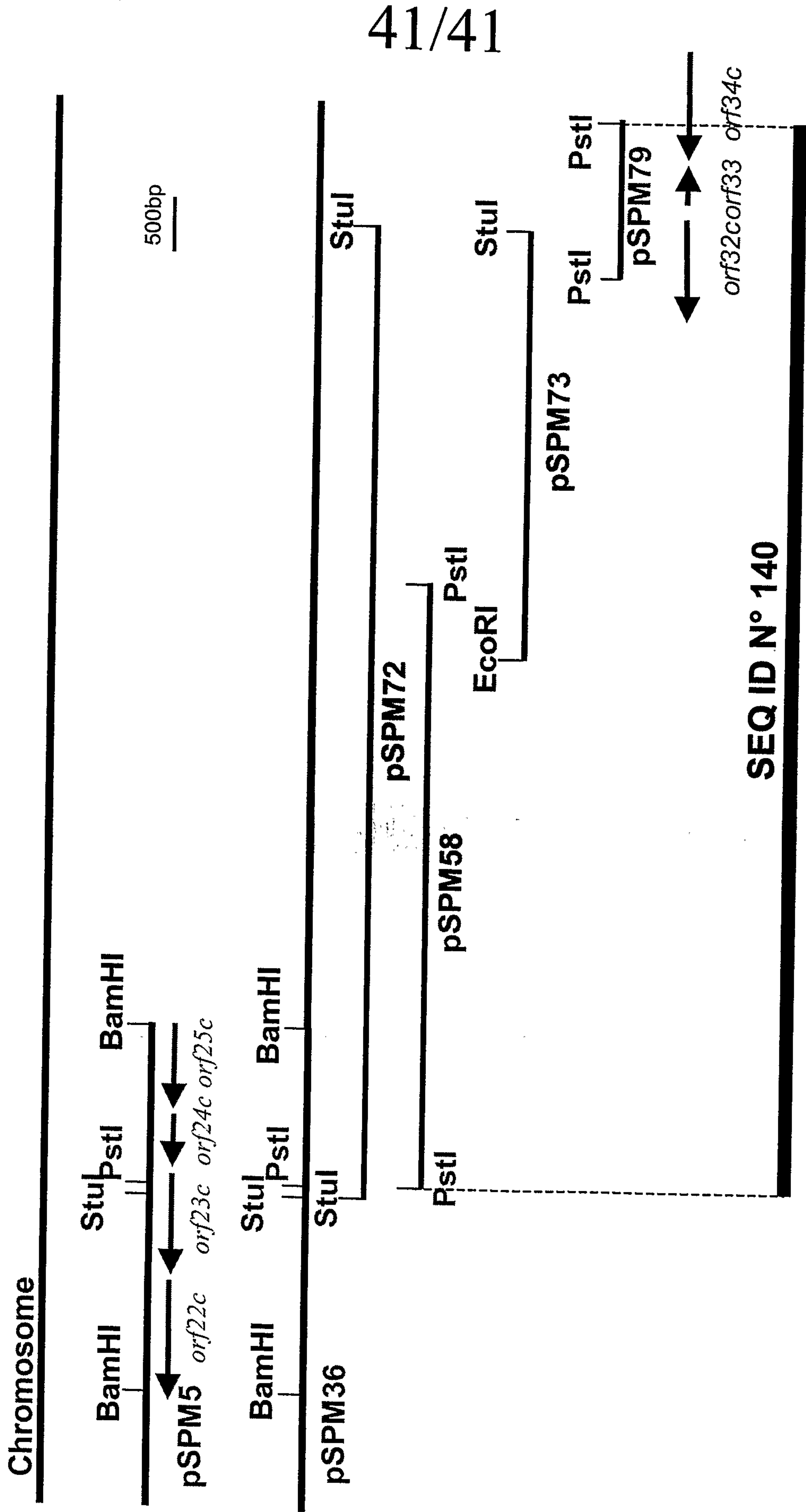


FIG. 41