



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0039342

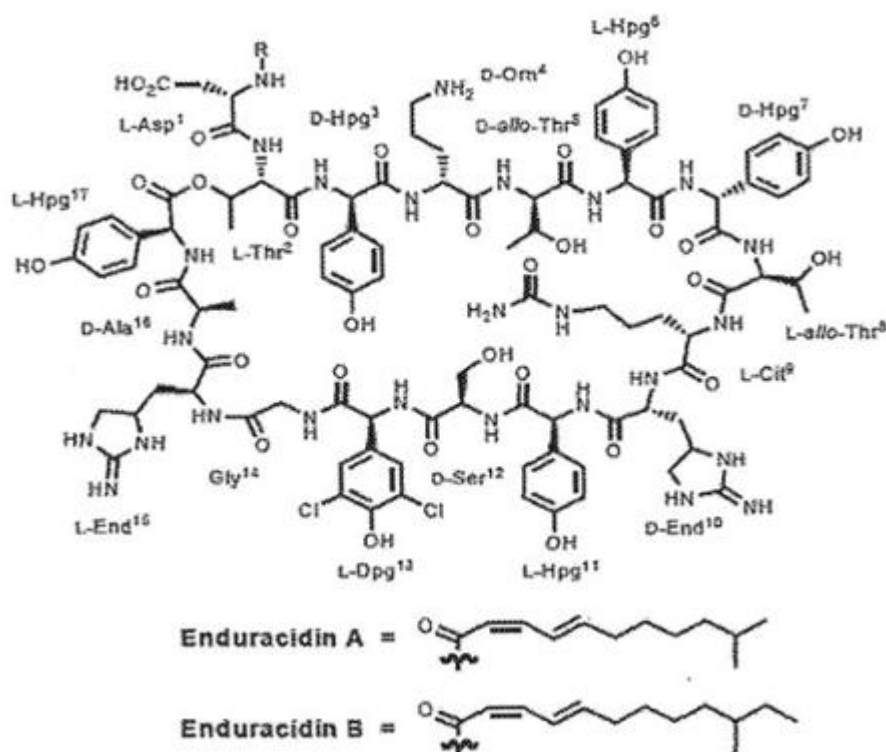
(51)^{2022.01} C12N 15/76; C12P 21/04; C12N 1/21;
C12N 15/31

(13) B

- (21) 1-2019-03212 (22) 01/12/2017
(86) PCT/US2017/064328 01/12/2017 (87) WO 2018/106545 14/06/2018
(30) 62/430,838 06/12/2016 US; 62/479,087 30/03/2017 US
(45) 25/04/2024 433 (43) 25/09/2019 378A
(73) OREGON STATE UNIVERSITY (US)
A312 Kerr Administration Building, Corvallis, OR 97331-2140, United States of America
(72) ZABRISKIE, Mark, T. (US); YIN, Xihou (US).
(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) CHỦNG STREPTOMYCES FUNGICIDICUS TÁI TỔ HỢP ĐỂ TĂNG CƯỜNG SẢN XUẤT ENDURACIDIN VÀ PHƯƠNG PHÁP SẢN XUẤT ENDURACIDIN

(57) Sáng chế đề cập đến chế phẩm và phương pháp để tăng cường sản xuất enduracidin ở chủng *Streptomyces fungicidicus* được thiết kế về mặt di truyền. Cụ thể, sáng chế mô tả thao tác di truyền của các gen điều hòa *orf24* và *orf18* kết hợp với cụm gen sinh tổng hợp enduracidin (enramycin) từ *Streptomyces fungicidicus* tạo ra cấu trúc vector và các chủng tái tổ hợp tạo ra hiệu suất enduracidin lớn hơn.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến sinh tổng hợp thuốc kháng sinh, cụ thể là, chế phẩm và phương pháp để tăng cường sản xuất enduracidin.

Tình trạng kỹ thuật của sáng chế

Các bệnh nhiễm trùng do vi khuẩn kháng nhiều thuốc xuất hiện trên toàn cầu kéo theo chi phí cực lớn cho việc chăm sóc sức khỏe và trở thành mối đe dọa chính đối với sức khỏe cộng đồng. Để đón đầu sự phát triển của tình trạng kháng lại thuốc chống vi khuẩn, cần tìm ra các loại thuốc kháng sinh mới cũng như các phương pháp sản xuất các thuốc kháng sinh này theo hướng đạt hiệu quả về chi phí hơn.

Bản chất kỹ thuật của sáng chế

Sáng chế khắc phục các vấn đề nêu trên kết hợp với việc sản xuất enduracidin (enramycin) có giới hạn ở các chủng *Streptomyces fungicidicus* kiểu dại, cũng như việc sản xuất giới hạn ở các chủng công nghiệp được phát triển bằng cách bức xạ và đột biến qua trung gian hóa chất thông thường đối với nhiễm sắc thể và nhiều vòng chọn lọc đột biến liên tiếp để sản xuất các mức tăng của thuốc kháng sinh enduracidin peptit mong muốn. Sáng chế bộc lộ thao tác di truyền của các gen điều hòa *orf24* và *orf18* kết hợp với cụm gen sinh tổng hợp enduracidin (enramycin) từ *Streptomyces fungicidicus* để tạo ra vector và chủng tái tổ hợp tạo ra hiệu suất thuốc kháng sinh peptit lớn hơn. Các chủng tái tổ hợp được tạo cấu trúc ở cả hai nguồn kiểu dại, *Streptomyces fungicidicus* B-5477 (ATCC 21013), và *Streptomyces fungicidicus* BM38-2 (ATCC PTA-122342), có nguồn gốc từ chủng kiểu dại và thường được sử dụng để sản xuất enduracidin công nghiệp. Trong sinh vật kiểu dại, sự hợp nhất đặc hiệu vị trí của plasmit pXY152-endorf24, dẫn đến sự biểu hiện quá mức của bản sao thứ hai của *orf24*, tạo ra chủng SfpXY152endorf24. Sự hợp nhất fosmid pXYF24D3 gây đột biến vào nhiễm sắc thể

kiểu đại thay thế *orf18* nguyên thể bằng bản sao đứt đoạn của gen và tạo ra đột biến SfpXYF24D3. Hoạt động trong nguồn *Streptomyces fungicidicus* BM38-2 (ATCC PTA-122342) thương phẩm, sự hợp nhất plasmit pXY152-endorf24 tạo ra chủng *Streptomyces fungicidicus* BM38-2.24/16 tái tổ hợp. Để tạo ra chủng có nguồn gốc BM38-2 (ATCC PTA-122342) thiếu *orf18* chức năng, plasmit pKS-T-*orf18pfrd*-AmR được tạo cấu trúc để xóa *orf18* và các vùng hai bên của nó, thay thế vùng này bằng chất chỉ thị kháng apramycin và tạo ra chủng *Streptomyces fungicidicus* BM38-2.18pfrd-AmR tái tổ hợp. Các chủng thao tác về mặt di truyền được chứng minh là tạo ra hiệu suất enduracidin cao hơn khoảng từ 1,2 đến 4,6 lần so với các chủng gốc tương ứng. Hiệu suất enduracidin tăng từ các chủng tái tổ hợp tạo ra sự sản xuất enduracidin hiệu quả hơn về mặt chi phí.

Các dấu hiệu và ưu điểm nêu trên và các dấu hiệu và ưu điểm khác của sáng chế sẽ trở nên rõ ràng hơn từ phần mô tả chi tiết dưới đây có dựa vào các hình vẽ kèm theo.

Mô tả vắn tắt các hình vẽ

FIG. 1 là cấu trúc hóa học của các enduracidin A và B.

FIG. 2 là bản đồ sự biểu hiện hợp nhất của plasmit pXY152-endorf24.

FIG. 3 là bản đồ xóa đoạn gen của plasmit pXY300-*orf18ifd*.

FIG. 4 là bản đồ xóa đoạn gen của plasmit pKS-T-*orf18ifd*.

FIG. 5 là bản đồ xóa đoạn gen của plasmit pKS-T-*orf18pfrd*-AmR.

FIG. 6 là bản đồ xóa đoạn gen của plasmit pKS-*orf18ifd*-T-AmR(NS).

FIG. 7 là bản đồ sự biểu hiện hợp nhất của plasmit pXY152-endorf24-camtsr.

FIG. 8 là bản đồ sự biểu hiện hợp nhất của plasmit pXY152-endorf24-blatsr.

FIG. 9 là sự sắp xếp chất hoạt hóa streptomycin StrR protein (SEQ ID NO: 25) với Orf24 (SEQ ID NO: 26).

Các FIG. 10A và 10B là các bản đồ của quá trình cài xen các plasmit pKS-T-orf18pfrd-AmR (a) và pXY300-orf18ifd (b). Trong cấu tạo pXY300-orf18ifd, trình tự bên trong của *orf18* từ vị trí nucleotit 25795 đến 26450 (Số hiệu lưu giữ Ngân hàng gen DQ403252) được xóa và được thay thế bằng vị trí giới hạn PacI (TTAATTAA, FIG. 10B). Orf18 đã xóa đoạn không làm lệch khung dịch mã thu được (GTGTTTAATTAATGA (SEQ ID NO: 27)) có thể được dịch mã thành peptit ba axit amin (VFN). Nói chung, việc xóa đoạn không làm lệch khung dịch mã bên trong bất kỳ theo chiều dài của *orf18* dẫn đến chức năng của Orf18 bị bất hoạt do tính không đầy đủ của nó.

FIG. 11 là sự sắp xếp orf24 với sáu protein gen trực giao (ortholog) hoạt hóa đặc hiệu con đường giống StrR đặc trưng về mặt chức năng từ actinomycetes. Orf24 (Số hiệu lưu giữ Ngân hàng gen DQ403252; SEQ ID NO: 26) từ cụm gen sinh tổng hợp enduracidin *S. fungicidicus*; StrR (Số hiệu lưu giữ Ngân hàng gen Y00459; SEQ ID NO: 25) từ cụm gen sinh tổng hợp streptomycin *S. griseus*; Tei15* (Số hiệu lưu giữ Ngân hàng gen AJ632270; SEQ ID NO: 32) từ cụm gen teicoplanin *Actinoplanes teichomyceticus*; Bbr (Số hiệu lưu giữ Ngân hàng gen Y16952; SEQ ID NO: 28) từ cụm gen sinh tổng hợp balhimycin DSM 5908 chủng *Amycolatopsis*; KasT (Số hiệu lưu giữ Ngân hàng gen BAF79690; SEQ ID NO: 29) từ cụm gen kasugamycin *S. kasugaensis*; NovG (Số hiệu lưu giữ Ngân hàng gen AF170880; SEQ ID NO: 30) từ cụm gen sinh tổng hợp novobiocin NCIMB 9219 chủng *S. niveus*; SgcR1 (Số hiệu lưu giữ Ngân hàng gen AY048670; SEQ ID NO: 31) từ cụm gen sinh tổng hợp C-1027 *S. globisporus*. Các axit amin đồng nhất (*), các axit amin bảo toàn (.) và sự thay thế các axit amin bảo toàn cao (:). Đặc điểm motif cấu trúc dạng xoắn-ngoặt-xoắn (helix-turn-helix: HTH) bảo toàn của protein liên kết ADN giống như StrR được gạch dưới.

FIG. 12 là sự sắp xếp Orf18 (SEQ ID NO: 36) với các gen trực giao điều hòa đáp ứng khác đặc trưng về mặt chức năng. SCO1745/AbrA2: Chất điều hòa đáp ứng hai thành phần A3(2) *S. coelicolor* (Số hiệu lưu giữ Ngân hàng gen CAB50960; SEQ ID NO: 33). SCO3226/AbsA2: chất điều hòa đáp ứng hai thành phần A3(2) *S. coelicolor* (Số hiệu lưu giữ Ngân hàng gen AAB08053; SEQ ID NO: 34). SCO3818: chất điều hòa

phiên mã đáp ứng hệ thống hai thành phần A3(2) *S. coelicolor* (Số hiệu lưu giữ Ngân hàng gen CAB46941; SEQ ID NO: 35).

Danh mục trình tự

Trình tự axit nucleic và axit amin được liệt kê ở đây và trong danh mục trình tự kèm theo được thể hiện bằng cách sử dụng các chữ viết tắt chuẩn đối với các gốc nucleotit, và mã ba ký tự đối với các axit amin, như được xác định trong 37 C.F.R. 1.822. Chỉ một sợi của trình tự axit nucleic được thể hiện, còn sợi bổ sung được hiểu là bao gồm bằng cách tham chiếu đến sợi được thể hiện. Trong danh mục trình tự kèm theo đây:

Các SEQ ID NO: 1 và 2 là đoạn mỗi oligonucleotit được sử dụng để tạo ra cài xen plasmit pXY152-endorf24.

SEQ ID NO: 3 là trình tự axit nucleic của plasmit pXY152-endorf24.

Các SEQ ID NO: 4-7 là đoạn mỗi oligonucleotit được sử dụng để tạo ra cài xen plasmit pXY300-orf18ifd.

SEQ ID NO: 8 là trình tự axit nucleic của plasmit pXY300-orf18ifd.

Các SEQ ID NO: 9 và 10 là đoạn mỗi oligonucleotit được sử dụng để tạo ra mảnh *oriT* của plasmit pKS-T-orf18pfrd.

SEQ ID NO: 11 là trình tự axit nucleic của plasmit pKS-T-orf18pfrd.

Các SEQ ID NO: 12 và 13 là đoạn mỗi oligonucleotit được sử dụng để tạo ra mảnh *amR* của plasmit pKS-T-orf18pfrd-AmR.

SEQ ID NO: 14 là trình tự axit nucleic của plasmit pKS-T-orf18pfrd-AmR.

Các SEQ ID NO: 15-18 là đoạn mỗi oligonucleotit được sử dụng để tạo ra các mảnh *oriT* và *amR* của plasmit pKS-orf18ifd-T-AmR(NS).

SEQ ID NO: 19 là trình tự axit nucleic của plasmit pKS-orf18ifd-T-AmR(NS).

SEQ ID NO: 20 là trình tự axit nucleic của plasmit pXY152-endorf24-camtsr.

Các SEQ ID NO: 21 và 22 là đoạn mỗi oligonucleotit được sử dụng để tạo ra mảnh *bla* của plasmit pXY152-endorf24-blatsr.

SEQ ID NO: 23 là trình tự axit nucleic của plasmit pXY152-endorf24-blatsr.

SEQ ID NO: 24 là đoạn mỗi oligonucleotit tương ứng với vùng gen kháng apramycin.

SEQ ID NO: 25 là trình tự axit amin của protein StrR hoạt hóa streptomycin.

SEQ ID NO: 26 là trình tự axit amin được mã hóa bởi ORF24.

SEQ ID NO: 27 là trình tự axit nucleic minh họa việc xóa đoạn không làm lệch khung dịch mã ở *orf18*.

SEQ ID NO: 28 là trình tự axit amin của cài xen Bbr.

SEQ ID NO: 29 là trình tự axit amin của cài xen KasT.

SEQ ID NO: 30 là trình tự axit amin của cài xen NovG.

SEQ ID NO: 31 là trình tự axit amin của cài xen SgcR1.

SEQ ID NO: 32 là trình tự axit amin của cài xen Teil5*.

SEQ ID NO: 33 là trình tự axit amin của gen trực giao điều hòa đáp ứng SCO1745/AbrA2 từ *S. coelicolor* A3(2) (Số hiệu lưu giữ Ngân hàng gen CAB50960).

SEQ ID NO: 34 là trình tự axit amin của gen trực giao điều hòa đáp ứng SCO/3226/AbsA2 từ *S. coelicolor* A3(2) (Số hiệu lưu giữ Ngân hàng gen AAB08053).

SEQ ID NO: 35 là trình tự axit amin của gen trực giao điều hòa đáp ứng SCO3818 từ *S. coelicolor* A3(2) (Số hiệu lưu giữ Ngân hàng gen CAB46941).

SEQ ID NO: 36 là trình tự axit amin được mã hóa bởi ORF18.

SEQ ID NO: 37 là trình tự axit nucleic của *orf18*.

SEQ ID NO: 38 là trình tự axit nucleic của *orf24*.

SEQ ID NO: 39 là trình tự axit nucleic của fosmid pXYF148 với *orf24* được đặt ở vị trí nucleotit 23109 đến 24044).

SEQ ID NO: 40 là trình tự axit nucleic của fosmid pXYF24 với *orf18* được đặt ở vị trí nucleotit 31091-31753).

Mô tả chi tiết sáng chế

I. Giới thiệu

Enduracidin (FIG. 1), còn gọi là enramycin, là thuốc kháng sinh 17 axit amin lipodepsipeptit được sản xuất bằng vi khuẩn trong đất *S. fungicidicus* B-5477 (ATCC 21013). Peptit được phân lập từ môi trường canh lên men và thể sợi chủ yếu là hỗn hợp enduracidin A và B, khác biệt ở một cacbon theo chiều dài của chuỗi lipit gắn vào. Về mặt cấu trúc, các enduracidin được phân biệt bằng nhóm axit béo C₁₂ hoặc C₁₃ mạch nhánh 2Z,4E được gắn vào bằng liên kết amit với gốc axit aspartic, và có mặt nhiều gốc axit amin không sinh protein như enduracididin (End), 4-hydroxyphenylglyxin (Hpg), 3,5-diclo-4-hydroxyphenylglyxin (Dpg), citrulin (Cit) và ornithin (Orn) (xem FIG. 1). Bảy trong số 17 axit amin có cấu hình D và sáu axit amin còn lại là Hpg hoặc Dpg dẫn xuất được clo hóa.

Enduracidin (các peptit sẽ được gọi một cách đơn giản) bộc lộ hoạt tính kháng khuẩn *in vitro* và *in vivo* hiệu nghiệm chống lại phổ rộng các sinh vật Gram dương, bao gồm *Staphylococcus aureus* kháng methixilin (methicillin-resistant *Staphylococcus aureus*: MRSA) và *Enterococcus* kháng vancomycin (vancomycin-resistant *Enterococcus*: VRE). Nồng độ ức chế tối thiểu (minimum inhibitory concentrations: MICs) là nồng độ thấp bằng 0,05 µg/mL và hiệu quả là diệt khuẩn. Nghiên cứu 100 chủng *S. aureus* thu được từ các sản phẩm bệnh lý khác nhau, và bao gồm 40% MRSA, thiết lập MIC nằm trong khoảng từ 0,09 đến 0,56 µg/mL thì không chủng nào có thể sống sót khi tiếp xúc với 1 µg/mL. Để so sánh, các MIC thông thường đối với các chủng *S. aureus* nhạy vancomycin nằm trong khoảng từ 0,5 đến 2 µg/mL. Ngoài ra, enduracidin

có profin độc tính rất tốt. Trong nghiên cứu ở chuột nhắt, thỏ, chó và khỉ, các LD₅₀ cấp tính là: trong tĩnh mạch, 30-125 mg/kg; trong màng bụng, 750-910 mg/kg; dưới da, trong cơ (intramuscular: i.m.) hoặc qua đường miệng, >5-10 g/kg. Trong nghiên cứu tương tự, khỉ nhận enduracidin i.m. trong 6 tháng và chuột cống mà được định liều tương tự trong 12 tháng được phát hiện chỉ có viêm cục bộ ở vị trí tiêm. Ở người, enduracidin được dùng i.m. (100 mg mỗi 12 giờ) cho 20 bệnh nhân là người trưởng thành nhập viện bị nhiễm MRSA. Peptit được báo cáo là không có tác dụng phụ và còn có hiệu quả cao để điều trị bệnh lây nhiễm qua đường tiêu và da do MRSA gây ra, nhưng không phải đối với bệnh lây nhiễm ở xương mạn tính (Peromet *et al.*, *Chemotherapy* 19:53-61, 1973).

Enduracidin ức chế sinh tổng hợp thành tế bào vi khuẩn peptidoglycan bằng cách tạo phức với Lipit II ngoại bào, tiền chất của cấu trúc thành tế bào vi khuẩn. Vị trí của phức Lipit II khác với vị trí được nhận diện bằng vancomycin và giải thích tác dụng của enduracidin chống lại vi khuẩn kháng vancomycin. Đến nay, không có tài liệu nào về sự kháng chéo của enduracidin với thuốc kháng sinh được sử dụng về mặt lâm sàng bất kỳ và không có dấu hiệu về cơ chế kháng phát triển, kháng mắc phải hoặc kháng chuyển. Không có dạng đã biết bất kỳ của cơ chế kháng chuyển, thiếu sinh khả dụng qua đường miệng, độc tính thấp, và hoạt tính của *Clostridium spp.* rất tốt để tạo ra enduracidin là kháng sinh peptit thương phẩm quan trọng được sử dụng làm chất phụ gia thức ăn cho gia cầm để kiểm soát bệnh viêm ruột do vi khuẩn clostridial.

Để thu được chủng của vi khuẩn sản xuất mà có thể cung cấp số lượng peptit cần để sử dụng thương mại, Japan Takeda Animal Health (hiện là một phần của Intervet/Merck Animal Health) đưa *S. fungicidicus* B-5477 vào các phương pháp cải thiện chủng truyền thống và lựa chọn các đột biến tạo ra hiệu suất enduracidin cao hơn. Thị trường trên toàn thế giới có nhu cầu về enduracidin tăng dẫn đến nỗ lực để cải thiện hơn nữa hiệu suất của thuốc kháng sinh này trong BM38-2 (ATCC PTA-122342). Với trình tự di truyền của cụm gen sinh tổng hợp enduracidin có sẵn (Số hiệu lưu giữ Ngân hàng gen DQ403252 được kết hợp ở đây để tham khảo ở dạng có sẵn trên trang tin điện tử toàn cầu vào ngày 3 tháng 10 năm 2006, BM38-2 (ATCC PTA-122342) được dùng làm chủng bắt đầu để thao tác di truyền hướng đích các gen điều hòa kết hợp với cụm

gen và tạo thành cơ sở của sáng chế này. Ở đây, sáng chế bộc lộ sản phẩm của *orf18* có tác dụng tiêu cực đến sự sản xuất enduracidin và sản phẩm gen *orf24* có tác dụng tích cực đến sự sản xuất enduracidin và các chủng tái tổ hợp có nguồn gốc từ cả sinh vật kiểu đại *S. fungicidicus* và BM38-2 (ATCC PTA-122342) mà lợi dụng các tác dụng điều hòa này để tạo ra hiệu suất enduracidin cao. Ngoài ra, sáng chế bộc lộ sự thay thế gen mới và vector biểu hiện hợp nhất lần lượt dựa vào pBluescript II KS và pSET152.

II. Chữ viết tắt và thuật ngữ

a. Chữ viết tắt

AA:	axit amin
Am:	apramycin
AmR:	chất chỉ thị kháng apramycin
<i>amRp</i> :	gen khởi đầu kháng apramycin nguyên thể
ATCC:	bộ sưu tập giống chuẩn của Mỹ
<i>bla</i> :	gen kháng ampixilin
BLAST:	công cụ tìm kiếm trình tự tương đồng cục bộ
<i>cam</i> :	gen kháng cloramphenicol
CFU	đơn vị hình thành khuẩn lạc
CTAB:	Xetyl Trimetyl Amoni Bromua
Cit:	L-citrulin
Dpg:	3,5-diclo-L-4-hydroxyphenylglyxin
EDTA:	dinatri EtylenDiaminTetra-Axetat
End:	enduracidin

Enradin:	Enduracidin, Enramycin
EPM:	môi trường sản xuất enduracidin
Hpg:	D- và L- 4-hydroxyphenylglyxin
HPLC:	Sắc ký lỏng hiệu năng cao
HTH:	xoắn-ngoặt-xoắn
IM:	trong cơ
ISP2:	phương pháp phân loại xạ khuẩn quốc tế 2
ISP4:	phương pháp phân loại xạ khuẩn quốc tế 4
LB:	môi trường canh Luria-Bertani
LD50:	liều gây chết, LD50 thể hiện liều riêng rẽ cần để diệt 50 phần trăm số lượng động vật thử nghiệm
MAH:	sức khỏe động vật Intervet/Merck
MeOH:	metanol
MICs:	nồng độ ức chế tối thiểu,
MRSA:	<i>Staphylococcus aureus</i> kháng methicillin
nm:	nanomet
NRPS:	peptit synthetaza không phải thể ribosom
ORF:	khung đọc mở
Orn:	D-ornithin
PCP:	protein mang peptidyl
PCR:	phản ứng chuỗi polymeraza

Pfrd:	xóa vùng sườn cộng (Plus Flanking Region Deletion)
SDS:	Natri Dodexyl Sulfat
SNP:	hiện tượng nhiều dạng nucleotit đơn
SPD:	quang phổ điôt
TFA:	axit trifloaxetic
TSB:	Môi trường canh tương tripsin
<i>tsr</i> :	gen kháng thiostrepton
UV:	tia cực tím
VRE:	enterococci kháng vancomycin

b. Thuật ngữ

Trừ khi có quy định khác, các thuật ngữ kỹ thuật được sử dụng theo cách dùng thông thường. Định nghĩa về các thuật ngữ thông thường trong lĩnh vực sinh học phân tử có thể được tìm thấy trong tài liệu của Benjamin Lewin *Genes V* được Oxford University Press xuất bản năm 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.) *The Encyclopedia of Molecular Biology*, được Blackwell Science Ltd. xuất bản năm 1994 (ISBN 0-632-02182-9); và Robert A. Meyers (ed.) *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, được VCH Publishers, Inc. xuất bản năm 1995 (ISBN 1-56081-569-8).

Dưới đây là các giải thích về các thuật ngữ cụ thể để thuận tiện cho việc xem xét các phương án khác nhau của sáng chế:

Cách dùng: Việc dùng theo con đường bất kỳ cho động vật. Khi được sử dụng ở đây, cách dùng thường là dùng qua đường miệng.

Biến thể alen: Dạng thay thế của polypeptit mà đặc trưng ở chỗ có sự thay thế, xóa đoạn, hoặc bổ sung một hoặc nhiều axit amin. Theo một ví dụ, biến thể không làm thay đổi chức năng sinh học của polypeptit.

Sự khuếch đại: Khi sử dụng cho các axit nucleic, các kỹ thuật làm tăng số lượng bản sao của phân tử axit nucleic trong mẫu vật hoặc tiêu bản. Ví dụ về sự khuếch đại này là phản ứng chuỗi polymeraza, trong đó mẫu sinh học thu được từ đối tượng được tiếp xúc với cặp đoạn mồi oligonucleotit, trong điều kiện cho phép lai hóa các đoạn mồi với khuôn mẫu axit nucleic trong mẫu. Các đoạn mồi được mở rộng trong các điều kiện thích hợp, được phân ly ra khỏi khuôn mẫu, và sau đó ủ lại, mở rộng, và phân ly để khuếch đại số lượng bản sao của axit nucleic. Sản phẩm của sự khuếch đại *in vitro* có thể được đặc trưng bởi hiện tượng điện di, mẫu phân cắt endonucleaza giới hạn, lai hóa hoặc thất oligonucleotit, và/hoặc xác định trình tự axit nucleic, sử dụng các kỹ thuật chuẩn. Các ví dụ khác về các kỹ thuật khuếch đại *in vitro* bao gồm sự khuếch đại thay thế sợi (xem U.S. Patent No. 5,744,311); sự khuếch đại đẳng nhiệt không phiên mã (xem U.S. Patent No. 6,033,881); sự khuếch đại phản ứng chuỗi sửa lỗi (xem WO 90/01069); sự khuếch đại phản ứng chuỗi ligaza (xem EP-A-320 308); sự khuếch đại phản ứng chuỗi ligaza lấp đầy khoảng trống (xem U.S. Patent No. 5,427,930); xóa đoạn ligaza ghép cặp và PCR (xem U.S. Patent No. 6,027,889); và sự khuếch đại không phiên mã ARN NASBA™ (xem U.S. Patent No. 6,025,134).

Chất tương tự, dẫn xuất hoặc chất bắt chước: Dạng tương tự là phân tử khác về cấu trúc hóa học so với hợp chất gốc, ví dụ chất đồng đẳng (khác về sự gia tăng cấu trúc hóa học, như sự khác biệt về độ dài của mạch alkyl), mảnh phân tử, cấu trúc mà khác một hoặc nhiều nhóm chức, và/hoặc sự thay đổi về việc ion hóa. Chất tương tự về cấu trúc thường được tìm thấy bằng cách sử dụng mối quan hệ định lượng giữa hoạt tính-cấu trúc (quantitative structure activity relationships: QSAR), bằng các kỹ thuật như được bộc lộ trong tài liệu Remington (*The Science and Practice of Pharmacology*, 19th Edition (1995), chapter 28). Khi sự thay đổi của hợp chất gốc là đáng kể, hoặc nhiều sự thay đổi tăng được kết hợp lại, thì hợp chất không còn là chất tương tự. Ví dụ, ramoplanin ở đây không được coi là chất tương tự của enduracidin: ramoplanin không có axit amin

enduracididin, bao gồm các axit amin khác nhau, và dù nó có mạch bên lipit, thì chiều dài mạch về cơ bản là ngắn. Các chất tương tự của enduracidin có thể được tạo ra bằng cách bổ sung hoặc xóa nhóm chức trên các axit amin cấu thành lipodepsipeptit, bằng cách thay thế axit amin này cho axit amin kia (ngoại trừ các axit amin enduracididin) hoặc sự kết hợp của việc biến đổi nhóm chức và thay thế axit amin. Các chất tương tự enduracidin làm ví dụ bao gồm tetrahydroenduracidin A, tetrahydroenduracidin B, deschloroenduracidin A, và deschloroenduracidin B.

Dẫn xuất là phân tử có hoạt tính sinh học có nguồn gốc từ cấu trúc gốc. Chất bắt chước là phân tử bắt chước hoạt tính của phân tử khác bằng cách bắt chước cấu trúc của phân tử đó, như phân tử có hoạt tính sinh học. Do đó, thuật ngữ "bắt chước" chỉ cấu trúc xác định liên quan đến hoạt tính.

Thuốc kháng sinh: Chất, ví dụ enduracidin, penixilin hoặc streptomycin, thường được sản xuất bằng hoặc có nguồn gốc từ một số loại nấm, vi khuẩn nhất định, và các sinh vật khác, mà có thể phá hủy hoặc ức chế sự phát triển của các vi sinh vật khác.

Đối nghĩa, có nghĩa, và đối gen: ADN sợi đôi (double-stranded DNA: dsADN) có hai sợi, sợi $5' \rightarrow 3'$, là ở dạng sợi cộng, và sợi $3' \rightarrow 5'$ (bổ sung ngược), là ở dạng sợi trừ. Bởi vì ARN polymeraza thêm các axit nucleic theo hướng $5' \rightarrow 3'$, nên sợi ADN trừ dùng làm khuôn mẫu cho ARN trong quá trình phiên mã. Do đó, ARN tạo thành sẽ có bổ sung trình tự vào sợi trừ và giống hệt sợi cộng (ngoại trừ U được thay thế cho T). Các phân tử đối nghĩa là các phân tử có thể lai hóa đặc hiệu hoặc bổ sung đặc hiệu vào ARN hoặc ADN sợi cộng. Các phân tử có nghĩa là các phân tử có thể lai hóa đặc hiệu hoặc bổ sung đặc hiệu vào sợi ADN trừ. Các phân tử đối gen là các phân tử đối nghĩa hoặc có nghĩa bổ sung vào đích dsADN. Theo một phương án, phân tử đối nghĩa lai hóa đặc hiệu với mARN đích và ức chế sự phiên mã mARN đích.

Liên kết hoặc liên kết bền: Phân tử, như oligonucleotit hoặc protein, liên kết hoặc liên kết bền với phân tử đích, như axit nucleic hoặc protein đích, nếu liên kết có thể được phát hiện. Theo một ví dụ, oligonucleotit liên kết hoặc liên kết bền với axit nucleic đích nếu lượng oligonucleotit đủ để tạo thành các cặp bazơ hoặc được lai hóa thành axit

nucleic của nó, cho phép phát hiện ra liên kết. Liên kết có thể được phát hiện bởi các đặc điểm vật lý hoặc chức năng của đích: phức oligonucleotit. Liên kết giữa đích và oligonucleotit có thể được phát hiện bằng phương pháp bất kỳ đã biết đối với người có hiểu biết trung bình về lĩnh vực kỹ thuật này, bao gồm cả thử nghiệm liên kết chức năng và vật lý. Liên kết có thể được phát hiện về mặt chức năng bằng cách xác định xem liên kết đó có tác dụng đáng kể trong quy trình sinh tổng hợp như sự biểu hiện gen, sao chép ADN, phiên mã, dịch mã và các quy trình tương tự hay không.

Các phương pháp vật lý phát hiện liên kết của sợi ADN hoặc ARN bổ sung đã biết rõ trong lĩnh vực này, và bao gồm các phương pháp như DNase I hoặc in dấu chân hóa học, thử nghiệm chuyển gel và thử nghiệm phân cắt ái lực, phương pháp lai thẩm tách Bắc (Northern blotting), phương pháp lai điểm (dot blotting) và phương pháp phát hiện dựa vào độ hấp thụ ánh sáng. Ví dụ, một phương pháp mà được sử dụng rộng rãi, vì nó rất đơn giản và đáng tin cậy, bao gồm bước quan sát sự thay đổi độ hấp thụ ánh sáng của dung dịch chứa oligonucleotit (hoặc chất tương tự) và axit nucleic đích ở 220 đến 300 nm khi nhiệt độ tăng chậm. Nếu oligonucleotit hoặc chất tương tự liên kết với đích của nó, thì độ hấp thụ tăng đột ngột ở nhiệt độ đặc trưng khi oligonucleotit (hoặc chất tương tự) và đích của nó tách ra khỏi nhau, hoặc tan ra.

Liên kết giữa oligome và axit nucleic đích của nó thường xuyên được đặc trưng bởi nhiệt độ (T_m) tại đó 50% oligome tan ra khỏi đích của nó. T_m cao hơn nghĩa là phức mạnh hơn hoặc bền hơn so với phức có T_m thấp hơn.

Liên kết giữa protein và protein đích của nó, như kháng thể của kháng nguyên thường xuyên được đặc trưng bởi sự xác định ái lực liên kết. Theo một phương án, ái lực được tính bằng sự thay đổi của phương pháp Scatchard được mô tả bởi Frankel *et al.*, *Mol. Immunol.*, 16:101-106, 1979. Theo phương án khác, ái lực liên kết được đo bằng tốc độ phân tách thụ thể chất liên kết đặc hiệu. Theo phương án khác nữa, ái lực liên kết cao được đo bằng thử nghiệm miễn dịch phóng xạ cạnh tranh. Trong một vài ví dụ, ái lực liên kết cao bằng ít nhất là khoảng 1×10^{-8} M. Theo các phương án khác, ái lực liên kết cao bằng ít nhất là khoảng $1,5 \times 10^{-8}$, ít nhất là khoảng $2,0 \times 10^{-8}$, ít nhất là khoảng

$2,5 \times 10^{-8}$, ít nhất là khoảng $3,0 \times 10^{-8}$, ít nhất là khoảng $3,5 \times 10^{-8}$, ít nhất là khoảng $4,0 \times 10^{-8}$, ít nhất là khoảng $4,5 \times 10^{-8}$, hoặc ít nhất là khoảng $5,0 \times 10^{-8}$ M.

Chức năng sinh học: (Các) chức năng của polypeptit trong các tế bào trong đó polypeptit có mặt tự nhiên. Polypeptit có thể có nhiều hơn một chức năng sinh học.

cADN (ADN bổ sung: complementary DNA): Một phần ADN thiếu các phần không mã hóa, bên trong (intron) và trình tự điều hòa phiên mã. cADN cũng có thể chứa các vùng không dịch mã (untranslated regions: UTRs) chịu trách nhiệm điều chỉnh dịch mã trong phân tử ARN tương ứng. cADN được tổng hợp trong phòng thí nghiệm bằng cách phiên mã ngược từ ARN thông tin được chiết xuất từ tế bào.

Thay thế bảo toàn: Sự thay thế axit amin gần như không làm thay đổi hoạt tính (tính đặc hiệu hoặc ái lực liên kết) của phân tử. Sự thay thế axit amin bảo toàn điển hình bao gồm sự thay thế axit amin này cho axit amin khác có cùng đặc điểm hóa học (ví dụ, điện tích hoặc tính kỵ nước). Bảng sau thể hiện sự thay thế axit amin bảo toàn làm mẫu:

Gốc ban đầu	Sự thay thế bảo toàn
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr

Gốc ban đầu	Sự thay thế bảo toàn
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Chủng *Streptomyces fungicidicus* đối chứng: Chủng kiểu dại xuất hiện tự nhiên, *Streptomyces fungicidicus* ATCC21013.

ADN (axit deoxyribonucleic): Polyme mạch dài chứa vật liệu di truyền của hầu hết sinh vật sống (một số virus có gen chứa axit ribonucleic (ARN)). Các đơn vị lặp lại trong các polyme ADN là bốn nucleotit khác nhau, mỗi nucleotit chứa một một trong số bốn bazơ, adenin, guanin, xytosin và thymin liên kết với đường deoxyriboza mà nhóm phosphat gắn vào. Bộ ba các nucleotit (gọi là codon) mã hóa mỗi axit amin trong polypeptit. Thuật ngữ codon còn được sử dụng đối với các trình tự tương ứng (và bổ sung) của ba nucleotit trong mARN vào vị trí mà trình tự ADN được dịch mã.

Trừ khi có quy định khác, sự tham chiếu bất kỳ đến phân tử ADN được dự tính bao gồm phần bổ sung ngược của phân tử ADN đó. Ngoại trừ trạng thái sợi đơn được yêu cầu trong sáng chế này, các phân tử ADN, dù chỉ để mô tả sợi đơn, bao gồm cả hai sợi của phân tử ADN sợi kép. Do đó, khi tham chiếu đến phân tử axit nucleic mã hóa protein đặc hiệu, hoặc mảnh của nó, thì bao gồm cả sợi có nghĩa và phần bổ sung ngược của nó. Do đó, chẳng hạn như thích hợp để tạo ra đoạn dò hoặc đoạn mồi từ trình tự bổ sung ngược của các phân tử axit nucleic đã bộc lộ.

Miền: Một phần của phân tử như các protein hoặc axit nucleic khác về mặt cấu trúc và/hoặc chức năng với phần khác của phân tử.

Mã hóa: Polynucleotit được gọi là "mã hóa" polypeptit nếu, ở trạng thái nguyên thể của nó hoặc khi được thao tác bằng các phương pháp mà người có hiểu biết trung bình về lĩnh vực kỹ thuật này biết rõ, thì có thể được phiên mã và/hoặc dịch mã để tạo

ra mARN cho và/hoặc polypeptit hoặc mảnh của nó. Sợi đối nghĩa là phần bổ sung của axit nucleic này, và trình tự mã hóa có thể được tạo ra từ đó.

Enduracidin: Các enduracidin A và B là 17 lipodepsipeptit axit amin được tìm ra vào cuối những năm 1960 từ việc lên men vi khuẩn trong đất *Streptomyces fungicidicus* B-5477 (ATCC 21013). Các peptit A và B là các chất đồng đẳng khác nhau bởi một cacbon theo chiều dài của chuỗi lipit gắn vào. Về mặt cấu trúc, các enduracidin được phân biệt bằng gốc axit béo C₁₂ hoặc C₁₃ mạch nhánh 2Z,4E và có mặt nhiều gốc axit amin không sinh protein như enduracididin (End), 4-hydroxyphenylglyxin (Hpg), 3,5-diclo-4-hydroxyphenylglyxin (Dpg), citrulin (Cit) và ornithin (Orn). Bảy trong số 17 axit amin có cấu hình D và sáu axit amin còn lại là Hpg hoặc Dpg tương tự được clo hóa.

Các mảnh chức năng và biến thể của polypeptit: Bao gồm các mảnh và biến thể này duy trì một hoặc nhiều chức năng của polypeptit gốc. Nhận ra rằng gen hoặc cADN mã hóa polypeptit có thể bị biến đổi đáng kể mà chủ yếu không làm thay đổi chức năng của một hoặc nhiều polypeptit. Thứ nhất, mã di truyền có tính thoái hóa, và do đó các codon khác nhau mã hóa các axit amin giống nhau. Thứ hai, ngay cả khi đưa vào sự thay thế axit amin, thì đột biến có thể là bảo toàn và không có tác động đáng kể đến (các) chức năng cần thiết của protein. Xem tài liệu của Stryer *Biochemistry 3rd Ed.*, (c) 1988. Thứ ba, một phần của chuỗi polypeptit có thể bị xóa mà không làm giảm hoặc loại bỏ tất cả các chức năng của nó. Thứ tư, việc cài xen hoặc bổ sung có thể được thực hiện trong chuỗi polypeptit, ví dụ, bổ sung các đuôi epitop, mà không làm giảm hoặc loại bỏ các chức năng của nó (Ausubel *et al. J. Immunol.* 159(5): 2502-12, 1997). Các biến đổi khác có thể được thực hiện mà chủ yếu không làm giảm một hoặc nhiều chức năng của polypeptit bao gồm, ví dụ, các biến đổi hóa học và hóa sinh *in vivo* hoặc *in vitro* hoặc sự kết hợp các axit amin khác thường. Các biến đổi này bao gồm, ví dụ, axetyl hóa, carboxyl hóa, phosphoryl hóa, glycosyl hóa, ubiquitin hóa, đánh dấu, ví dụ, bằng nuclit phóng xạ, và các biến đổi enzym khác nhau, sẽ dễ dàng đạt được bởi người có hiểu biết trung bình về lĩnh vực kỹ thuật này. Các phương pháp khác nhau để đánh dấu polypeptit, và các ký hiệu đánh dấu là hữu dụng cho các mục đích này, bao gồm các chất đồng vị phóng xạ như ³²P, các phối tử mà liên kết với hoặc được liên kết bằng các thành phần liên kết đặc

hiệu được đánh dấu (ví dụ, kháng thể), thuốc nhuộm huỳnh quang (fluorophore), chất phát quang hóa học, enzym, và kháng phôi tử. Các mảnh chức năng và các biến thể có thể có chiều dài thay đổi. Ví dụ, một số mảnh có ít nhất 10, 25, 50, 75, 100, 200 gốc axit amin, hoặc thậm chí nhiều hơn.

Lượng có hiệu quả: Lượng hoặc nồng độ của hợp chất hoặc chế phẩm đặc hiệu đủ để đạt được hiệu quả mong muốn ở đối tượng. Lượng có hiệu quả có thể phụ thuộc ít nhất một phần vào các loài động vật cần điều trị, kích thước của động vật, và/hoặc bản chất của hiệu quả mong muốn.

Cụm gen: Tập hợp các thành phần di truyền được nhóm với nhau trên nhiễm sắc thể, các sản phẩm protein của tập hợp này có chức năng liên quan, như tạo thành con đường sinh tổng hợp sản phẩm tự nhiên.

Khác loại: Khi đề cập đến các trình tự axit nucleic như trình tự mã hoá và trình tự kiểm soát, thì “khác loại” chỉ các trình tự thường không được kết hợp với vùng cấu trúc tái tổ hợp, và/hoặc thường không được kết hợp với tế bào cụ thể. Do đó, vùng “khác loại” của cấu trúc axit nucleic là đoạn axit nucleic nhận biết được trong hoặc được gắn với phân tử axit nucleic khác mà không tìm thấy khi kết hợp với phân tử khác trong tự nhiên. Ví dụ, vùng khác loại của cấu trúc có thể bao gồm trình tự mã hoá được tạo bên sườn bằng các trình tự không tìm thấy khi kết hợp với trình tự mã hoá trong tự nhiên. Ví dụ khác về trình tự mã hoá khác loại là cấu trúc trong đó bản thân trình tự mã hoá không được tìm thấy trong tự nhiên (ví dụ, trình tự tổng hợp có các codon khác so với gen nguyên bản). Tương tự, tế bào chủ được biến nạp với cấu trúc mà thường không có mặt trong tế bào chủ sẽ được coi là dạng khác loại đối với các mục đích của sáng chế.

Trình tự axit amin tương đồng: Polypeptit bất kỳ mà được mã hoá, toàn bộ hoặc một phần, bởi trình tự axit nucleic lai hoá với phần bất kỳ của trình tự axit nucleic vùng mã hoá. Trình tự axit amin tương đồng là trình tự khác với trình tự axit amin thể hiện trong danh mục trình tự ở một hoặc nhiều sự thay thế axit amin bảo toàn. Trình tự này còn bao gồm các biến thể alen (được xác định ở trên) cũng như các trình tự chứa sự xoá đoạn hoặc cài xen mà giữ các đặc tính chức năng của polypeptit. Tốt hơn là, trình tự này

đồng nhất ít nhất là 75%, tốt hơn nữa là 80%, tốt hơn nữa là 85%, tốt hơn nữa là 90%, tốt hơn nữa là 95%, tốt nhất là 98% với trình tự bất kỳ trong số các trình tự axit amin.

Các trình tự axit amin tương đồng bao gồm các trình tự đồng nhất hoặc gần như đồng nhất với các trình tự axit amin của danh mục trình tự. "Trình tự axit amin gần như đồng nhất" nghĩa là trình tự đồng nhất ít nhất là 90%, tốt hơn là 95%, tốt hơn nữa là 97%, tốt nhất là 99% với trình tự axit amin tham chiếu và tốt hơn là khác trình tự tham chiếu chủ yếu ở sự thay thế axit amin bảo toàn. Phù hợp với khía cạnh này của sáng chế, các polypeptit có trình tự tương đồng với trình tự bất kỳ trong số các trình tự axit amin của danh mục trình tự bao gồm các biến thể alen xuất hiện tự nhiên, cũng như các đột biến hoặc các biến thể có không tự nhiên khác bất kỳ mà giữ được đặc tính vốn có của polypeptit bất kỳ của các trình tự được bộc lộ ở đây. Độ tương đồng có thể được xác định bằng phần mềm phân tích trình tự như Gói phần mềm phân tích trình tự của The Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705. Các trình tự axit amin có thể được sắp xếp để tối đa hoá độ đồng nhất. Các khoảng trống cũng có thể được đưa vào trình tự bằng cách nhân tạo để đạt được sự sắp xếp tối ưu. Khi sự sắp xếp tối ưu được cài đặt, mức độ tương đồng được thiết lập bằng cách ghi lại tất cả các vị trí trong đó các axit amin của cả hai trình tự là đồng nhất, so với tổng số vị trí. Các trình tự polynucleotit tương đồng được xác định theo cách tương tự. Tốt hơn là, trình tự tương đồng là trình tự đồng nhất ít nhất là 45%, 50%, 60%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% với trình tự bất kỳ trong số các trình tự mã hoá.

Lai hoá: Các oligonucleotit và các axit nucleic khác lai hoá bằng cách liên kết hydro, mà bao gồm liên kết Watson-Crick, liên kết hydro Hoogsteen hoặc liên kết hydro Hoogsteen ngược, giữa các bazơ bổ sung. Nói chung, axit nucleic bao gồm các bazơ nitơ là pyrimidin (xytosin (C), uraxil (U), và thymin (T)) hoặc purin (adenin (A) và guanin (G)). Các bazơ nitơ này tạo thành liên kết hydro giữa pyrimidin và purin, và liên kết giữa pyrimidin với purin được gọi là cặp bazơ. Đặc biệt hơn, A sẽ liên kết hydro với T hoặc U, và G sẽ liên kết với C. Phần bổ sung là cặp bazơ xuất hiện giữa hai trình tự axit nucleic riêng biệt hoặc hai vùng riêng biệt của cùng trình tự axit nucleic.

Phần bổ sung có thể lai hoá đặc hiệu và phần bổ sung đặc hiệu là các thuật ngữ cho thấy mức độ bổ sung đủ để xuất hiện liên kết bền và đặc hiệu giữa axit nucleic thứ nhất (như, oligonucleotit) và đích ADN hoặc ARN. Axit nucleic thứ nhất (như, oligonucleotit) không cần bổ sung 100% vào trình tự đích của nó để có thể lai hoá đặc hiệu. Axit nucleic thứ nhất (như, oligonucleotit) có thể lai hoá đặc hiệu khi có mức độ bổ sung đủ để tránh liên kết không đặc hiệu của axit nucleic thứ nhất (như, oligonucleotit) với các trình tự không phải đích trong các điều kiện mà liên kết đặc hiệu được mong muốn. Liên kết này được gọi là sự lai hoá đặc hiệu.

Các điều kiện lai hoá dẫn đến các mức độ đặc biệt nghiêm ngặt sẽ thay đổi phụ thuộc vào bản chất của phương pháp lai hóa lựa chọn và chế phẩm và chiều dài của trình tự axit nucleic lai hoá. Nói chung, nhiệt độ của sự lai hoá và cường độ ion (đặc biệt là nồng độ Na^+) của đệm lai hóa sẽ xác định mức độ nghiêm ngặt của sự lai hoá, dù thời gian rửa vẫn ảnh hưởng đến sự nghiêm ngặt. Sự tính toán liên quan đến các điều kiện lai hoá được yêu cầu để đạt đến các mức độ đặc biệt nghiêm ngặt được thảo luận bởi Sambrook *et al.* (ed.) *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, chapters 9 and 11.

Sau đây là tập hợp các điều kiện lai hoá làm ví dụ và không nhằm mục đích làm giới hạn sáng chế.

Mức độ nghiêm ngặt rất cao (phát hiện các trình tự có chung độ đồng nhất trình tự 90%)

Lai hoá: 5x SSC ở 65EC trong 16 giờ

Rửa hai lần: 2x SSC ở nhiệt độ trong phòng (RT) trong mỗi lần 15 phút

Rửa hai lần: 0,5x SSC ở 65EC trong mỗi lần 20 phút

Mức độ nghiêm ngặt cao (phát hiện các trình tự có chung độ đồng nhất trình tự 80% hoặc lớn hơn)

Lai hoá: 5x-6x SSC ở 65EC-70EC trong 16-20 giờ

Rửa hai lần: 2x SSC ở RT trong mỗi lần 5-20 phút

Rửa hai lần: 1x SSC ở 55EC-70EC trong mỗi lần 30 phút

Mức độ nghiêm ngặt thấp (phát hiện các trình tự có chung độ đồng nhất trình tự lớn hơn 50%)

Lai hoá: 6x SSC ở RT đối với 55EC trong 16-20 giờ

Rửa ít nhất hai lần: 2x-3x SSC ở RT đối với 55EC trong mỗi lần 20-30 phút.

Phân lập: Thành phần sinh học phân lập (như phân tử axit nucleic hoặc protein) là thành phần về cơ bản được tách hoặc tinh chế ra khỏi các thành phần sinh học khác trong tế bào của sinh vật trong đó thành phần này xuất hiện tự nhiên, như ADN và ARN ngoài nhiễm sắc thể và nhiễm sắc thể khác, protein và cơ quan tế bào. Đối với các axit nucleic và/hoặc polypeptit, thuật ngữ này có thể đề cập đến các axit nucleic hoặc polypeptit mà không còn được gắn bên sườn bởi các trình tự thường gắn sườn chúng trong tự nhiên. Các axit nucleic và protein được phân lập bao gồm các axit nucleic và protein được tinh chế bằng các phương pháp tinh chế chuẩn. Thuật ngữ này còn bao gồm các axit nucleic và protein được tạo ra bằng cách biểu hiện tái tổ hợp trong tế bào chủ cũng như các axit nucleic được tổng hợp về mặt hoá học.

Đột biến: Quy trình gây ra sự thay đổi về trình tự của vật liệu di truyền (thường là ADN hoặc ARN) của tế bào hoặc sinh vật. Đột biến có thể cố tình được đưa vào vật liệu di truyền bằng cách sử dụng các kỹ thuật phân tử đã biết rõ trong lĩnh vực này (ví dụ, gây đột biến vị trí, gây đột biến PCR và các kỹ thuật khác).

Peptit không phải ribosom (Nonribosomal peptide: NRP): Nhóm chất chuyển hoá thứ cấp, thường tạo ra bởi vi sinh vật, như vi khuẩn và nấm. Không giống các polypeptit được tổng hợp trên ribosom, các peptit này được tổng hợp bằng peptit không phải ribosom synthetaza (nonribosomal peptide synthetases: NRPS) từ các axit amin.

Bộ phận khung chính peptit không phải ribosom: Bước thứ hai trong quy trình sinh tổng hợp peptit không phải ribosom, bao gồm sự hình thành (kết tụ) liên kết amit của trình tự peptit.

Peptit không phải ribosom synthetaza (NRPS): Protein đa chức năng lớn mà tổng hợp polypeptit bằng cơ chế không phải ribosom, thường được biết đến là quy trình tổng hợp thiotemplat (Kleinkauf and von Doehren *Ann. Rev. Microbiol.* 41: 259-289, 1987). Các polypeptit không phải ribosom này có thể có cấu trúc mạch thẳng, vòng, hoặc vòng có nhánh và thường chứa các axit amin không có mặt trong protein hoặc axit amin bị biến đổi do quá trình methyl hoá hoặc epime hoá. Ví dụ cụ thể, NRPS tạo ra dipeptit.

Sự biến đổi peptit không phải ribosom: Bước thứ ba trong quy trình sinh tổng hợp peptit không phải ribosom. Có nhiều axit amin tiền chất mới được tìm thấy trong peptit không phải ribosom và nhiều đơn thể được tạo ra hoặc được biến đổi trong khi gắn với các miền PCP của protein đặc hiệu hoặc NRPS. Sự biến đổi sau khi tổng hợp có thể xuất hiện sau khi hình thành liên kết amit của khung chính peptit. Các biến đổi làm ví dụ bao gồm epime hoá α -cacbon, *N*-methyl hoá, tạo dị vòng của các gốc Cys hoặc Ser/Thr với thiazolin và oxazolin, và halogen hoá hoặc hydroxyl hoá mạch bên. Các biến đổi khác như oxy hoá, alkyl hoá, axyl hoá và glycosyl hóa có thể xuất hiện sau khi giải phóng peptit mới sinh ra khỏi phức NRPS và thường cần cho toàn bộ hoạt tính sinh học.

Sinh tổng hợp axit amin tiền chất không phải ribosom: Bước thứ nhất trong quy trình sinh tổng hợp peptit không phải ribosom. Peptit không phải ribosom thường có các axit amin không tìm thấy trong các peptit và protein được gắn trên các ribosom. Các axit amin không sinh protein góp phần vào sự đa dạng của các peptit này và thường đóng vai trò trong hoạt tính sinh học của chúng. Quá trình sinh tổng hợp các axit amin này có thể xảy ra qua chất trung gian liên kết protein hoặc ở dạng các loại hoà tan, tự do.

Axit nucleic: Polyme deoxyribonucleotit hoặc ribonucleotit ở dạng sợi đơn hoặc sợi đôi, và trừ khi có giới hạn khác, bao gồm các chất tương tự nucleotit tự nhiên đã biết mà lai hoá với các axit nucleic theo cách tương tự như các nucleotit xuất hiện tự nhiên.

Nucleotit: Thuật ngữ này bao gồm, nhưng không bị giới hạn ở, monome chứa bazơ liên kết với đường, như pyrimidin, purin hoặc các chất tương tự tổng hợp của chúng, hoặc bazơ liên kết với axit amin, như trong axit nucleic peptit. Nucleotit là một monome trong polynucleotit. Trình tự nucleotit là trình tự của các bazơ trong polynucleotit.

Oligonucleotit: Nhiều nucleotit nối được nối bằng các liên kết phosphodiester nguyên thể, có chiều dài nằm trong khoảng từ 6 đến 300 nucleotit. Chất tương tự oligonucleotit là các gốc có chức năng tương tự với các oligonucleotit nhưng có các phần xuất hiện không tự nhiên. Ví dụ, các chất tương tự oligonucleotit có thể bao gồm các phần xuất hiện không tự nhiên, như các gốc đường thay thế hoặc liên kết giữa các đường (inter-sugar), như phosphorothioat oligodeoxynucleotit. Các chất tương tự chức năng của polynucleotit xuất hiện tự nhiên có thể liên kết với ARN hoặc ADN, và bao gồm các phân tử axit nucleic peptit.

Các oligonucleotit cụ thể và các chất tương tự oligonucleotit có thể bao gồm các trình tự mạch thẳng có chiều dài lên đến khoảng 200 nucleotit, ví dụ, trình tự (như ADN hoặc ARN) dài ít nhất là 6 bazơ, ví dụ, ít nhất là 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 hoặc thậm chí là 200 bazơ, hoặc nằm trong khoảng từ 6 đến 50 bazơ, ví dụ, khoảng từ 10 đến 25 bazơ, như 12, 15, hoặc 20 bazơ.

Khung đọc mở (ORF): Một dãy bộ ba nucleotit (các codon) mã hoá các axit amin mà không có các codon kết thúc bên trong bất kỳ. Các trình tự này thường có thể dịch mã thành peptit. Ví dụ, ORF, khung đọc mở, và enduracidin ORF là khung đọc mở trong cụm gen sinh tổng hợp enduracidin vì được phân lập từ *Streptomyces fungicidicus*. Thuật ngữ này còn bao gồm các ORF giống nhau có mặt trong các sinh vật tổng hợp bằng enduracidin khác. Thuật ngữ này còn bao gồm các biến thể alen và hiện tượng nhiều dạng nucleotit đơn (single nucleotide polymorphisms: SNPs). Trong một số trường hợp nhất định, thuật ngữ enduracidin ORF được sử dụng đồng nghĩa với polypeptit được mã hóa bởi enduracidin ORF và có thể bao gồm sự thay thế bảo toàn trong polypeptit đó. Việc sử dụng cụ thể sẽ được làm rõ theo ngữ cảnh.

Khung đọc mở mà đã bị bất hoạt là khung đọc mở được làm cho không có chức năng thông qua việc xoá đoạn, cài xen hoặc gây đột biến một hoặc nhiều nucleotit trong trình tự mã hoá.

Streptomyces fungicidicus bao gồm khung đọc mở-18 (*orf18*) giảm là sinh vật bị giảm, như giảm 2 lần, hoặc thậm chí mất hoàn toàn chức năng sinh học của sản phẩm gen *orf18*, so với *Streptomyces fungicidicus* kiểu dại, ví dụ, thông qua sự biến đổi gen của *orf18*, bao gồm *orf18* bị bất hoạt như được minh hoạ bằng ví dụ dưới đây, và/hoặc thông qua thao tác điều hoà, ví dụ, biến đổi, cài xen vào, loại bỏ, và/hoặc thay thế các vùng không mã hoá của gen mã hoá ORF18 dẫn đến giảm sự biểu hiện của sản phẩm gen *orf18*. Ví dụ, gen khởi đầu kiểu dại của *orf18* có thể được biến đổi sao cho về cơ bản làm giảm sự phiên mã của *orf18*.

Streptomyces fungicidicus bao gồm khung đọc mở-24 (*orf24*) tăng là sinh vật tăng, như gấp 2 lần hoặc nhiều hơn, về chức năng sinh học của sản phẩm gen *orf24*, so với *Streptomyces fungicidicus* kiểu dại, ví dụ, thông qua sự biến đổi gen của *orf24* để tăng cường chức năng sinh học của sản phẩm gen *orf24* và/hoặc bằng cách thao tác điều hoà, ví dụ, biến đổi, cài xen vào, loại bỏ, và/hoặc thay thế các vùng không mã hoá của gen mã hoá ORF24 dẫn đến tăng sự biểu hiện của sản phẩm gen *orf24*. Ví dụ, gen khởi đầu kiểu dại đối với *orf24* được thay thế bằng gen khởi đầu cơ định mạnh giúp tăng cường sự phiên mã của *orf24*, như được minh hoạ bằng ví dụ dưới đây.

Gen biến đổi: Trình tự gen bao gồm sự biến đổi khi so sánh với trình tự gen được tìm thấy trong gen xuất hiện tự nhiên (kiểu dại).

Liên kết theo cách có điều khiển: Trình tự axit nucleic thứ nhất được liên kết theo cách có điều khiển với trình tự axit nucleic thứ hai khi trình tự axit nucleic thứ nhất được đặt trong mối liên hệ chức năng với trình tự axit nucleic thứ hai. Chẳng hạn, gen khởi đầu được liên kết theo cách có điều khiển với trình tự mã hoá nếu gen khởi đầu ảnh hưởng đến sự phiên mã hoặc sự biểu hiện của trình tự mã hoá. Nói chung, trình tự ADN liên kết theo cách có điều khiển là liên tục và, khi cần thiết để nối hai vùng mã hoá protein, trong cùng khung đọc.

Gen trực giao: Hai trình tự axit nucleic hoặc axit amin là gen trực giao của nhau nếu chúng dùng chung trình tự di truyền thông thường và phân kỳ khi các loài mang trình tự di truyền tách thành hai loài. Trình tự gen trực giao cũng là trình tự tương đồng.

Polypeptit: Polyme trong đó các monome là các gốc axit amin được nối với nhau qua liên kết amit. Khi các axit amin là axit amin alpha, thì có thể sử dụng chất đồng phân quang L hoặc chất đồng phân quang D, các chất đồng phân L được ưu tiên trong một số trường hợp. Thuật ngữ polypeptit hoặc protein như được sử dụng ở đây bao gồm trình tự axit amin bất kỳ và bao gồm các trình tự biến đổi như glycoprotein. Đặc biệt, thuật ngữ polypeptit được dự tính bao gồm các protein có trong tự nhiên (dù được sản xuất bằng cơ chế ribosom hay không phải ribosom), cũng như các protein được sản xuất tái tổ hợp hoặc tổng hợp.

Thuật ngữ mảnh polypeptit là một phần của polypeptit biểu hiện ít nhất một epitop hữu dụng. Cụm từ mảnh chức năng của polypeptit là tất cả các mảnh polypeptit giữ lại hoạt tính (như hoạt tính sinh học), hoặc phần đo được của hoạt tính, của polypeptit mà mảnh được lấy từ đó. Ví dụ, các mảnh có thể thay đổi kích thước từ mảnh polypeptit nhỏ bằng epitop có khả năng liên kết phân tử kháng thể với polypeptit lớn có khả năng tham gia vào sự cảm ứng đặc trưng hoặc lập chương trình thay đổi kiểu hình trong tế bào.

Thuật ngữ polypeptit gần như tinh khiết như được sử dụng ở đây là polypeptit gần như không chứa các protein, lipid, carbohydrat khác hoặc vật liệu khác mà được kết hợp tự nhiên với polypeptit này. Theo một phương án, polypeptit không chứa ít nhất là 50%, ví dụ, không chứa ít nhất là 80% protein, lipid, carbohydrat khác hoặc vật liệu khác mà được kết hợp tự nhiên với polypeptit này. Theo phương án khác, polypeptit không chứa ít nhất là 90% protein, lipid, carbohydrat khác hoặc vật liệu khác mà được kết hợp tự nhiên với polypeptit này. Theo phương án khác nữa, polypeptit không chứa ít nhất là 95% protein, lipid, carbohydrat khác hoặc vật liệu khác mà được kết hợp tự nhiên với polypeptit này.

Đoạn dò và đoạn mồi: Đoạn dò và đoạn mồi axit nucleic có thể dễ dàng được điều chế dựa vào các phân tử axit nucleic được đề xuất trong bản mô tả này. Đoạn dò bao

gồm axit nucleic được phân lập gắn với chất đánh dấu có thể phát hiện hoặc phân tử thông báo. Các chất đánh dấu điển hình bao gồm chất đồng vị phóng xạ, chất nền enzym, các đồng yếu tố, phối tử, chất phát quang hoá học hoặc chất huỳnh quang, bán kháng nguyên (hapten), và enzym. Các phương pháp đánh dấu và hướng dẫn lựa chọn chất đánh dấu thích hợp cho các mục đích khác nhau được thảo luận, ví dụ, trong tài liệu của Sambrook *et al.* (Trong *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) và Ausubel *et al.* (Trong *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

Các đoạn mồi là các phân tử axit nucleic ngắn, tốt hơn là ADN oligonucleotit, dài 10 nucleotit hoặc hơn. Tốt hơn nữa là, ADN oligonucleotit dài hơn có thể là dài khoảng 15, 17, 20, hoặc 23 nucleotit hoặc hơn. Các đoạn mồi có thể được ủ với sợi ADN đích bổ sung bằng cách lai hoá axit nucleic để tạo ra thể lai giữa đoạn mồi và sợi ADN đích, và sau đó mở rộng đoạn mồi dọc theo sợi ADN đích bằng enzym ADN polymeraza. Các cặp đoạn mồi có thể được sử dụng để khuếch đại trình tự axit nucleic, ví dụ, bằng phản ứng chuỗi polymeraza (PCR) hoặc bằng các phương pháp khuếch đại axit nucleic khác đã biết trong lĩnh vực này.

Các phương pháp để điều chế và sử dụng đoạn dò và đoạn mồi được mô tả, ví dụ, trong tài liệu của Sambrook *et al.* (Trong *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989), Ausubel *et al.* (Trong *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1998), và Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). Các cặp đoạn mồi PCR có thể có nguồn gốc từ trình tự đã biết, ví dụ, bằng cách sử dụng các chương trình máy tính nhằm mục đích đó như Primers (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). Tính đặc hiệu của đoạn dò hoặc đoạn mồi cụ thể tăng theo chiều dài của nó. Do đó, để thu được tính đặc hiệu tốt hơn, đoạn dò và đoạn mồi có thể được chọn để bao gồm ít nhất 17, 20, 23, 25, 30, 35, 40, 45, 50 hoặc hơn 50 nucleotit liên tục của trình tự nucleotit mong muốn.

Protein: Phân tử sinh học được biểu hiện bằng gen và bao gồm các axit amin.

Tinh chế: Thuật ngữ tinh chế không yêu cầu độ tinh khiết tuyệt đối; đúng hơn là, đó là thuật ngữ tương đối. Do đó, ví dụ, việc điều chế protein tinh chế là quy trình trong đó protein được đề cập đến là tinh khiết hơn protein trong môi trường tự nhiên của nó trong tế bào.

Tái tổ hợp: Axit nucleic có trình tự mà không xuất hiện tự nhiên hoặc có trình tự mà được tạo ra bằng sự kết hợp nhân tạo của hai đoạn trình tự phân tách khác. Sự kết hợp nhân tạo này có thể được thực hiện bằng cách tổng hợp hoá học hoặc, thông thường hơn, bằng cách thao tác nhân tạo các đoạn phân lập của axit nucleic, ví dụ, bằng các kỹ thuật thiết kế di truyền. “Tái tổ hợp” còn được sử dụng để mô tả các phân tử axit nucleic được thao tác nhân tạo, nhưng chứa các trình tự đối chứng và các vùng mã hoá giống nhau được tìm thấy trong sinh vật mà gen được phân lập từ sinh vật này.

Điều chỉnh sự sản xuất kháng sinh: Để tạo ra sự thay đổi, như làm tăng hoặc giảm, về lượng, loại hoặc chất lượng sản xuất kháng sinh. Được mô tả ở đây là các chủng tái tổ hợp của *Streptomyces fungicidicus* với sự sản xuất enduracidin tăng cường.

Độ đồng nhất trình tự: Sự giống nhau giữa hai trình tự axit nucleic hoặc giữa hai trình tự axit amin được biểu hiện về mức độ đồng nhất trình tự chung giữa các trình tự. Độ đồng nhất trình tự thường được biểu hiện theo phần trăm đồng nhất; phần trăm càng cao thì hai trình tự càng giống nhau.

Các phương pháp sắp xếp trình tự để so sánh là đã biết trong lĩnh vực này. Các chương trình và thuật toán sắp xếp khác nhau được mô tả trong tài liệu của: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444, 1988; Higgins and Sharp, *Gene* 73:237-244, 1988; Higgins and Sharp, *CABIOS* 5:151-153, 1989; Corpet *et al.*, *Nucleic Acids Research* 16:10881-10890, 1988; Huang, *et al.*, *Computer Applications in the Biosciences* 8:155-165, 1992; Pearson *et al.*, *Methods in Molecular Biology* 24:307-331, 1994; Tatiana *et al.*, (1999), *FEMS Microbiol. Lett.*, 174:247-250, 1999. Altschul *et al.* mô tả sự nghiên cứu chi tiết về các phương pháp sắp xếp trình tự và tính độ tương đồng (*J. Mol. Biol.* 215:403-410, 1990).

Công cụ tìm kiếm trình tự tương đồng cục bộ (BLASTTM, Altschul *et al.*, *J. Mol. Biol.* 215:403-410, 1990) của Trung tâm thông tin sinh học quốc gia (The National Center for Biotechnology Information: NCBI) có sẵn từ một số nguồn, bao gồm trung tâm thông tin sinh học quốc gia (NCBI, Bethesda, MD) và trên Internet, để sử dụng trong việc kết nối với các chương trình phân tích trình tự blastp, blastn, blastx, tblastn và tblastx. Mô tả về cách xác định độ đồng nhất trình tự sử dụng chương trình này là có sẵn trên internet ở phần trợ giúp của BLASTTM.

Để so sánh các trình tự axit amin lớn hơn khoảng 30 axit amin, chức năng "Blast 2 sequences" của chương trình BLASTTM (Blastp) được dùng bằng cách sử dụng ma trận BLOSUM62 thiết lập để mặc định các thông số (phí để mở khoảng trống [mặc định = 5]; phí để mở rộng khoảng trống [mặc định = 2]; hình phạt cho sự không khớp [mặc định = -3]; phần thưởng cho sự ghép khớp [mặc định = 1]; giá trị mong đợi (E) [mặc định = 10,0]; cỡ từ [mặc định = 3]; số lượng mô tả một dòng (V) [mặc định = 100]; số lượng sắp xếp để hiển thị (B) [mặc định = 100]). Khi sắp xếp các peptit ngắn (ít hơn khoảng 30 axit amin), sự sắp xếp cần được thực hiện bằng cách sử dụng chức năng "Blast 2 sequences", dùng ma trận PAM30 thiết lập để mặc định các thông số (điểm phạt khoảng trống mở 9, khoảng trống mở rộng 1). Protein (hoặc axit nucleic) có mức độ giống với trình tự tham khảo lớn hơn sẽ thể hiện phần trăm đồng nhất tăng khi đánh giá bằng phương pháp này, như độ đồng nhất trình tự ít nhất là 50%, ít nhất là 60%, ít nhất là 70%, ít nhất là 80%, ít nhất là 85%, ít nhất là 90%, hoặc ít nhất là 95%.

Để so sánh các trình tự axit nucleic, chức năng "Blast 2 sequences" của chương trình BLASTTM (Blastn) được dùng bằng cách sử dụng ma trận BLOSUM62 thiết lập để mặc định các thông số (phí mở khoảng trống [mặc định = 11]; phí mở rộng khoảng trống [mặc định = 1]; giá trị mong đợi (E) [mặc định = 10,0]; cỡ từ [mặc định = 11]; số lượng mô tả một dòng (V) [mặc định = 100]; số lượng sắp xếp để hiển thị (B) [mặc định = 100]). Trình tự axit nucleic có mức độ giống với trình tự tham khảo lớn hơn sẽ thể hiện phần trăm đồng nhất tăng khi đánh giá bằng phương pháp này, như độ đồng nhất trình tự ít nhất là 60%, ít nhất là 70%, ít nhất là 75%, ít nhất là 80%, ít nhất là 85%, ít nhất là 90%, ít nhất là 95%, hoặc ít nhất là 98%.

Dấu hiệu khác cho thấy hai phân tử axit nucleic có liên quan mật thiết là hai phân tử lai hoá với nhau trong các điều kiện nghiêm ngặt (xem phần “Lai hoá” ở trên).

Tuy nhiên, trình tự axit nucleic không thể hiện mức độ đồng nhất cao có thể mã hoá các trình tự axit amin giống nhau, do sự thoái hoá mã di truyền. Cần hiểu rằng có thể tạo ra sự thay đổi về trình tự axit nucleic bằng cách sử dụng sự thoái hoá này để tạo ra nhiều phân tử axit nucleic mà tất cả đều mã hoá về cơ bản cùng protein.

Chuyển nạp: Quy trình đưa phân tử axit nucleic vào tế bào, chẳng hạn bằng các kỹ thuật sinh học phân tử, dẫn đến tế bào được chuyển nạp (hoặc được biến nạp). Khi được sử dụng ở đây, thuật ngữ chuyển nạp bao gồm tất cả các kỹ thuật mà có thể đưa phân tử axit nucleic vào tế bào này, bao gồm sự tải nạp với vector virus, chuyển nạp với vector plasmit, và đưa vào ADN bằng cách điện di, chuyển nạp lipid, và tăng tốc súng bắn hạt.

Biến nạp: Tế bào biến nạp là tế bào mà phân tử axit nucleic được đưa vào đó bằng các kỹ thuật sinh học phân tử. Thuật ngữ này bao gồm tất cả các kỹ thuật để có thể đưa phân tử axit nucleic vào tế bào, bao gồm chuyển nạp với vector virus, biến nạp với vector plasmit, và đưa vào ADN trần bằng cách điện di, biến nạp lipid, và tăng tốc súng bắn hạt.

Gen nhảy: Yếu tố di truyền di động có trình tự lặp lại gần như đồng nhất ở một trong hai đầu, và chứa ít nhất một gen mã hoá transposaza (enzym cần để cài xen gen nhảy trong trình tự ADN). Các gen nhảy có thể được hợp nhất vào các vị trí khác nhau trong bộ gen của tế bào, hoặc qua mẫu AND plasmit, cosmid, hoặc fosmid phân lập in vitro. Gen nhảy còn có thể chứa các gen khác không phải gen cần cho sự cài xen.

Vector: Phân tử axit nucleic khi được đưa vào tế bào chủ, từ đó tạo ra tế bào chủ được chuyển nạp. Các vector ADN tái tổ hợp là các vector có ADN tái tổ hợp. Vector có thể bao gồm các trình tự axit nucleic cho phép vector này sao chép trong tế bào chủ, như bản sao chép gốc. Vector cũng có thể bao gồm một hoặc nhiều gen đánh dấu có thể chọn lọc và các thành phần di truyền khác đã biết trong lĩnh vực này. Vector virus là vector ADN tái tổ hợp có ít nhất một số trình tự axit nucleic có nguồn gốc từ một hoặc nhiều virus. Plasmit là vector.

Trừ khi có giải thích khác, tất cả các thuật ngữ kỹ thuật và khoa học được sử dụng ở đây có cùng nghĩa như được hiểu thông thường bởi người có hiểu biết trung bình về lĩnh vực kỹ thuật này. Các thuật ngữ số ít bao gồm cả số nhiều trừ khi có quy định khác. Tương tự, từ “hoặc” được dự định để bao gồm “và” trừ khi có quy định rõ ràng khác. Còn được hiểu là tất cả các kích thước bazơ hoặc kích thước axit amin, và tất cả giá trị trọng lượng phân tử hoặc khối lượng phân tử, đưa ra đối với các axit nucleic hoặc polypeptit là xấp xỉ, và được đề xuất trong bản mô tả. Mặc dù các phương pháp và vật liệu tương tự hoặc tương đương với các phương pháp và vật liệu được mô tả ở đây có thể được sử dụng trong thực tế hoặc thử nghiệm của sáng chế, các phương pháp và vật liệu thích hợp được mô tả dưới đây. Thuật ngữ “gồm” nghĩa là “bao gồm”. Trong trường hợp xung đột, bản mô tả sáng chế, bao gồm các giải thích về thuật ngữ, sẽ quyết định. Ngoài ra, vật liệu, phương pháp, và các ví dụ chỉ minh họa mà không làm giới hạn sáng chế.

Các phương pháp và vật liệu thích hợp để thực hiện các phương án được bộc lộ của sáng chế được mô tả dưới đây. Ngoài ra, phương án hoặc kỹ thuật thích hợp bất kỳ đã biết đối với người có hiểu biết trung bình về lĩnh vực này có thể được sử dụng trong việc thực hiện các phương án của sáng chế. Một số phương pháp và kỹ thuật thông thường có thể áp dụng cho sáng chế được mô tả, ví dụ, trong tài liệu của Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Press, 2001; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates, 1992 (and Supplements to 2000); Ausubel *et al.*, *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed., Wiley & Sons, 1999; Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1990; Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1999; and Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A.: *Practical Streptomyces genetics*, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 &UH, England, 2000.

Tất cả công bố, đơn sáng chế, bằng sáng chế, và các tài liệu tham khảo khác nêu ở đây được kết hợp để tham khảo theo toàn bộ nội dung của chúng. Trong trường hợp xung đột, bản mô tả sáng chế, bao gồm các giải thích về thuật ngữ, sẽ quyết định. Ngoài ra, vật liệu, phương pháp, và các ví dụ chỉ minh họa mà không làm giới hạn sáng chế.

III. Các vector biểu hiện tái tổ hợp được thiết kế của *Streptomyces fungicidicus*

Sáng chế bộc lộ các vector plasmit biểu hiện *Streptomyces fungicidicus* tái tổ hợp được thiết kế. Trong một số phương án, vector *Streptomyces fungicidicus* tái tổ hợp được thiết kế bao gồm ít nhất một khung đọc mở chọn lọc của *Streptomyces fungicidicus*. Trong một số phương án, vector *Streptomyces fungicidicus* tái tổ hợp được thiết kế bao gồm ít nhất một khung đọc mở chọn lọc của *Streptomyces fungicidicus* được biểu hiện trong điều kiện kiểm soát của gen khởi đầu. Trong một số ví dụ, gen khởi đầu là gen khởi đầu *Streptomyces* cơ định mạnh dẫn đến tăng cường sản xuất enduracidin khi vector được biểu hiện ở chủng *Streptomyces fungicidicus*. Trong một số phương án, khung đọc mở được liên kết theo cách có điều khiển với gen khởi đầu khác loại thay vì gen khởi đầu nguyên thể của nó. Ví dụ, có thể được liên kết theo cách có điều khiển với gen khởi đầu cơ định, như gen khởi đầu biểu hiện cơ định mạnh hoặc gen khởi đầu cảm ứng được. Trong một số ví dụ, gen khởi đầu cơ định mạnh là *ermE**p từ nguồn sản sinh erythromycin. Trong một số ví dụ, gen khởi đầu cảm ứng được là *tipA*. Trong một số ví dụ, hệ thống P(*nitA*)-NitR (Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S, Kobayashi M, Proc Natl Acad Sci U S A.2004. 101(39):14031-5) hoặc gen khởi đầu streptomycete SF14 được dùng. Trong một số ví dụ, gen khởi đầu nguyên thể của gen kháng apramycin (*amRp*) được dùng. Trong một số ví dụ, *P_{hrdB}*, *P_{tcp830}*, *P_{SF14}*, *P_{ermE}** và/hoặc *P_{neos}* được dùng.

Trong một số phương án, vector tái tổ hợp được thiết kế bao gồm khung đọc mở *orf 24* (SEQ ID NO: 38) và/hoặc khung đọc mở *orf18* (SEQ ID NO: 37) mà đã bị bất hoạt. Trong một số ví dụ, khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã, dịch khung và/hoặc đột biến điểm.

Trong một số phương án, vector tái tổ hợp được thiết kế bao gồm khung đọc mở *orf24* từ cụm gen enduracidin của *Streptomyces fungicidicus*. Trong một số ví dụ, khung đọc mở *orf24* (SEQ ID NO: 38) được liên kết theo cách có điều khiển với gen khởi đầu khác loại. Ví dụ, được liên kết với gen khởi đầu cơ định mạnh như *ermE**p. Trong các ví dụ khác, khung đọc mở *orf24* được liên kết theo cách có điều khiển với gen khởi đầu *tipA*, SF14, *amRp*, *P_{hrdB}*, *P_{tcp830}*, *P_{SF14}*, *P_{ermE}** và/hoặc *Pneos*.

Theo phương án khác, vector tái tổ hợp được thiết kế bao gồm khung đọc mở *orf18* mà tập trung vào vùng ngược dòng của cụm gen enduracidin. Khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách phá vỡ cài xen, xóa đoạn không làm lệch khung dịch mã, dịch khung và/hoặc đột biến điểm. Trong một số ví dụ, khung đọc mở *orf18* bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã, như xóa đoạn không làm lệch khung dịch mã như được minh họa trên FIG. 9B. Theo một ví dụ, khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã. Ví dụ, khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã của các axit nucleic từ 5 đến 660 của *orf18* (SEQ ID NO: 37). Nói chung, việc xóa đoạn không làm lệch khung dịch mã bên trong bất kỳ qua *orf18* dẫn đến chức năng của Orf18 bị bất hoạt do sự không đầy đủ của nó. Trong một số ví dụ, sự xóa đoạn không làm lệch khung dịch mã bao gồm việc xóa ít nhất 3 axit nucleic trong *orf18* (SEQ ID NO: 37), như ít nhất 3 axit nucleic, bao gồm 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, 132, 135, 138, 141, 144, 147, 150, 153, 156, 159, 162, 165, 168, 171, 174, 177, 180, 183, 186, 189, 192, 195, 198, 201, 204, 207, 210, 213, 216, 219, 221, 224, 227, 230, 233, 236, 239, 242, 245, 248, 251, 254, 257, 260, 263, 266, 269, 272, 275, 278, 281, 284, 287, 290, 293, 296, 299, 302, 305, 308, 311, 314, 317, 320, 323, 326, 329, 332, 335, 338, 341, 344, 347, 350, 353, 356, 359, 362, 365, 368, 371, 374, 377, 380, 383, 386, 389, 392, 395, 398, 401, 404, 407, 410, 413, 416, 419, 421, 424, 427, 430, 433, 436, 439, 442, 445, 448, 451, 454, 457, 460, 463, 466, 469, 472, 475, 478, 481, 484, 487, 490, 493, 496, 499, 502, 505, 508, 511, 514, 517, 520, 523, 526, 529, 532, 535, 538, 541, 544, 547, 550, 553, 556, 559, 562, 565, 568, 571, 574, 577, 580, 583, 586, 589,

592, 595, 598, 601, 604, 607, 610, 613, 616, 619, 621, 624, 627, 630, 633, 636, 639, 642, 645, 648, 651, hoặc 654 axit nucleic nằm trong khoảng các axit nucleic từ 5 đến 660 của *orf18* (SEQ ID NO: 37).

Trong các phương án có liên quan, vectơ plasmit tái tổ hợp được thiết kế bao gồm hai hoặc nhiều khung đọc mở từ cụm gen *enduracidin* và/hoặc các vùng hai bên sườn của cụm gen hoặc từ các chủng *actinomycete* khác. Hai hoặc nhiều khung đọc mở có thể được liên kết với gen khởi đầu đơn. Theo cách khác, chúng có thể được liên kết theo cách có điều khiển với hai gen khởi đầu khác nhau. Hai gen khởi đầu có thể là gen khởi đầu cùng loại. Theo cách khác, chúng có thể là hai loại gen khởi đầu khác nhau.

Theo các phương án khác, các khung đọc mở bổ sung hoặc thay thế mà có thể tăng cường sản xuất *enduracidin* có thể được đưa vào, hoặc làm bất hoạt, ở chủng *Streptomyces fungicidicus* được thiết kế.

Trong một số ví dụ, plasmit tái tổ hợp là pXY152-endorf24 (SEQ ID NO:3). Trong một số ví dụ, plasmit tái tổ hợp là pXY300-orf18ifd (SEQ ID NO: 8). Trong một số ví dụ, plasmit tái tổ hợp là pKS-T-orf18ifd (SEQ ID NO: 11). Trong một số ví dụ, plasmit tái tổ hợp là pKS-T-orf18pfrd-AmR (SEQ ID NO: 14). Trong một số ví dụ, plasmit tái tổ hợp là pKS-orf18ifd-T-AmR(NS)(SEQ ID NO: 19). Trong một số ví dụ, plasmit tái tổ hợp là pXY152-endorf24-camtsr (SEQ ID NO: 20). Trong một số ví dụ, plasmit tái tổ hợp là pXY152-endorf24-blatsr (SEQ ID NO: 23).

IV. Các chủng tái tổ hợp được thiết kế của *Streptomyces fungicidicus*

Sáng chế mô tả các chủng *Streptomyces fungicidicus* tái tổ hợp được thiết kế có khả năng sản xuất tăng cường *enduracidin* so với chủng đối chứng (như chủng *Streptomyces fungicidicus* kiểu dại hoặc chủng gốc công nghiệp). Trong một số phương án, chủng *Streptomyces fungicidicus* tái tổ hợp được thiết kế bao gồm ít nhất một khung đọc mở chọn lọc từ *Streptomyces fungicidicus* được đưa lên nhiễm sắc thể và được biểu hiện trong điều kiện kiểm soát của gen khởi đầu, như gen khởi đầu *Streptomyces* cơ định mạnh, mà dẫn đến sản sinh tăng cường *enduracidin* ở chủng được thiết kế. Trong một số phương án, sự biểu hiện của khung đọc mở được đưa vào trong *Streptomyces*

fungicidicus được dẫn bởi gen khởi đầu khác loại thay vì gen khởi đầu nguyên thể của nó. Ví dụ, nó có thể được liên kết theo cách có điều khiển với gen khởi đầu cơ định, như gen khởi đầu biểu hiện cơ định mạnh hoặc gen khởi đầu cảm ứng được. Trong một số ví dụ, gen khởi đầu cơ định mạnh là *ermE**p từ nguồn sản sinh erythromycin. Trong một số ví dụ, gen khởi đầu cảm ứng được là *tipA*. Trong một số ví dụ, hệ thống P(*nitA*)-NitR (xem Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S, Kobayashi M, Proc Natl Acad Sci U S A., 2004. 101(39):14031-5) hoặc gen khởi đầu streptomycete SF14 được dùng. Trong một số ví dụ, gen khởi đầu biểu hiện cơ định là *amRp*. Trong một số ví dụ, các gen khởi đầu *P_{hrdB}*, *P_{tcp830}*, *P_{SF14}*, *P_{ermE}** và/hoặc *P_{neos}* được dùng.

Trong một số phương án, chủng thiết kế bao gồm khung đọc mở *orf24* từ cụm gen enduracidin của *Streptomyces fungicidicus*. Trong một số ví dụ, khung đọc mở *orf24* được liên kết theo cách có điều khiển với gen khởi đầu khác loại. Ví dụ, nó được liên kết với gen khởi đầu cơ định mạnh như *ermE**p. Trong các ví dụ khác, khung đọc mở *orf24* được liên kết theo cách có điều khiển với gen khởi đầu *tipA*, SF14, *amRp*, *P_{hrdB}*, *P_{tcp830}*, *P_{SF14}*, *P_{ermE}** và/hoặc *P_{neos}*.

Theo phương án khác, chủng thiết kế liên quan đến khung đọc mở *orf18* tập trung ở vùng ngược dòng của cụm gen enduracidin. Khung đọc mở *orf18* bị bất hoạt bằng cách phá vỡ cài xen, xóa đoạn không làm lệch khung dịch mã, dịch khung và/hoặc đột biến điểm. Trong một số ví dụ, khung đọc mở *orf18* bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã, như xóa đoạn không làm lệch khung dịch mã như được minh họa trên FIG. 9B. Theo một ví dụ, khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã. Ví dụ, khung đọc mở *orf18* (SEQ ID NO: 37) bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã của các axit nucleic từ 5 đến 660 của (SEQ ID NO: 37). Nói chung, việc xóa đoạn không làm lệch khung dịch mã bên trong bất kỳ qua *orf18* dẫn đến chức năng của Orf18 bị bất hoạt do sự không đầy đủ của nó.

Theo các phương án có liên quan, chủng thiết kế bao gồm hai hoặc nhiều khung đọc mở từ cụm gen enduracidin và/hoặc các vùng hai bên sườn của cụm gen hoặc từ các

chủng actinomycete khác. Hai hoặc nhiều khung đọc mở có thể được liên kết với gen khởi đầu đơn. Theo cách khác, chúng có thể được liên kết theo cách có điều khiển với hai gen khởi đầu khác nhau. Hai gen khởi đầu có thể là gen khởi đầu cùng loại. Theo cách khác, chúng có thể là hai loại gen khởi đầu khác nhau.

Theo các phương án khác, các khung đọc mở bổ sung hoặc thay thế mà có thể tăng cường sản xuất enduracidin có thể được đưa vào, hoặc làm bất hoạt, ở chủng thiết kế của *Streptomyces fungicidicus*.

Trong một số phương án, chủng thiết kế của *Streptomyces fungicidicus* có nguồn gốc từ chủng gốc kiểu dại, như, nhưng không bị giới hạn ở, *Streptomyces fungicidicus* công ty nuôi cấy mô Mỹ (American Tissue Culture Company: ATCC) 21013. Theo các phương án khác, chủng thiết kế của *Streptomyces fungicidicus* có nguồn gốc từ chủng gốc công nghiệp, như, nhưng không bị giới hạn ở BM38-2 (ATCC PTA-122342). Theo các phương án khác, chủng thiết kế của *Streptomyces fungicidicus* có nguồn gốc từ các chủng đột biến thông thường, như, nhưng không bị giới hạn ở *Streptomyces fungicidicus* ATCC 31729, *Streptomyces fungicidicus* ATCC 31730 và *Streptomyces fungicidicus* ATCC 31731.

Trong một số phương án, việc sản xuất tăng cường enduracidin là tăng gấp ít nhất 1,2 lần, như ít nhất 1,5 lần, ít nhất 2 lần, ít nhất 2,5 lần, ít nhất 3 lần, ít nhất 3,5 lần, ít nhất 4 lần, ít nhất 4,5 lần, bao gồm, nhưng không bị giới hạn ở tăng gấp 1,2 đến 10 lần, tăng gấp 1,2 đến 4,6 lần, tăng gấp 2 đến 5 lần, như tăng gấp 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 2,1, 2,2, 2,3, 2,4, 2,5, 2,6, 2,7, 2,8, 2,9, 3,0, 3,1, 3,2, 3,3, 3,4, 3,5, 3,6, 3,7, 3,8, 3,9, 4,0, 4,1, 4,2, 4,3, 4,4, 4,5, 4,6, 4,7, 4,8, 4,9, 5,0, 5,5, 6,0, 6,5, 7,0, 7,5, 8,0, 8,5, 9,0, 9,5 và 10 lần khi sản xuất enduracidin so với chủng *Streptomyces fungicidicus* đối chứng. Trong một số phương án, chủng *Streptomyces fungicidicus* đối chứng là chủng *Streptomyces fungicidicus* kiểu dại, bao gồm, nhưng không bị giới hạn ở, *Streptomyces fungicidicus* công ty nuôi cấy mô Mỹ (ATCC) 21013 hoặc chủng gốc công nghiệp, như, nhưng không bị giới hạn ở, BM38-2 (ATCC PTA-122342), hoặc chủng đột biến thông thường, như, nhưng không bị giới hạn ở *Streptomyces fungicidicus* ATCC 31729,

Streptomyces fungicidicus ATCC 31730 và *Streptomyces fungicidicus* ATCC 31731. Theo một ví dụ, đối chứng là *Streptomyces fungicidicus* ATCC 21013 và tăng sự sản xuất enduracidin tăng cường gấp ít nhất 1,2 lần, như gấp tăng 1,2 đến 4,6 lần. Theo một ví dụ, đối chứng là *Streptomyces fungicidicus* BM38-2 (ATCC PTA-122342) và tăng sự sản xuất enduracidin tăng cường gấp ít nhất 1,2 lần, như tăng gấp 1,2 đến 4,6 lần.

V. Cấu trúc của các chủng tái tổ hợp được thiết kế của *Streptomyces fungicidicus*

Theo các phương án, các chủng tái tổ hợp của *Streptomyces fungicidicus* có thể được tạo cấu trúc bằng cách hợp nhất plasmit tái tổ hợp chứa ít nhất một sự sản xuất enduracidin tăng cường khung đọc mở vào nhiễm sắc thể của chủng gốc *Streptomyces fungicidicus*. Vector liên hợp hợp nhất có thể có, hoặc có thể được thiết kế để có, gen khởi đầu *Streptomyces* cơ định mạnh. Trong một số phương án, plasmit có thể thiếu replicon của streptomycete và có thể được hợp nhất vào nhiễm sắc thể bằng cách tái tổ hợp tương đồng giao nhau ở vị trí đặc hiệu đơn. Theo các phương án khác, plasmit có thể có mặt ở dạng plasmit tự do. Trong một số phương án, vector liên hợp có thể được thiết kế trong đó sự cài xen plasmit mang một phần hoặc toàn bộ gen đã xóa đoạn quan tâm, và các vùng hai bên của nó, mà có thể được hợp nhất vào nhiễm sắc thể sau khi tái tổ hợp tương đồng giao nhau kép để tạo ra đột biến xóa đoạn không làm lệch khung dịch mã.

VI. Sản xuất enduracidin từ các chủng tái tổ hợp được thiết kế của *Streptomyces fungicidicus*

Các chủng tái tổ hợp được thiết kế của *Streptomyces fungicidicus* được tạo ra theo sáng chế đề xuất phương pháp sản xuất các mức enduracidin tăng cường. Sự tiến bộ về mặt kỹ thuật trong lĩnh vực này cho phép tiết kiệm chi phí đáng kể kết hợp với việc sản xuất enduracidin. Trong một số ví dụ, các phương pháp sản xuất enduracidin bao gồm bước nuôi cấy chủng *Streptomyces fungicidicus* tái tổ hợp trong các điều kiện đủ để sản xuất enduracidin. Trong một số ví dụ, phương pháp này còn bao gồm bước phân lập enduracidin từ môi trường nuôi cấy sau khi nuôi cấy. Trong một số ví dụ, phương pháp này còn bao gồm bước xác định hoạt tính kháng khuẩn của enduracidin được sản xuất,

như bằng cách phân tích HPLC hoặc thử nghiệm sinh học sử dụng *S. aureus* ATCC 29213 hoặc *Bacillus subtilis* ATCC 6633 làm vi sinh vật chỉ thị.

Trong một số ví dụ, enduracidin được sản xuất bằng chủng *Streptomyces fungicidicus* đã bộc lộ bằng cách dùng các điều kiện lên men như được mô tả trước đó để sản xuất enduracidin (Higashide *et al.* *J. Antibiot.* 21: 126-137, 1968). Sau khi sản xuất, các hợp chất có thể được tinh chế và/hoặc được phân tích bao gồm phương pháp phân tích HPLC như được mô tả trong Ví dụ 1. Các phương pháp sản xuất enduracidin và thu hoạch hợp chất này từ môi trường sinh trưởng có thể được tìm thấy ở Patent Mỹ số 4,465,771, toàn bộ nội dung của tài liệu này được kết hợp ở đây để tham khảo.

Trong một số ví dụ, chủng *Streptomyces fungicidicus* được nuôi cấy trong môi trường canh tương tripsin (TSB) trên máy khuấy (như ở 225 vòng/phút (rounds per minute: rpm) và 30°C trong 48 giờ) và sau đó chuyển đến môi trường sản xuất enduracidin (EPM, bảng 1 dưới đây) trong khoảng thời gian lên men liên tục, như trong ít nhất năm ngày và đến mười một ngày, bao gồm 5, 6, 7, 8, 9, 10 hoặc 11 ngày lên men liên tục. Trong một số ví dụ, sự sản xuất enduracidin bằng các chủng kiểu dại và dẫn xuất được thực hiện trong các thiết bị lên men tự động.

Bảng 1. Chế phẩm môi trường sản xuất enduracidin (Enduracidin Production Medium: EPM) (độ pH 6,7)

Thành phần	Nồng độ (%)
Tinh bột tan	1,5
Glucoza	1,0
Bột ngô	2,5
Bột gluten ngô	2,0
Dịch ngâm ngô	0,25
Natri clorua	0,25
NaH ₂ PO ₄	1,3
KH ₂ PO ₄	0,05
(NH ₄) ₂ SO ₄	0,15

CaCO ₃	0,5
Lactoza	0,5
ZnCl ₂	0,005
Dầu gà	0,7

Trong một số ví dụ, sinh khối *Streptomyces fungicidicus* được tạo ra bằng quy trình lên men trong thiết bị lên men công nghiệp được thiết kế dạng bể sâu dễ vệ sinh với các hệ thống để giám sát và kiểm soát độ pH, nhiệt độ, oxy, sự thông khí, khuấy trộn. Ví dụ, mỗi mẻ *S. fungicidicus* lên men được bắt đầu từ kho hạt giống chế biến được định rõ đặc điểm và được kiểm soát của hạt giống sản xuất được bảo quản ở nơi an toàn và giữ trong môi trường nhiệt độ thấp.

Trong một số ví dụ, quy trình lên men diễn ra theo một hoặc nhiều giai đoạn, như ba giai đoạn sau và có thể tùy ý theo sau là quy trình xử lý xuôi dòng tiếp theo:

Giai đoạn I:

Môi trường nuôi cấy hạt giống chế biến được thiết lập đặc điểm được sử dụng để bắt đầu mẻ lên men. Một đến năm lọ trong số các lọ hạt giống đông lạnh được lấy từ kho chứa nhiệt độ thấp và làm tan giá tự nhiên hoặc đặt trong bể nước ở 28°C- 32°C cho đến khi tan hết các thành phần. (Các) môi trường nuôi cấy tan giá được chuyển vô trùng vào nước vô trùng giữ ở nhiệt độ trong phòng và khuấy nhẹ để tái huyền phù môi trường nuôi cấy.

Giai đoạn II:

Môi trường nuôi cấy tái huyền phù được chuyển vô trùng vào môi trường hạt giống 0,005 m³-0,05 m³. Môi trường hạt giống bao gồm glucoza (0,1-1,0 g/L), Dextrin (0,1-3 g/L), dịch ngâm ngô (0-5.0 mL/L), bột đậu tương (1-5,0 g/L), amoni sulfat (0,1-0-0,5 g/L), mono-kali phosphat (0,13-0,54 g/L), sắt (II) sulfat (0,00-0,5 g/L), kali hydroxit (0,13 mL/L), canxi carbonat (1-2 g/L), chất khử tạo bọt gốc silicon (0,1 mL/L), nước thêm vừa đủ. Môi trường được vô trùng ở 125°C-128°C trong 30-45 phút và sau

đó làm mát đến 28°C-32°C. Thể tích môi trường được điều chỉnh bằng cách sử dụng nước vô trùng đến thể tích chế biến mong muốn. Độ pH được điều chỉnh đến 6,5-7,0.

Các thông số hoạt động của chu trình gieo hạt giống ở quy mô lớn bao gồm: Nhiệt độ nuôi cấy bằng 28°C±2°C, áp suất bên trong bằng 1,0±0,5 kg/cm², tốc độ thông khí bằng 3±2 Nm³/phút, và tốc độ khuấy trộn xấp xỉ 80 vòng/phút, phụ thuộc vào kích thước cấu hình của bình. Độ pH, mức tiêu thụ oxy và độ nhớt được giám sát nhưng không kiểm soát. Môi trường nuôi cấy được sinh trưởng trong 40-80 giờ trước khi chuyển vào thiết bị lên men sản xuất chính. Độ nhớt ở thời điểm chuyển cần nằm trong khoảng từ 200-600 centipoise (centipoise - cps), và độ pH cần ≤6,0, và cần tăng mức độ tiêu thụ oxy. Môi trường nuôi cấy hạt giống được chuyển vô trùng vào môi trường lên men chính để hoàn thành chu trình lên men.

Giai đoạn III:

Chế phẩm môi trường lên men sản xuất (10 m³-250 m³) bao gồm các thành phần tự nhiên và hoá học như bột ngô (13,0-15,0% trọng lượng/thể tích), bột gluten ngô (3,0-6,0% trọng lượng/thể tích), bột hạt bông (0,1-0,3% trọng lượng/thể tích), dịch ngâm ngô (0,1-0,6% thể tích/thể tích), natri clorua (0,3% trọng lượng/thể tích), amoni sulfat (0,25-0,6% trọng lượng/thể tích), axit lactic (0-0,5% thể tích/thể tích), kẽm clorua (0,01% trọng lượng/thể tích), sắt (II) sulfat (0,0-0,02% trọng lượng/thể tích), kali hydroxit (0,20-0,5% thể tích/thể tích), canxi sulfat (0,0-0,5% trọng lượng/thể tích), canxi carbonat (0,5% trọng lượng/thể tích), amylaza (0,02-0,06% trọng lượng/thể tích), kali hydroxit (0,05% thể tích/thể tích), dầu thực vật (0,5-2,0% thể tích/thể tích), chất khử tạo bọt, và nước đến lượng vừa đủ. Các thành phần được thêm vào theo thứ tự được liệt kê. Thêm nước vào các thành phần này sau đó gia nhiệt đến 70-90°C để cho phép enzym phá vỡ phức carbohydrat trong 15 phút ở nhiệt độ này. Thêm các thành phần còn lại, điều chỉnh độ pH đến 6,6-6,8, và thêm nước đến lượng vừa đủ, vô trùng ở 125°C-128°C trong 25-50 phút để vô trùng môi trường. Làm mát môi trường đến 25°C-32°C, và thêm nước vừa đủ đến thể tích chế biến.

Chuyển các thành phần từ thiết bị lên men hạt giống vào môi trường lên men chính và cài đặt thiết bị lên men theo các điều kiện sau: nhiệt độ $28^{\circ}\text{C} \pm 3^{\circ}\text{C}$, tốc độ thông khí 20-60 Nm³/phút, áp suất bên trong 0,1-1,0 kg/cm², tốc độ khuấy trộn tương đương với khoảng 1,85 kW/m³. Tốc độ thông khí, áp suất bên trong và tốc độ khuấy trộn được điều chỉnh đến tốc độ cần để đảm bảo oxy hoà tan không phải là tốc độ giới hạn xác định. Kiểm soát tạo bọt cẩn thận trong suốt chu trình để tránh nhiễm bẩn hoặc tránh thất thoát. Bắt đầu kiểm soát độ pH sau khi nhu cầu oxy tăng lên. Các thông số sau được kiểm soát và/hoặc được giám sát trong suốt chu trình lên men: độ pH, sự thông khí, oxy hoà tan, CO₂, độ nhớt, độ tinh khiết, tốc độ khuấy trộn, áp suất bên trong, và đường dư. Duy trì độ pH ở 6,8 cho đến khi vi khuẩn dừng phát triển, sau đó cho phép độ pH thay đổi tự nhiên cho đến khi thu hoạch. Chu trình lên men điển hình là 210-300 giờ. Môi trường nuôi cấy sẵn sàng để thu hoạch khi hiệu lực lớn hơn 5000 µl/L, độ pH tăng lên đến 7,5 hoặc cao hơn, độ nhớt giảm, và nhu cầu oxy dừng.

Thu hoạch quá trình lên men bằng cách gia nhiệt môi trường nuôi cấy đến 70°C trong 30 phút để bất hoạt vi khuẩn, và sau đó làm nguội dịch thu hoạch đến 25°C - 32°C .

Trong một số ví dụ, việc xử lý xuôi dòng bao gồm việc loại nước ra khỏi sinh khối, làm khô sinh khối và tạo công thức sinh khối khô thành dạng trộn trước.

Nộp lưu vật liệu sinh học

Các vật liệu sinh học sau được nộp lưu theo Hiệp ước Budapest với Trung tâm lưu giữ giống nuôi cấy Hoa Kỳ (American Type Culture Collection), và có các số hiệu lưu giữ sau:

<u>Vật nộp lưu</u>	<u>Số hiệu đăng ký</u>	<u>Ngày nộp lưu</u>
BM38-2-18pfrd	PTA-124007	02-03-2017
BM38-2-24/16	PTA-124006	02-03-2017

Các chủng ở trên đã được nộp lưu trong các điều kiện nhằm đảm bảo rằng có thể tiếp cận được quy trình nuôi cấy trong quá trình chưa xử lý đơn patent này cho người được quyết định bởi Ủy viên hội đồng Sáng chế và Nhãn hiệu (Commissioner of Patents and Trademarks) được trao quyền theo điều 37 C.F.R. §1.14 và 35 U.S.C. §122. Các vật

nộp lưu thể hiện quy trình nuôi cấy gần như tinh khiết so với các chủng được nộp lưu. Các vật nộp lưu có giá trị theo yêu cầu của luật patent ngược ngoài ở các nước trong đó các đơn đồng dạng, hoặc đơn tách của nó được nộp. Tuy nhiên, cần hiểu rằng tính khả dụng của vật nộp lưu không tạo ra li-xăng để thực hiện sáng chế trong việc xâm phạm quyền patent do chính phủ cấp.

Các ví dụ không giới hạn sau được đưa ra nhằm minh họa các dấu hiệu và/hoặc phương án cụ thể xác định. Các ví dụ này không được coi là làm giới hạn sáng chế ở các dấu hiệu và phương án cụ thể được mô tả này.

Ví dụ thực hiện sáng chế

Ví dụ 1

Vật liệu và phương pháp để sản xuất Enduracidin tăng cường

Ví dụ này đề xuất phương pháp điển hình để sản xuất enduracidin tăng cường.

Chủng vi khuẩn, plasmit, fosmit và điều kiện nuôi cấy.

Streptomyces fungicidicus B-5477 (ATCC 21013) và *Escherichia coli* S17-1 (ATCC 47055) được mua từ ATCC. Chủng *S. fungicidicus* BM38-2 (ATCC PTA-122342) và các mẫu chuẩn enduracidin A và B được cung cấp bởi Intervet/Merck Animal Health (MAH). Các chủng *E. coli* DH5 α (Life Technologies, Inc.), EPI300 (Epicentre) và XL10-Gold (Stratagene) được sử dụng làm vật chủ cho plasmit, fosmit *E. coli* và vectơ con thoi *E. coli*-*Streptomyces*. Các plasmit pSET152 (Bierman *et al.*, Gene 116: 43-49, 1992, toàn bộ nội dung của tài liệu này được kết hợp ở đây để tham khảo) và pIJ773 được cung cấp bởi Giáo sư Keith Chater (JIC, Norwich, UK). Plasmit pWHM860 mang gen khởi đầu *ermE**p được cung cấp bởi Giáo sư Bradley Moore (UCSD, San Diego). ISP2 (Môi trường ISP Difco™ 2 - Difco™ ISP Medium 2), ISP4 và TSB (Môi trường canh đậu tương trypsin Bacto™-Bacto™ Tryptic Soy Broth) được mua từ VWR. Các đoạn mồi được sử dụng cho PCR và xác định trình tự ADN được tổng hợp từ Fisher and Sigma-Aldrich. Môi trường và điều kiện nuôi cấy để sinh trưởng

S. fungicidicus được mô tả bởi Higashide *et al.* (*Journal of Antibiotics*, 21:126-137, 1968). Tất cả các quy trình *E. coli* đều được thực hiện theo protocol chuẩn.

Phân lập và thao tác ADN

Để điều chế ADN bộ gen từ *S. fungicidicus* B-5477, BM38-2 (ATCC PTA-122342) và các chủng dẫn xuất tái tổ hợp và đột biến để xác định trình tự, cấu trúc thư viện fosmit, tách dòng phụ và PCR, các bào tử vừa thu hoạch được từ các chủng riêng rẽ được ủ và sinh trưởng trong 100 mL môi trường lỏng TSB có bổ sung 5 mM MgCl₂ và 0,5% glyxin. Quy trình nuôi cấy điển hình được thực hiện trong bình Erlenmeyer 500 mL trên thiết bị ủ lắc tròn ở 225 rpm và 30°C trong 48 đến 72 giờ. Tế bào sợi nấm được thu hoạch bằng cách ly tâm ở 4000 rpm và 4°C trong 15 phút. Dịch nổi được loại bỏ và viên cặn lần lượt được rửa một lần bằng 10,3% sucroza và hai lần bằng 10 mM Tris-HCl và 1 mM dinatri EtylenDiaminTetra-Axetat (EDTA), pH 8,0 (dung dịch đệm TE). Tế bào ướt, tương đương với thể tích 80 µL nước được phân bổ vào ống vi ly tâm vô trùng 1,5 mL. Sau khi bổ sung 300 µL dung dịch phân giải chứa 200 µL Tris-HCl 10 mM và EDTA 1 mM, pH 8,0 và sucroza 0,3 M (dung dịch đệm TES), 50 µL EDTA 0,5 M, 50 µL lysozym (50 mg/mL), ống được ủ ở 37°C trong khoảng từ 30 đến 60 phút cho đến khi dung dịch trở nên nhớt. Tiếp theo, 5 µL proteinaza K (20 mg/mL) và 180 µL natri dodecyl sulfat (SDS) 10% được bổ sung vào mỗi ống. Sau khi trộn nhẹ mà kỹ lưỡng, dung dịch được ủ ở 37°C trong 90 phút. Tiếp theo, 80 µL Xetyl Trimetyl Amoni Bromua (CTAB) 10% được bổ sung. Sau khi trộn kỹ, ống được ủ ở 65°C trong 10 phút. Dung dịch được chiết hai lần bằng 600 µL phenol/clorofom/rượu isoamylic (25/24/1). ADN hệ gen trong các pha nước phía trên được thu hồi và được kết tủa bằng 0,6 thể tích isopropanol. ADN hệ gen đã thu hoạch được rửa hai lần bằng 70% etanol. Sau khi sấy khô ở nhiệt độ trong phòng trong 10 phút, ADN hệ gen được hòa tan trong 50 đến 100 µL nước vô trùng. Chất lượng cao của chế phẩm ADN hệ gen được xác nhận bằng cách cắt bằng *Hind*III và *Sau*3AI cho thấy cắt hoàn toàn và không có sự phân hủy ADN hệ gen không cắt nhờ điện di trên gel agarosa 0,8%. ADN hệ gen hỗn hợp được cắt tiếp bằng RNaza để loại bỏ ARN lẫn vào. Độ tinh khiết và lượng ADN hệ gen được xác định bằng quang phổ kế Nanodrop. Các thao tác ADN streptomycete chung bao gồm

điện di trên gel agarosa được thực hiện và kit QIAprep Spin Miniprep (Qiagen) được sử dụng để điều chế plasmit và fosmit từ các chủng *E. coli*. Các endonucleaza giới hạn, ADN ligaza, ADN polymeraza, transposaza, enzym Klenow, phosphatasa kiềm và ligaza được mua từ Biolabs, Invitrogen, Epicentre and Roche, và được sử dụng theo khuyến cáo của nhà sản xuất. Các mảnh ADN được tinh chế bằng cách sử dụng kit QIAquick Gel Extraction.

PCR

PCR khuẩn lạc được thực hiện như sau: bào tử từ các khuẩn lạc tuyển chọn đột biến độc lập được ủ trong môi trường nuôi cấy lỏng TSB. Sau khi sinh trưởng trong 48 đến 72 giờ, hệ sợi nấm được thu hoạch bằng cách ly tâm và được rửa hai lần bằng dung dịch đệm TE (10 mM Tris, 1 mM EDTA), pH 8,0. Hệ sợi nấm được tạo huyền phù lại trong H₂O vô trùng và được sử dụng làm khuôn trong hỗn hợp phản ứng PCR với thể tích cuối 100 μ L chứa 60 μ L hệ sợi nấm, 150 pmol mỗi môi, 20 μ L dung dịch đệm A giàu GC 5X AccuPrime (Invitrogen), và 1 μ L Polymix (được bổ sung ở 80°C) từ hệ thống Expand long template PCR (Roche). PCR được thực hiện như sau: 1 chu kỳ ở 95°C trong 3 phút, 30 chu kỳ ở 95°C trong 1 phút, ở 55°C trong 1 phút, và ở 72°C trong 2 phút. Phản ứng được kết thúc bằng một chu kỳ kéo dài ở 72°C trong 10 phút. Sản phẩm PCR được tinh chế trên gel và được xác định trình tự. PCR tổng quát được thực hiện tương tự như đã mô tả ở trên, chỉ khác ở chỗ ADN hệ gen, ADN plasmit/fosmit phân lập được được sử dụng làm khuôn thay vì sử dụng trực tiếp ADN giải phóng từ các khuẩn lạc hệ sợi nấm mà không tinh chế trước.

Cấu trúc plasmit biểu hiện hợp nhất pXY152-endorf24

Để biểu hiện lệch vị gen điều hòa giả định *orf24* từ cụm gen enduracidin ở các chủng *S. fungicidicus* kiểu dại và BM38-2 (ATCC PTA-122342), *orf24* được tách dòng trong plasmit hợp nhất pXY152 có nguồn gốc từ pXY152aR20 (Yin *et al.*, *J. Natural Products*, 73: 583-589, 2010, toàn bộ nội dung của tài liệu này được kết hợp ở đây để tham khảo) *orf24* được khuếch đại bằng PCR từ ADN hệ gen *S. fungicidicus* bằng cách sử dụng đoạn mồi xuôi (End24Ndpf: 5'-

CCACCACATATGGAAATAAGTTCGCTCTCCA-3' (SEQ ID NO:1, vị trí *NdeI* được in đậm) và đoạn môi ngược (End24ERpr:5'-GTGTGTGAATTCCTCGTTCACCCGGCCAGATG-3' (SEQ ID NO: 2, vị trí *EcoRI* được in đậm). Sản phẩm PCR được cắt bằng *NdeI* và *EcoRI*. Tiếp theo, mảnh *orf24* đã tinh chế trên gel được gắn với vector giới hạn tương tự pXY152. Plasmid thu được được ký hiệu là pXY152-endorf24 (FIG. 2; SEQ ID NO: 3). Mảnh cài xen *orf24* được xác nhận là không có lỗi bằng cách xác định trình tự.

Cấu trúc plasmid pXY300-*orf18ifd* để xóa đoạn không làm lệch khung dịch mã *orf18*

pXY300-*orf18ifd* được tạo cấu trúc bằng cách tách dòng hai mảnh nằm hai bên *orf18* và dùng để xóa đoạn thành pXY300, một vector con thoi nhạy nhiệt kiểu tiếp hợp *E. coli-Streptomyces* chứa gen kháng thiostrepton (*tsr*) để chọn lọc trong *S. fungicidicus*. Trình tự nằm hai bên “ngược” 2 kb và “xuôi” 2 kb, lần lượt được ký hiệu là *orf18ifdNP* và *orf18ifdPH*, nằm hai bên *orf18* được tạo ra bằng PCR bằng cách sử dụng ADN hệ gen *S. fungicidicus* làm khuôn và hai bộ đoạn môi. Mảnh *orf18ifdPH* được khuếch đại bằng cách sử dụng các đoạn môi xuôi và ngược (*Ifdenorf18pf1*, 5'-TTATTGAAGCTTGCCGGGGCCGACGCGGCGGGCGGCCT-3'

(SEQ ID NO: 4), *Ifdendorf18pr1*, 5'-GTTGTTTTTAATTAAACACCAGGCCTCCTGGGGTG-3'

(SEQ ID NO: 5), các vị trí *HindIII* và *PacI* được in đậm). Mảnh *orf18ifdNP* được khuếch đại bằng cách sử dụng các môi xuôi và ngược (*Ifdendorf18pf2*, 5'

-TTTATATTAATTAAATGACCCTTCCGTCCCGCCCCCGAT-3'

(SEQ ID NO: 6), *Ifdendorf18pr2*, 5'

-TTTGGTGCTAGCTGGTCGTGGCGCTGTTCC-3'

(SEQ ID NO: 7), các vị trí *PacI* và *NheI* được in đậm). Hai mảnh PCR này được giới hạn thích hợp và được gắn đồng thời với vector pXY300 điều chế được bằng cách cắt bằng *NheI* và *HindIII*, thu được plasmid pXY300-*orf18ifd* (FIG. 3; SEQ ID NO: 8). Đoạn

cài xen xóa đoạn không làm lệch khung dịch mã không lỗi pXY300-orf18ifd được xác nhận bằng cách xác định trình tự.

Cấu trúc plasmit pKS-T-orf18pfrd-AmR để xóa đoạn *orf18* và các vùng hai bên của nó.

Mảnh *oriT* được khuếch đại bằng PCR từ plasmit pIJ773 bằng cách sử dụng mỗi xuôi (Oritnhexbahd3f, 5'

-AGCACAGCTAGCTTCTAG**AAGCTTC**ATTCAAAGGCCGGCA-3' (SEQ ID NO: 9), vị trí *HindIII* được in đậm) và mỗi ngược (Oriter1pstxhor, 5'

-GCCAGTGAATTCTGCAG**CTCGAG**CAGAGCAGGATTCCCGTTGA-3' (SEQ ID NO: 10), vị trí *XhoI* được in đậm). Mảnh *oriT* được cắt bằng *HindIII* và *XhoI*, được tinh chế trên gel và tiếp theo được gắn vào vector giới hạn tương tự dẫn xuất pBluescript II KS thu được plasmit pKS-T (Alting-Mees and Short, Nucleic acids Research, 17: 9494, 1989). Mảnh cài xen của plasmit pXY300-orf18ifd được cắt ra bằng cách cắt bằng *NheI* và *HindIII*, được tinh chế trên gel và tiếp theo được gắn với plasmit pKS-T đã chuyển thành mạch thẳng nhờ *NheI* và *HindIII* thu được plasmit pKS-T-orf18ifd (FIG. 4; SEQ ID NO: 11). Mảnh 1 kb mang *aac(3)IV*, gen kháng apramycin (*amR*), được khuếch đại từ pIJ773 bằng cách sử dụng đoạn mỗi xuôi (ApraNcoIpf, 5'-GAATGG**CCATGG**TTCATGTGCAGCTCCAT-3' (SEQ ID NO: 12), vị trí *NcoI* được in đậm) và đoạn mỗi ngược (ApraBamHIpr, 5'-TCTCG**AGGATCC**GAATAGGAAGTTCGGAAT-3' (SEQ ID NO: 13), vị trí *BamHI* được in đậm). Việc cắt mảnh AmR và plasmit pKS-T-orf18ifd bằng *NcoI* và *BamHI* tạo ra cả mảnh cài xen và vector dùng để gắn. Plasmit thu được được ký hiệu là pKS-T-orf18pfrd-AmR (FIG. 5; SEQ ID NO: 14).

Cấu trúc plasmit pKS-T-orf18ifd-AmR(NS) để xóa đoạn không làm lệch khung dịch mã *orf18*.

Mảnh cài xen pXY300-orf18ifd được cắt ra bằng cách cắt bằng *NheI* và *HindIII*, được tinh chế trên gel và tiếp theo được gắn với vector pBluescript II KS đã chuyển

thành mạch thẳng nhờ *SpeI* và *HindIII* tạo ra plasmit pKS-orf18ifd. Mảnh *oriT* được khuếch đại bằng PCR bằng cách sử dụng đoạn xuôi (*Oritnhexahd3f*, 5'-AGCACAGCTAGCTTCTAGAA**AGCTTC**ATTCAAAGGCCGGCA-3' (SEQ ID NO: 15), vị trí *HindIII* được in đậm) và đoạn ngược (*oriTXhNdSpr*, 5'-AGGCAG**CTCGAGCATATGACTAGT**CAGAGCAGGATTCCCGTTGA-3' (SEQ ID NO: 16), các vị trí *XhoI*, *NdeI* và *SpeI* được in đậm). Mảnh *oriT* được cắt bằng *XhoI* and *HindIII*, được tinh chế trên gel và tiếp theo được gắn với plasmit giới hạn tương tự pKS-orf18ifd thu được plasmit pKS-orf18ifd-T. Mảnh 1 kb mang gen *aac(3)IV* tạo tính kháng apramycin (AmR) được khuếch đại từ pIJ773 bằng PCR bằng cách sử dụng đoạn xuôi (*ApraNdepf*, 5'-GAATGG**CATATGG**TTCATGTGCAGCTCCAT-3' (SEQ ID NO: 17), vị trí *NdeI* được in đậm) và đoạn ngược (*ApraSpeIpr*, 5'-TCTAGAA**CTAGTGA**ATAGGAACTTCGGAAT-3' (SEQ ID NO: 18), vị trí *SpeI* được in đậm). Plasmit pKS-orf18ifd-T được chuyển thành mạch thẳng bằng cách cắt bằng *NdeI* và *SpeI* và tiếp theo được gắn với mảnh giới hạn tương tự AmR tạo ra plasmit pKS-orf18ifd-T-AmR(NS) (FIG. 6; SEQ ID NO: 19).

Các quy trình tiếp hợp liên giống, làm đứt quãng gen trên cơ sở pXY300 và trên cơ sở pKS.

Các plasmit đứt quãng gen được đưa riêng rẽ vào *E. coli* S17-1 bằng cách biến nạp và tiếp theo được chuyển sang *S. fungicidicus* hoặc dẫn xuất của nó nhờ tiếp hợp. Nói vắn tắt, các bào tử *S. fungicidicus* vừa thu hoạch được cho nảy mầm trước và tế bào *E. coli* S17-1 được sinh trưởng qua đêm ở 37°C trong môi trường canh Terrific. Các dung dịch pha loãng theo bậc của huyền phù bào tử nảy mầm được tạo ra và 100 mL mỗi dung dịch pha loãng này được trộn với thể tích bằng nhau của *E. coli* S17-1 mang plasmit đứt quãng trên cơ sở pXY300. Các dung dịch được cấy lên đĩa thạch ISP4 có bổ sung 10 mM MgCl₂ và được ủ trong 22 giờ ở 30 hoặc 37°C. Mỗi đĩa được phủ 3 mL thạch dinh dưỡng mềm chứa natri nalidixat và apramycin (0,5 mg/mL) và được ủ tiếp ở 30°C trong khoảng 1 tuần. Các thể tiếp hợp ngoài phân lập được mà sống sót qua việc chọn lọc bằng chất kháng sinh được tinh chế bằng cách cấy vạch trên đĩa thạch ISP4 có bổ sung natri nalidixat và apramycin (50 µg/mL each).

Để thực hiện nghiên cứu làm đứt quãng gen bằng plasmit trên cơ sở pXY300, các thể tiếp hợp ngoài trước hết được nuôi cấy trong môi trường lỏng TSB chứa apramycin (5 µg/mL) ở 30°C trong 24 giờ mà tại thời điểm đó hệ sợi nấm được thu hoạch, được làm đồng nhất và được sử dụng để ủ môi trường lỏng TSB có bổ sung apramycin (5 µg/mL). Sau 3–6 ngày ủ ở 40°C, hệ sợi nấm được làm đồng nhất và được cấy lên đĩa thạch ISP4 chứa apramycin (50 µg/mL) và được ủ ở 30°C trong 1 tuần. ADN hệ gen được phân lập từ các khuẩn lạc sống sót riêng rẽ được chọn lọc ngẫu nhiên và được phân tích bằng PCR hoặc thẩm tách Southern để xác nhận rằng sự đứt quãng chéo đơn hoặc kép đã xảy ra. Đối với các plasmit đứt quãng gen trên cơ sở pKS và xóa đoạn không làm lệch khung dịch mã, các thể tiếp hợp ngoài được cho qua ba vòng ủ liên tiếp trên đĩa thạch ISP4 để tạo bào tử mà không bổ sung bước chọn lọc kháng sinh bất kỳ để kích thích quá trình chuyển hóa thành các thể tái tổ hợp chéo kép. Các thể tiếp hợp ngoài trên cơ sở pKS không được cho qua bước chọn lọc nhiệt độ 40°C. Cấu trúc chính xác của tất cả các thể đột biến đều được xác nhận bằng PCR và/hoặc phân tích thẩm tách Southern.

Cấu trúc plasmit biểu hiện hợp nhất pXY152-endorf24-camtsr và pXY152-endorf24-blatsr.

Để biểu hiện lệch vị *orf24* trong thể đột biến kháng apramycin mang đoạn xóa *orf18* và các vùng hai bên của nó, plasmit biểu hiện hợp nhất pXY152-endorf24-blatsr được thiết kế. Để tạo cấu trúc plasmit này, cassette (camtsr) mang gen kháng chloramphenicol và gen kháng thiostrepton (*tsr*) được cắt ra từ dẫn xuất plasmit pUC57 bằng cách cắt bằng *SacI* và *NheI*. Tiếp theo, cassette camtsr được gắn với plasmit pXY152-endorf24 đã chuyển thành mạch thẳng bằng *SacI* và *NheI* thu được cấu trúc mới pXY152-endorf24-camtsr (FIG. 7; SEQ ID NO: 20). Gen kháng ampicillin (*bla*) được khuếch đại bằng PCR từ pBluescript KS bằng cách sử dụng đoạn môi xuôi (amp2956SwaIpf, 5'-GTGGCAATTTAAATGGAAATGTGCGCGGAA-3' (SEQ ID NO: 21), vị trí *SwaI* được in đậm) và đoạn môi ngược (amp1973SacIpr, 5'-TATATAGAGCTCAACTTGGTCTGACAGTTAC-3' (SEQ ID NO: 22), vị trí *SacI* được in đậm). *bla* tiếp theo được tách dòng vào các vị trí *SacI* và *SwaI* của pXY152-

endorf24-camtsr để thay thế casset camtsr bằng blatsr. Plasmid biểu hiện tiếp hợp thu được được ký hiệu là pXY152-endorf24-blatsr (FIG.8; SEQ ID NO: 23).

Cấu trúc casset Tn5AT để gây đột biến gen nhảy *in vitro*

Casset Tn5AT được thiết kế để kết hợp ba yếu tố di truyền: gen nhảy Tn5, *oriT* và *aac3(IV)*. Tn5 được nhận biết đặc hiệu và duy nhất bằng Tn5 transposaza (Epicentre) và dễ dàng cài xen vào ADN giàu G + C của *Streptomyces* đã tách dòng trong plasmid và fosmit của *E. coli* (cũng tham chiếu đến patent Mỹ số 8,188,245 được đưa vào để tham khảo). *oriT* cần cho việc chuyển tiếp hợp ADN từ *E. coli* S17-1 sang *Streptomyces* và *aac(3)IV* là chất đánh dấu chọn lọc hai chức năng *E. coli-Streptomyces* tạo tính kháng apramycin. Cả *oriT* và *aac3(IV)* đều được cắt ra từ plasmid pIJ773 dưới dạng mảnh *XbaI* và tiếp theo được tách dòng vào plasmid cho gen nhảy pMODTM-2(MCS) (Epicentre), được chuyển thành mạch thẳng trước đó bằng *XbaI*. Plasmid thu được pXYTn5ATa và pXYTn5ATb chỉ khác nhau bởi hướng của mảnh *XbaI* và được sử dụng để điều chế casset Tn5AT bằng cách cắt bằng *PvuII* theo hướng dẫn của nhà sản xuất.

Đột biến gen nhảy *in vitro* và chọn lọc fosmit được gây đột biến pXYF24D3 và pXYF148D12

Để tạo ra thư viện fosmit được gây đột biến ngẫu nhiên mang các đoạn của cụm sinh tổng hợp enduracidin cho nghiên cứu thay thế gen, nghiên cứu đột biến cài xen gen nhảy *in vitro* của fosmit pXYF24 và pXYF148 được thực hiện. Hai gen điều hòa sinh tổng hợp enduracidin giả định, *orf18* và *orf24*, lần lượt nằm trên các đoạn cài xen của fosmit pXYF24 và pXYF148 (số hiệu lưu giữ Ngân hàng gen DQ403252). Phản ứng gen nhảy *in vitro* được thực hiện ở 37°C trong 2 giờ sau khi trộn 10 µL (0.5 µg) ADN khuôn fosmit, 2 µL (20 ng) ADN casset Tn5AT, 2 µL 10 x dung dịch đệm phản ứng, 1 µL Tn5 transposaza và 5 µL nước vô trùng. Biến nạp tế bào có hiệu lực của *E. coli* EPI300TM-T1^R (Epicentre) bằng hỗn hợp phản ứng gen nhảy được thực hiện bằng điện di. Fosmit được gây đột biến được chọn lọc trên đĩa thạch LB có bổ sung 100 µg/mL apramycin. Đĩa được ủ qua đêm ở 37°C và các khuẩn lạc sống sót được

chọn ngẫu nhiên và được sinh trưởng trong môi trường lỏng LB có bổ sung 100 µg/mL apramycin. ADN fosmit được gây đột biến từ các khuẩn lạc này và fosmit đối chứng pXYF24 hoặc pXYF148 được cắt bằng *Hind*III và được phân tích bằng điện di trên gel agarosa 1%. Casset Tn5AT chứa vị trí *Hind*III duy nhất mà hữu ích khi sàng lọc một số với nhiều sự kiện đứt quãng qua đoạn cài xen fosmit. Không có vị trí *Hind*III nào có mặt trong các đoạn cài xen fosmit pXYF24 hay pXYF148, và chỉ có một vị trí *Hind*III có mặt trong vector fosmit. Do đó, việc cắt bằng *Hind*III dễ dàng nhận diện fosmit có đoạn cài xen Tn5AT duy nhất bằng sự có mặt của hai băng trong gel. Các khuẩn lạc mang fosmit được gây đột biến có đoạn cài xen gen nhảy duy nhất được chọn ngẫu nhiên và được sinh trưởng trong môi trường lỏng LB cho phép phân lập và nhận diện fosmit có gen đứt quãng. Việc sàng lọc được thực hiện bằng cách phân tích trình tự bằng cách sử dụng đoạn mồi 5'-AAGGAGAAGAGCCTTCAGAAGGAA-3' (SEQ ID NO: 24), tương ứng với vùng gen kháng apramycin. Theo cách này, fosmit pXYF24D3 và pXYF148D12 được phát hiện là có Tn5AT được cài xen lần lượt vào *orf18* ở vị trí nucleotit 26386 và *orf24* ở vị trí nucleotit 34333 (số hiệu lưu giữ Ngân hàng gen DQ403252).

Đứt quãng đoạn cài xen *orf18* và *orf24* trong *S. fungicidicus* kiểu đại ATCC 21013.

Fosmit thay thế gen pXYF24D3 và pXYF148D12 được biến nạp riêng rẽ vào *E. coli* S17-1 bằng cách điện di và tiếp theo được đưa vào *S. fungicidicus* bằng tiếp hợp liên giống (Mazodier *et al.*, *J. Bacteriology* 171: 3583-3585, 1989, toàn bộ nội dung của tài liệu này được kết hợp ở đây để tham khảo). Các khuẩn lạc tiếp hợp ngoài sống sót sau quá trình chọn lọc apramycin được cho qua ba vòng tạo bào tử liên tiếp mà không chọn lọc kháng sinh trên đĩa thạch ISP2 để tạo ra chủng đột biến ổn định thông qua sự tái tổ hợp tương đồng chéo kép. Các bào tử thu được được gom lại, pha loãng và cấy lên đĩa thạch ISP2 có bổ sung 50 µg/mL apramycin để xác nhận tính kháng apramycin và để dùng trong nuôi cấy hạt giống và lên men sản xuất enduracidin. Chủng đột biến có đứt quãng đoạn cài xen *orf18* trong *S. fungicidicus* kiểu đại được ký hiệu là SfpXYF24D3 và chủng đột biến có đứt quãng đoạn cài xen *orf24* trong *S. fungicidicus* kiểu đại được ký hiệu là SfpXYF148D12.

Sản xuất enduracidin trên quy mô phòng thí nghiệm và trong bình lên men 10 L.

Các điều kiện lên men trong bình lắc phòng thí nghiệm cho quá trình sản xuất enduracidin ở *S. fungicidicus* kiểu dại, BM38-2 (ATCC PTA-122342) và các chủng dẫn xuất là như được mô tả bởi Higashide *et al.* (*J. Antibiotics*, 21: 126-137, 1968) ngoại trừ môi trường sản xuất enduracidin được bộc lộ trong patent (Patent Mỹ số 4465771). Để lên men trên quy mô phòng thí nghiệm, 5 mL TSB được sử dụng để cấy giống hạt có bào tử streptomycete vừa thu hoạch. Thường là từ 5 đến 10 mL giống hạt được ủ trên thiết bị lắc quay tròn với tốc độ 225 rpm và ở 30°C trong 48 giờ và tiếp theo được chuyển sang môi trường sản xuất enduracidin 50 mL trong 10 ngày lên men liên tục. Việc sản xuất enduracidin bằng các chủng kiểu dại và dẫn xuất trong các điều kiện kiểm soát chặt chẽ cũng được thực hiện trong thiết bị lên men tự động 10 L.

Bảng 2: So sánh hiệu suất enduracidin (enramycin) ở các chủng *Streptomyces fungicidicus* kiểu dại, đột biến và được thao tác di truyền

Chủng <i>S. fungicidicus</i>	Điều kiện lên men	Hiệu suất (HPLC)
Kiểu dại (ATCC21013)	Bình lắc	5-30 mg/L
BM38-2	Bình lắc	60-90 mg/L
SfpXY52endorf24	Bình lắc	60 mg/L
SfpXYF24D3	Bình lắc	40 mg/L
BM38-2.orf18pfrd-AmR	Bình lắc	67 mg/L
BM38-2.24/16	Bình lắc	30-130 mg/L
BM38-2	Thiết bị lên men 10L	80-145 mg/L
BM38-2.24/16	Thiết bị lên men 10L	375 mg/L

Chiết enduracidin từ sản phẩm lên men cho phân tích HPLC.

Để chiết các chất chuyển hóa cho phân tích HPLC về sự sản xuất enduracidin, hệ sợi nấm mới được thu hoạch bằng cách ly tâm và được rửa bằng nước đã loại ion và được tạo huyền phù lại trong 5x thể tích (tỷ lệ metanol trong nước (mL) so với trọng lượng hệ sợi nấm ướt (g)) 70% metanol trong nước (pH được điều chỉnh đến 3,5 bằng 1 N HCl). Huyền phù được lắc với tốc độ 200 rpm ở nhiệt độ trong phòng qua đêm và tiếp theo được ly tâm với tốc độ 4000 rpm và ở 4°C trong 20 phút. Tiếp theo, 1,4 mL dịch nổi từ mỗi mẫu được chuyển sang các ống vi ly tâm 1,5 mL riêng rẽ và được ly tâm với tốc độ 13000 rpm ở nhiệt độ trong phòng trong 10 phút. Phần lọc được cho qua bộ lọc ống tiêm 0,22 μ m và tiếp theo được phân tích bằng HPLC. Việc chiết chất chuyển hóa từ hệ sợi nấm thu được trong thiết bị lên men 10 L được thực hiện trên quy mô nhỏ tương đương với lên men trong phòng thí nghiệm.

Phân tích HPLC và xác định hiệu suất enduracidin.

50 μ L mẫu HPLC điều chế như đã mô tả ở trên được bơm lên cột Gemini C₁₈ (5 μ m, 4,6x150 mm, Phenomenex, Torrance, CA) gắn với Shimadzu HPLC. Việc tách đạt được bằng cách sử dụng gradien tuyến tính theo bậc 18 phút với dung môi A: nước + 0,1% TFA và dung môi B: axetonitril. Lưu lượng là 1 mL/phút bắt đầu bằng 10% B, gia tăng đến 40% B trong 10 phút, và tiếp theo gia tăng tiếp đến 95% B trong 8 phút. Vùng UV từ 200 đến 300 nm được quét bằng bộ dò mạng diot phát quang SPD M20A. Hiệu suất enduracidin được tính bằng cách so sánh với đường cong chuẩn được vẽ từ dung dịch gốc enduracidin chuẩn trong 70% metanol. Loạt bơm bao gồm 2, 4, 6, 8, 10 và 12 μ g enduracidin được sử dụng để vẽ đường cong chuẩn bằng cách sử dụng tổng diện tích hấp thu đối với các enduracidin A và B ở 230 nm. Phương trình hồi quy được tạo ra từ đường cong chuẩn và được sử dụng để tính hiệu suất enduracidin.

Đánh giá hoạt tính kháng khuẩn.

Staphylococcus aureus (ATCC 29213) được sử dụng làm vi sinh vật chỉ thị trong thử nghiệm sinh học. Các tế bào được sử dụng để cấy môi trường canh LB, được sinh trưởng ở 37°C qua đêm, và tiếp theo 100 μ L môi trường nuôi cấy được trộn với 5 mL lớp thạch trên cùng (hỗn hợp gồm các thể tích bằng nhau của thạch dinh dưỡng và môi

trường canh dinh dưỡng). Lớp thạch trên cùng được phủ lên đĩa thạch dinh dưỡng trong đó các giếng có khoảng cách thích hợp được tạo ra bằng cách cắt các nút thạch. Các mẫu chuẩn enduracidin và các phần mẫu chiết nuôi cấy được hòa tan hoặc được pha loãng trong MeOH 50% ở nồng độ 20 $\mu\text{g/mL}$, và 100 μL mỗi dung dịch được nạp vào các giếng. Sau khi ủ đĩa ở 37°C trong 16 giờ, các vùng ức chế được quan sát và so sánh, và các đĩa được chụp ảnh hoặc được bảo quản ở 4°C.

Ví dụ 2

Làm đứt quãng *orf18* và *orf24* trong *S. fungicidicus* kiểu dại và ảnh hưởng đến việc sản xuất enduracidin

Ví dụ này mô tả việc làm đứt quãng *orf18* và *orf24* trong *S. fungicidicus* kiểu dại và ảnh hưởng đến việc sản xuất enduracidin.

Trình tự ADN 116000 bp từ *S. fungicidicus* kiểu dại ATCC 21013 mang cụm gen sinh tổng hợp enduracidin và các vùng hai bên của nó (Patent Mỹ số 8,188,245, toàn bộ nội dung của tài liệu này được kết hợp ở đây để tham khảo) được nhận diện trước đó và có thể sử dụng trong Ngân hàng gen (số hiệu lưu giữ DQ403252). Trong số 48 orf được chú thích có 8 gen điều hòa giả định: *orf5*, *orf12*, *orf18*, *orf22*, *orf24*, *orf41*, *orf42* và *orf43*. Để giải mã vai trò của mỗi sản phẩm gen trong sản xuất enduracidin, các đoạn cài xen fosmit mang các đoạn của cụm enduracidin chứa các gen điều hòa giả định này được gây đột biến ngẫu nhiên bằng cách sử dụng phương pháp cài xen qua trung gian gen nhảy gen chỉ thị kháng apramycin như được mô tả trong Ví dụ 1.

Bước sàng lọc tiếp theo về tính kháng apramycin và vị trí cài xen trong số các khuẩn lạc *E. coli* mang các fosmit được gây đột biến nhận diện pXYF24D3 mang *orf18* đứt quãng và pXYF148D12 mang *orf24* đứt quãng. Đột biến cài xen duy nhất trong mỗi fosmit này và vị trí cài xen được xác nhận bằng cách xác định trình tự. Tiếp theo, hai fosmit được gây đột biến này được đưa riêng rẽ vào chủng *S. fungicidicus* kiểu dại bằng cách tiếp hợp. Tiếp theo, các thể tiếp hợp ngoài thể hiện tính kháng apramycin được cho qua ba vòng tạo bào tử trên thạch ISP2 mà không bổ sung bước chọn lọc kháng sinh bất kỳ để thúc đẩy sự chuyển hóa của quá trình tái tổ hợp tương đồng chéo đơn thành đột

biến chéo kép. Các chủng đột biến ổn định thu được SfpXYF24D3 và SfpXYF148D12 được lên men trong môi trường sản xuất enduracidin (EPM) trên quy mô phòng thí nghiệm trong bình lắc. Phân tích HPLC đối với hệ sợi nấm chiết bằng metanol 70% từ 10 ngày lên men cho thấy hiệu suất enduracidin bởi chủng SfpXYF24D3 có *orf18* bị đứt quãng gia tăng 1,3 lần và sự sản xuất enduracidin bởi chủng SfpXYF148D12 có *orf24* bị đứt quãng là mất hoàn toàn. Phần chiết hệ sợi nấm cũng được đánh giá về hoạt tính đối với *S. aureus*. SfpXYF24D3 có *orf18* bị đứt quãng vẫn giữ được hoạt tính trong khi SfpXYF148D12 có *orf24* bị đứt quãng mất hoàn toàn hoạt tính đối với *S. aureus*.

Ví dụ 3

Cấu trúc của chủng tái tổ hợp SfpXY152-endorf24 và ảnh hưởng đến việc sản xuất enduracidin

Ví dụ này mô tả việc cấu trúc của chủng tái tổ hợp SfpXY152-endorf24 và khả năng chủng này tạo ra enduracidin.

Việc mất khả năng sản xuất enduracidin ở chủng đột biến SfpXYF148D12 chỉ ra vai trò điều hòa có thể có đối với *orf24*. Tìm kiếm BLAST trên cơ sở dữ liệu của Ngân hàng gen bằng cách sử dụng trình tự protein *orf24* cho thấy mức giống nhau cao về trình tự so với protein điều hòa đặc hiệu với con đường, StrR, protein này tham gia vào quá trình sinh tổng hợp streptomycin. Việc sắp xếp hàng trình tự giữa Orf24 và StrR cho thấy các protein này có mức giống nhau đáng kể (độ đồng nhất axit amin 54% , FIG. 9). Việc mất khả năng sản xuất enduracidin khi đứt quãng *orf24* và sự giống với StrR chỉ ra rằng Orf24 có thể đóng vai trò làm yếu tố hoạt hóa đặc hiệu với con đường trong sản xuất enduracidin.

Để nghiên cứu vai trò của *orf24* làm đích điều hòa dương đối với việc cải thiện chủng, plasmit biểu hiện hợp nhất pXY152-endorf24 (FIG. 2) được tạo cấu trúc (Ví dụ 1). Plasmit pXY152-endorf24 được đưa vào *S. fungicidicus* kiểu dại bằng cách tiếp hợp và các thể tiếp hợp ngoài được sàng lọc về kiểu hình kháng apramycin, dẫn đến nhận diện chủng tái tổ hợp mới SfpXY152-endorf24. Ít nhất 10 khuẩn lạc tiếp hợp ngoài độc lập từ chủng này được chọn ngẫu nhiên và được tinh chế. Các chủng khuẩn lạc này

mang plasmid pXY152-endorf24 được hợp nhất vào vị trí *attB* trên nhiễm sắc thể *S. fungicidicus* bằng cách tái tổ hợp tương đồng chéo đơn với vị trí *attP* trên plasmid này.

Để nghiên cứu các chất chuyển hóa được sản xuất bởi các chủng tái tổ hợp, các bào tử từ hai chủng khuẩn lạc được ủ trong môi trường nuôi cấy giống TSB và tiếp theo được chuyển sang môi trường sản xuất enduracidin để lên men quy mô phòng thí nghiệm. Phân tích HPLC đối với hệ sợi nấm chiết bằng metanol 70% cho thấy mức sản xuất enduracidin tăng hai lần (60 mg/L) bởi cả hai chủng tái tổ hợp so với chủng kiểu dại (30 mg/L). Hiệu suất enduracidin gia tăng quan sát được ở hai chủng khuẩn lạc này có khả năng biểu hiện quá mức *orf24* là bằng chứng khác về vai trò điều hòa dương gen này có được trong sản xuất enduracidin và các kết quả này là phù hợp với các kết quả thu được từ sự đứt quãng *orf24* dẫn đến mất khả năng sản xuất enduracidin.

Ví dụ 4

Cấu trúc của chủng BM38-2.24/16 biểu hiện quá mức *orf24* trong *S. fungicidicus* BM38-2 (ATCC PTA-122342) và ảnh hưởng đến việc sản xuất enduracidin

Ví dụ này mô tả cấu trúc của chủng BM38-2.24/16 (Nộp lưu ATCC số PTA-124006), biểu hiện quá mức *orf24* trong *S. fungicidicus* BM38-2 (ATCC PTA-122342) và ảnh hưởng đến việc sản xuất enduracidin.

Để nghiên cứu tiếp vai trò điều hòa dương của Orf24, plasmid pXY152-endorf24 được hợp nhất vào nhiễm sắc thể của chủng sản xuất thương mại *S. fungicidicus* BM38-2 (ATCC PTA-122342), như được mô tả ở trên đối với sinh vật kiểu dại. Việc chọn các thể tiếp hợp ngoài biểu hiện kiểu hình kháng apramycin thu được một số chủng khuẩn lạc tái tổ hợp, bao gồm *S. fungicidicus* BM38-2.24/16, có khả năng sản xuất enduracidin ở mức cao lên đến 200 mg/L (tăng 3,3 lần so với BM38-2 (ATCC PTA-122342)) trong các quy trình nuôi cấy bình lắc phòng thí nghiệm. *S. fungicidicus* BM38-2-24/16 được chọn để đánh giá tiếp về khả năng sản xuất enduracidin trên cơ sở hiệu suất trong quá trình sàng lọc sơ bộ.

Sự sản xuất enduracidin bởi chủng tái tổ hợp *S. fungicidicus* BM38-2.24/16 trong các quy trình nuôi cấy trong bình lắc phòng thí nghiệm cho thấy rõ khả năng cải thiện đáng kể so với BM38-2 (ATCC PTA-122342) và hiệu suất cũng được ghi nhận là thay đổi lớn. Để kiểm soát chặt chẽ hơn các điều kiện nuôi cấy trong khoảng thời gian sinh trưởng 10 ngày, bao gồm độ pH và oxy hòa tan mà không dễ kiểm soát trong bình lắc, sự sản xuất được đánh giá qua nhiều mẻ trong các bình lên men 10L. Trong các điều kiện được kiểm soát chặt chẽ hơn này, hiệu suất là ổn định hơn và trung bình của ba mẻ lên men 10L là 375 mg/mL (gấp 4,6 lần BM38-2 (ATCC PTA-122342)). Hiệu suất enduracidin gia tăng ở chủng tái tổ hợp *S. fungicidicus* BM38-2.24/16 (Nộp lưu ATCC số PTA-124006) củng cố thêm vai trò điều hòa dương tăng của Orf24 trong sản xuất enduracidin.

Ví dụ 5

Cấu trúc của chủng đột biến mất đoạn BM38-1.18pfrd-AmR và ảnh hưởng đến việc sản xuất enduracidin

Ví dụ này mô tả cấu trúc chủng đột biến mất đoạn BM38-2.18pfrd-AmR (Nộp lưu ATCC số PTA-124007) và ảnh hưởng đến việc sản xuất enduracidin.

orf18 nằm ở vùng ngược dòng của cụm gen sinh tổng hợp enduracidin (số hiệu lưu giữ Ngân hàng gen DQ403252). Orf18 dường như có vai trò âm trong sản xuất enduracidin do sự đứt quãng cài xen của gen ở chủng đột biến SfpXYF24D3 làm gia tăng hiệu suất enduracidin. Trên cơ sở quan sát này, các cấu trúc được thiết kế để xóa đoạn *orf18* một mình và *orf18* và các phần của các vùng hai bên của nó. Nhằm mục đích này, plasmit pKS-T-*orf18pfrd-AmR* được tạo cấu trúc (FIG. 5). Plasmit có nguồn gốc từ vector pKS này không có replicon của streptomycete cũng như không phải yếu tố để hợp nhất vào nhiễm sắc thể của streptomycete. Nó chỉ có thể trao đổi đoạn cài xen của nó với đoạn xác định của ADN trong nhiễm sắc thể vật chủ nhờ tái tổ hợp tương đồng chéo kép. Bản đồ cài xen của plasmit này được thể hiện trên FIG. 10. *orf18* và các vùng hai bên của nó chứa toàn bộ *orf19* và vùng mã hóa phần đầu tận cùng N của *orf17* bị xóa trong plasmit pKS-T-*orf18pfrd-AmR*. Nhánh trái 1-kb chứa vùng mã

hóa phần đầu tận cùng C của *orf17* và vùng xuôi dòng của nó và nhánh phải 1-kb chứa một phần đoạn *orf20* mã hóa vùng đầu tận cùng N. Do đó, việc xóa đoạn sau tái tổ hợp tương đồng chéo kép tạo ra chủng tái tổ hợp mà trong đó toàn bộ *orf18* + *orf19* và vùng mã hóa phần đầu tận cùng N của *orf17* bị xóa và được thay bằng gen kháng apramycin.

Plasmid pKS-T-*orf18pfrd*-AmR được đưa bằng cách tiếp hợp vào *S. fungicidicus* BM38-2 (ATCC PTA-122342) và tái tổ hợp tương đồng chéo đơn và kép được thúc đẩy trên đĩa thạch ISP4 mà không bổ sung apramycin. Các thể tiếp hợp ngoài mà có khả năng sống sót trước việc chọn lọc apramycin tiếp theo được tinh chế và chủng tái tổ hợp mới này được ký hiệu là BM38-2.18pfrd-AmR (Nộp lưu ATCC số PTA-124007). Bào tử từ chủng này được cấy vào môi trường TSB để nuôi cấy hạt và tiếp theo được chuyển vào môi trường sản xuất enduracidin. Sau 10 ngày lên men, hệ sợi nấm được thu hoạch, được xử lý và được phân tích bằng HPLC. So với chủng gốc BM38-2 (ATCC PTA-122342), mức sản xuất enduracidin tăng 1,2 lần được ghi nhận từ các mẻ lên men quy mô phòng thí nghiệm này. Hiệu suất tăng tương đối là giống với hiệu suất ghi nhận được với chủng có nguồn gốc từ kiểu đại SfpXYF24D3 và kết quả chỉ ra rằng *orf19* và *orf17*, nằm hai bên *orf18* và bị tác động trong cấu trúc của BM38-2.18pfrd-AmR, ít hoặc không ảnh hưởng đến việc sản xuất enduracidin. Do đó, mức sản xuất enduracidin gia tăng ở chủng tái tổ hợp BM38-2.18pfrd-AmR là do việc loại bỏ vai trò điều hòa âm của Orf18.

Liên quan đến việc xóa đoạn *orf18* một mình với plasmid pXY300-*orf18ifd* trong BM38-2 (ATCC PTA-122342), các khó khăn gặp phải khi chọn lọc dương các thể tiếp hợp ngoài và các thể đột biến đơn/kép với gen đánh dấu kháng thiostrepton. Do đó, vector thay thế pBluescript KS II được sử dụng để cấu trúc plasmid phân phối thay thế gen không đánh dấu như pKS-T-*orf18ifd* (FIG. 4) hoặc pKS-*orf18ifd*-T, pKS-*orf18ifd*-T-AmR(NS) (gen kháng apramycin được thực hiện trên vector này thay vì cài vào *orf18*, xem FIG. 6).

Ví dụ 6

Phát triển loạt vector bất hoạt gen có nguồn gốc từ pKS, vector pKS-T-orf18pfrd-AmR

Ví dụ này mô tả việc phát triển loạt vector bất hoạt gen có nguồn gốc từ pKS pKS-T-orf18pfrd-AmR.

Loạt vector bất hoạt gen có nguồn gốc từ pKS đã được phát triển (FIG. 4, 5 và 6) có chức năng tiếp hợp và không cần qua các thể biến nạp nhờ chọn lọc nhiệt độ cao để loại bỏ plasmit như các vector đứt quãng gen khác nhất định cần đến. Các vector có nguồn gốc từ pKS này mang replicon phi-streptomycete cho phép sao chép ở *E. coli* và có thể duy trì và được chọn lọc bằng gen đánh dấu kháng apramycin ở *Streptomyces* và *E. coli* hoặc ampicillin ở *E. coli*. Chúng sản xuất nhiều bản sao plasmit tái tổ hợp ổn định ở *E. coli* cho việc tiếp hợp và chúng đã được thiết kế với một vài vị trí giới hạn hiếm và duy nhất được tìm thấy trong ADN của streptomycete, như *PacI*, *HindIII*, *NheI*, và *XbaI*, các vị trí này có thể được sử dụng thích hợp để lắp ráp ADN đích vào plasmit cho các nghiên cứu đứt quãng gen cài và xóa đoạn không làm lệch khung dịch mã.

Ví dụ 7

Phát triển các vector biểu hiện gen hợp nhất có nguồn gốc từ pSET152, pXY152-endorf24-camtsr và pXY152-endorf24-blatsr

Ví dụ này mô tả việc phát triển các vector biểu hiện gen hợp nhất có nguồn gốc từ pSET152, là các vector pXY152-endorf24-camtsr (SEQ ID NO: 20) và pXY152-endorf24-blatsr (SEQ ID NO: 23).

Hai vector mới, pXY152-endorf24-camtsr (FIG. 7) và pXY152-endorf24-blatsr (FIG. 8) được phát triển. Chúng có chức năng tiếp hợp và hợp nhất giống như vector pSET152, vector hợp nhất được sử dụng rộng rãi nhất cho việc biểu hiện và hỗ trợ gen streptomycete. Cả hai vector này đều mang một vài vị trí giới hạn hiếm trong ADN của *Streptomyces* thích hợp cho việc tách dòng và lắp ráp cấu trúc biểu hiện. Vector pXY152-endorf24-camtsr có thể được duy trì và chọn lọc ở *E. coli* bằng cloramphenicol ở nồng độ 12,5 µg/mL và ở *Streptomyces* bằng thiostrepton ở nồng độ

50 µg/mL. Vector pXY152-endorf24-blatsr có thể được duy trì và chọn lọc ở *E. coli* bằng ampicillin và ở *Streptomyces* bằng thiostrepton.

Tóm tắt các ví dụ từ 1 đến 6:

Thao tác di truyền các gen điều hòa và sinh tổng hợp của *Streptomyces* cho việc cải thiện chủng

Trong số các vi sinh vật sản xuất sản phẩm tự nhiên, khoảng 75% chất kháng sinh của vi sinh vật là được sản xuất bởi actinomycetes. *Streptomyces*, vi khuẩn đất Gram dương dạng sợi, là thành viên của họ actinomycete và được biết là vô địch về khả năng tạo ra vô số chất chuyển hóa bậc hai có hoạt tính dược lý và sinh học, đa dạng về cấu trúc. Polyketit được sản xuất bằng polyketit synthaza (PKS) và các sản phẩm tự nhiên peptit được sản xuất bằng peptit synthetaza ngoài ribosom (NRPS) là các đại diện.

Nghiên cứu về sinh tổng hợp chất kháng sinh tự nhiên có thách thức chung nhất định: trước hết, cách khắc phục mức sản xuất thường thấp của các hợp chất gốc hoặc được cải biến cấu trúc được sản xuất bởi các chủng kiểu đại hoặc được thao tác di truyền; thứ hai, cách hoạt hóa nhiều con đường sinh tổng hợp chất chuyển hóa bậc hai bí ẩn hoặc mờ côi nhận dạng được từ các trình tự hệ gen để có thể nghiên cứu chức năng sinh học của các sản phẩm này. Tiến bộ trong nghiên cứu sinh tổng hợp chất kháng sinh tự nhiên trong hàng thập kỷ qua chỉ ra rằng việc sản xuất chất chuyển hóa bậc hai được điều hòa bởi nhiều con đường. Ví dụ, các gen sinh tổng hợp tiền chất và lắp ráp cấu trúc (như PKS và NRPS), các gen điều hòa và các gen tự kháng có thể được tạo cụm trên nhiễm sắc thể vi khuẩn. Việc sản xuất chất kháng sinh có thể được điều hòa bởi các gen điều hòa đặc hiệu với con đường, bao gồm các gen hoạt hóa và/hoặc các gen ức chế, các gen điều hòa lệch vị đa hướng, và các hệ thống điều hòa hai thành phần. Các đột biến xảy ra trong gen hoặc hệ thống bất kỳ trong số các gen hoặc hệ thống điều hòa này đều có thể làm gia tăng, làm giảm hoặc loại bỏ hoàn toàn việc sản xuất chất kháng sinh. Con đường sinh tổng hợp đa hướng có thể được hoạt hóa bởi đột biến không dự đoán được dẫn đến việc sản xuất sản phẩm chưa từng biết đến trước đó.

Việc cải thiện chủng đóng vai trò quan trọng trong sản xuất chất kháng sinh hoặc các chất chuyển hóa bậc hai khác của vi sinh vật ở quy mô công nghiệp có hiệu quả kinh tế. Các chủng đột biến có thể gia tăng hiệu suất chất chuyển hóa cụ thể có thể được tạo ra nhờ các đột biến ngẫu nhiên hoặc bằng cách làm đứt quãng có chủ đích các gen đặc hiệu hoặc bằng cách đưa (các) gen loại bỏ các điểm nghẽn vào con đường sinh tổng hợp. Thao tác di truyền các gen điều hòa dương và âm, cũng như các gen sinh tổng hợp, tạo ra mức sản xuất chất chuyển hóa bậc hai đích cao hơn đã được chứng minh là chiến lược mạnh mẽ và rất thành công trong cải thiện chủng actinomycete.

Trong bản mô tả này, vai trò điều hòa dương của *orf24* và vai trò điều hòa âm của *orf18* đối với việc sản xuất enduracidin đã được chứng minh. Việc bất hoạt đoạn cài xen có chủ đích *orf24* dẫn đến làm mất hoàn toàn sự sản xuất enduracidin ở chủng tái tổ hợp SfpXYF148D12. Việc biểu hiện quá mức *orf24* sau đó dưới sự kiểm soát của gen khởi đầu cơ định mạnh *ermE**p ở các chủng tái tổ hợp SfpXY152-endorf24 và BM38-2.24/16 dẫn đến hiệu suất enduracidin tăng gấp khoảng 2 đến 4,6 lần. Việc xóa đoạn *orf18* và các vùng hai bên của nó, bao gồm toàn bộ *orf19* và một phần của *orf17*, làm tăng hiệu suất enduracidin 1,2 lần. Các kết quả này cung cấp bằng chứng di truyền mạnh mẽ củng cố vai trò của *orf24* và *orf18* lần lượt làm gen hoạt hóa dương và gen ức chế âm, trong sinh tổng hợp enduracidin.

Các gen trực giao *orf24* đã được xác nhận về mặt chức năng từ các con đường sinh tổng hợp chất kháng sinh khác

Truy vấn BLAST với trình tự protein Orf24 theo cơ sở dữ liệu ngân hàng gen chỉ ra hàng trăm kết quả đúng (số hiệu lưu giữ ngân hàng gen DQ403252). Nhiều kết quả thể hiện mức giống nhau về axit amin rất cao (độ đồng nhất từ 60% đến 99%) và được chú giải là các yếu tố điều hòa phiên mã trong sinh tổng hợp chất kháng sinh aminoglycosit streptomycin. Tuy nhiên, không có gen nào trong nhóm gen này có chức năng được xác nhận bằng thực nghiệm. Phân tích kết quả BLAST nhận diện được một vài protein có liên quan có chung mức giống nhau thấp hơn (độ đồng nhất axit amin lớn hơn 40% nhưng nhỏ hơn 60%) với Orf24 đã được xác định chức năng. Các protein

này bao gồm protein StrR đã biết rõ có chung mức giống nhau thấp hơn nhưng đáng kể (độ đồng nhất axit amin 54% ở 311 axit amin chồng lấp) với Orf24. StrR đã được chứng minh về mặt di truyền và sinh hóa là hoạt động dưới dạng yếu tố hoạt hóa dương đặc hiệu với con đường trong việc biểu hiện các gen sinh tổng hợp streptomycin ở *Streptomyces griseus*. StrR đại diện cho họ yếu tố hoạt hóa đặc hiệu với con đường, một số ít trong số này được xác định bằng thao tác di truyền hoặc nghiên cứu sinh hóa. FIG. 11 thể hiện sự sắp xếp Orf24 với sáu protein kiểu StrR của actinomycete đã được xác định chức năng. Miền liên kết với ADN xoắn-ngoặt-xoắn (helix-turn-helix: HTH) điển hình và bảo toàn cao có mặt trong cả bảy protein như được gạch chân trên FIG. 11. Orf24 cũng có chung mức giống nhau về trình tự đáng kể (độ đồng nhất axit amin 54%) với Teil5*, yếu tố hoạt hóa đặc hiệu với con đường điều tiết quá trình sinh tổng hợp của chất kháng sinh glycopeptit được tạo ra ngoài ribosom teicoplanin. Teil5* điều hòa dương quá trình phiên mã của ít nhất 17 gen trong cụm teicoplanin. *Actinoplanes teichomyceticus* kiểu dại tạo ra khoảng 100 mg/L teicoplanin trong khi các chủng tái tổ hợp di truyền, có nguồn gốc từ *A. teichomyceticus* gốc và mang *teil5** được biểu hiện dưới sự kiểm soát của các gen khởi đầu khác nhau, gia tăng hiệu suất teicoplanin đến 1 g/L trong trường hợp của gen khởi đầu *ermE**p và đến 4 g/L trong trường hợp của gen khởi đầu kháng apramycin nguyên bản.

Như được minh họa trên FIG. 11, Orf24 cũng có chung mức giống nhau về trình tự đáng kể (độ đồng nhất axit amin 54%) với Bbr, từ cụm sinh tổng hợp chất kháng sinh glycopeptit balhimycin; với KasT (độ đồng nhất axit amin 50%) điều tiết sự biểu hiện của các gen sinh tổng hợp chất kháng sinh aminoglycosit kasugamycin; và yếu tố hoạt hóa đặc hiệu với con đường NovG (độ đồng nhất axit amin 45%) tham gia vào quá trình sinh tổng hợp novobiocin. Thử đột biến $\Delta novG$ chỉ tạo ra 2% novobiocin so với kiểu dại và sự biểu hiện quá mức *novG* từ plasmit nhiều bản sao ở chủng tái tổ hợp dẫn đến làm tăng mức sản xuất novobiocin lên 3 lần. Orf24 cũng có chung độ đồng nhất axit amin 42% với SgcR1, một trong bốn gen điều hòa (*sgcR1*, *sgcR2*, *sgcR3* và *sgcR*) được xác nhận bằng thực nghiệm là tham gia vào quá trình sản xuất chất kháng sinh kháng u C-1027 ở *S. globisporus*. Sự biểu hiện quá mức *sgcR1* ở *S. globisporus*

SB1022 làm tăng hiệu suất C-1027 khoảng 7 lần so với chủng kiểu dại. Sự biểu hiện quá mức gen điều hòa dương *sgcR3* ở chủng tái tổ hợp dẫn đến làm tăng mức sản xuất C-1027 lên 30-40%. Trái lại, sự bất hoạt gen điều hòa âm *sgcR* dẫn đến làm tăng sản xuất cả C-1027 và heptaen. Hơn nữa, sự biểu hiện quá mức *sgcR1* ở chủng đột biến Δ *sgcR* dẫn đến làm tăng mức sản xuất C-1027 khoảng 7 lần. *sgcR3* điều hòa ở mức cao hơn do sự kiểm soát của *sgcR1* và *sgcR2* trong điều hòa phân cấp đối với sự sản xuất C-1027. Tóm lại, các tác dụng làm đứt quãng và biểu hiện của *orf24* và việc so sánh Orf24 với các gen trực giao đã xác định chức năng khác cho thấy Orf24 đóng vai trò làm gen điều hòa/hoạt hóa dương đặc hiệu với con đường trong sản xuất enduracidin.

Orf18 là gen điều hòa đáp ứng mô côi không điển hình giả định và sắp thẳng hàng với các gen trực giao đã xác nhận chức năng

Việc sản xuất chất kháng sinh ở loài *Streptomyces* được điều hòa chặt chẽ bởi mạng lưới di truyền phức tạp làm giới hạn khả năng của nhiều loài sản xuất chất kháng sinh kiểu dại từ việc tạo ra sản lượng cần cho sản xuất công nghiệp quy mô lớn, có hiệu quả kinh tế. Một cơ chế điều hòa quan trọng là hệ thống dẫn truyền tín hiệu hai thành phần. Hệ thống hai thành phần bao gồm kinaza cảm biến và yếu tố điều hòa đáp ứng cùng nguồn. Kinaza cảm biến đáp ứng với các yếu tố kích thích/tín hiệu của môi trường bên ngoài cụ thể như căng thẳng, dinh dưỡng và hóa chất, v.v., và tiếp theo chuyển tiếp tín hiệu này đến yếu tố điều hòa đáp ứng tế bào chất làm khởi động và hoạt hóa quá trình phiên mã của các gen đích. Yếu tố điều hòa đáp ứng không bắt cặp với kinaza cảm biến được gọi là yếu tố điều hòa đáp ứng mô côi.

Hệ thống hai thành phần và yếu tố điều hòa đáp ứng mô côi có mặt trong hệ gen của streptomycete và có thể có chức năng ức chế quá trình sản xuất chất chuyển hóa bậc hai. Trong cụm gen enduracidin từ *S. fungicidicus*, *orf18* mã hóa yếu tố điều hòa đáp ứng mô côi giả định có chung mức giống nhau về trình tự từ thấp đến vừa phải với ba yếu tố điều hòa đáp ứng khác đã được xác định của *Streptomyces* bao gồm một yếu tố điều hòa đáp ứng mô côi, SCO3818, từ *S. coelicolor* (FIG. 12). Orf18 có trình tự đầu tận cùng N dài hơn so với các protein sắp thẳng hàng khác và dường như là yếu tố điều

hòa đáp ứng mô côi không điển hình do lysin bảo toàn cao ở vị trí 118 (tương ứng với vị trí 105 thông thường) không có mặt trong Orf18 và được thay bằng threonin. Lysin này được cho là cần cho việc tạo túi phosphoryl hóa.

Chỉ có một số yếu tố điều hòa đáp ứng của streptomycete được xác định chức năng. Hệ gen của *S. coelicolor* chứa tổng cộng năm yếu tố điều hòa đáp ứng mô côi không điển hình và bảy yếu tố điều hòa đáp ứng mô côi điển hình. Orf18 có chung độ đồng nhất axit amin 26% trong 191 axit amin chồng lấp với AbsA2. Việc xóa đoạn AbsA2 ở *S. coelicolor* dẫn đến gia tăng sản xuất hai chất kháng sinh, actinorhodin và undecylprodigiosin. Orf18 có độ đồng nhất axit amin 32% trong 176 axit amin chồng lấp với SCO3818. Việc xóa đoạn *sco3818* dẫn đến tăng cường sản xuất actinorhodin. Orf18 có chung độ đồng nhất axit amin 29% trong 166 axit amin chồng lấp với SCO1745 (AbrA2). Việc xóa đoạn AbrA2 chứa operon điều hòa đáp ứng dẫn đến tăng 100% chất kháng sinh kháng u oviedomycin ở chủng tái tổ hợp *S. coelicolor* M145 so với chủng sản xuất kiểu dại. Vai trò điều hòa âm quan sát được này của Orf18 trong sản xuất enduracidin là phù hợp với các hoạt tính đã được chứng minh của các yếu tố điều hòa âm liên quan (FIG. 12). Ngoài ra, lưu ý rằng Orf18 có chung mức giống nhau về trình tự protein cao nhất với các thành viên của họ LuxR là các yếu tố điều hòa phiên mã trong phép tìm kiếm BLAST.

Không có tác dụng phân cực ở BM38-1.orf18pfrd-AmR đột biến

Vùng bị xóa bỏ trong chủng BM38-1.18pfrd-AmR đột biến bao gồm ba gen, *orf18*, vùng mã hóa phần đầu tận cùng N của *orf17* nằm xuôi dòng *orf18*, và toàn bộ *orf19* nằm ngược dòng *orf18* (FIG. 5 và 10). *orf17* được dự đoán là mã hóa ribonucleaza đường như không có chức năng nào liên quan đến quá trình sinh tổng hợp hoặc điều hòa enduracidin. Tương tự, gen kháng apramycin thay thế *orf18* và vùng hai bên nó được phiên mã phân kỳ với *orf17* và không thể tạo ra các sự kiện đọc hết bất kỳ từ gen

khởi đầu kháng apramycin. Do đó, không có tác dụng phân cực nào tạo ra từ việc xóa một phần *orf17*.

orf19 được phiên mã và dịch mã theo hướng giống như *orf18*. Gen này được chú giải là mã hóa protein chưa biết chức năng. Chủng đột biến SfpXYF24D3 mang gen đứt quãng *orf18* một mình và BM38-1.18pfrd-AmR đột biến mang đoạn mất *orf18* và *orf19* cùng nhau có các tác dụng tăng cường giống nhau đối với việc sản xuất enduracidin ám chỉ rằng *orf19* không có vai trò hoặc có vai trò không đáng kể trong sản xuất enduracidin. Gene *orf20* nằm ngược dòng *orf19* được phiên mã và dịch mã theo hướng giống như gen chỉ thị kháng apramycin được cài xen vào (FIG. 10) *orf20* vẫn nguyên vẹn trong BM38-1.18pfrd-AmR và sản phẩm này dường như không có vai trò trong sản xuất enduracidin. Do đó tác dụng phân cực bất kỳ đến việc biểu hiện của *orf20* không được tin là có vai trò tăng cường sản xuất enduracidin ở BM38-1.18pfrd-AmR.

Ví dụ 7

Các ứng dụng và thao tác khác đối với *orf24* và/hoặc *orf18* đối với các chủng sản xuất enduracidin tăng cường

Ngoài các ví dụ nêu trên, có thể có một số cách khác để sử dụng vai trò điều hòa của *orf24* và *orf18* để cải thiện mức sản xuất enduracidin.

i. Biểu hiện *orf24* dưới gen khởi đầu thay thế, cơ định hoặc biểu hiện quá mức cảm ứng được

pXY152-endorf24 (được thể hiện trên FIG. 2) được tạo cấu trúc để biểu hiện lạc vị hợp nhất *orf24* dưới sự kiểm soát của *ermE**p, một gen khởi đầu biểu hiện cơ định mạnh của streptomycete được sử dụng rộng rãi. Sự biểu hiện quá mức của *orf24* cũng có thể được điều khiển bởi các gen khởi đầu cơ định hoặc cảm ứng được khác. Gen khởi đầu *tipA* là gen khởi đầu của streptomycete biểu hiện quá mức có thể cảm ứng bởi thiostrepton. Plasmid con thoi *E. coli-Streptomyces* chứa gen khởi đầu *tipA* nhiều bản sao, pXY200, được phát triển mà đã được sử dụng thành công để biểu hiện quá mức

gen streptomycete. Đối với các ứng dụng liên quan đến sáng chế, gen khởi đầu *tipA* có thể được cắt từ pXY200 và được tách dòng vào pXY152 để thay thế *ermE**p và điều khiển quá trình biểu hiện của *orf24*. Tương tự, *orf24* có thể dễ dàng được chuyển từ pXY152-endorf24 sang pXY200 cho việc biểu hiện trên cơ sở plasmit. Các tùy chọn gen khởi đầu khác bao gồm, nhưng không giới hạn ở, hệ thống P(*nitA*)-NitR và gen khởi đầu streptomycete SF14. Mới đây, plasmit hợp nhất pKC1139 và gen khởi đầu nguyên bản của gen kháng apramycin đã được sử dụng thành công để biểu hiện các gen điều hòa để tăng sản xuất chất kháng sinh peptit teicoplanin. Gen điều hòa *sanG* mã hóa yếu tố hoạt hóa đặc hiệu với con đường để sản xuất nikkomycin. Việc biểu hiện bản sao phụ của *sanG* dưới sự kiểm soát của 5 gen khởi đầu khác nhau (P_{hrdB} , P_{tcp830} , P_{SF14} , P_{ermE*} và P_{neos}) dẫn đến gia tăng hiệu suất nikkomycin lần lượt đến 69%, 51%, 26%, 22%, và 13% (xem Du *et al.*, Applied Microbiology and Biotechnology 97: 6383-6396, 2013).

ii. Chủng đột biến kép *S. fungicidicus* có xóa đoạn *orf18* và sự biểu hiện quá mức *orf24*

Với cả hai chủng đột biến xóa đoạn *orf18* và biểu hiện quá mức *orf24* biểu hiện mức sản xuất enduracidin gia tăng, chủng đột biến kép chứa cả hai có thể được tạo ra và quan sát xem liệu có tác dụng phụ nào đến hiệu suất của chất kháng sinh peptit này hay không. Chủng đột biến kép có thể được tạo ra bằng cách đưa plasmit pXY152-endorf24-blatsr biểu hiện quá mức (FIG. 8) vào BM38-2.18pfrd-AmR đột biến. pXY152-endorf24-blatsr là plasmit hợp nhất tiếp hợp mang gen kháng thiostrepton (*tsr*) để chọn lọc ở streptomycetes và gen kháng ampicillin (*bla*) để chọn lọc ở *E. coli*. Do chủng *E. coli* S17-1 dùng để tiếp hợp là kháng cloramphenicol (*cam*) một cách tự nhiên, nên gen đánh dấu kháng cloramphenicol trong pXY152-endorf24-camtsr (xem ở trên) đã được thay thế bằng kháng ampicillin (*bla*) để chọn các thể biến nạp S17-1. Theo cách khác, pXY152-endorf24-camtsr và các dẫn xuất có thể được đưa vào streptomycetes bằng cách sử dụng chủng *E. coli* tiếp hợp khác, ET12567/pUZ8002.

Bằng cách sử dụng plasmit pXY152-endorf24-blatsr hoặc pXY152-endorf24-camtsr để đưa bản sao thứ hai của *orf24* vào thể đột biến khuyết *orf18*, có thể chọn được thể đột biến kép bởi tính kháng thiostrepton. Để tạo ra thể đột biến xóa đoạn không làm lệch khung dịch mã *orf18* bất hoạt BM38-2 (ATCC PTA-122342), plasmit pXY300-*orf18ifd* (FIG. 3) và pKS-*orf18ifd*-T-AmR(NS) (FIG. 6) được tạo cấu trúc nhằm mục đích này. pXY300-*orf18ifd* cho phép chọn thể đột biến xóa đoạn không làm lệch khung dịch mã *orf18* bằng thiostrepton trong khi pKS-*orf18ifd*-T-AmR(NS) sử dụng apramycin để chọn các thể đột biến xóa đoạn không làm lệch khung dịch mã. Mặc dù các chủng đột biến *S. fungicidicus* kiểu đại dễ dàng chọn được bằng cách sử dụng gen đánh dấu kháng thiostrepton, nhưng đã gặp phải các khó khăn do sử dụng gen đánh dấu kháng này ở chủng BM38-2 (ATCC PTA-122342). Do đó, hai plasmit, pXY300-*orf18ifd* và pKS-*orf18ifd*-T-AmR(NS), được tạo cấu trúc nhằm mục đích tương tự.

Do nhiều phương án có thể có mà các quy tắc theo sáng chế có thể được áp dụng, cần hiểu rằng các phương án được minh họa chỉ là các ví dụ ưu tiên theo sáng chế và không làm giới hạn phạm vi của sáng chế. Hơn nữa, hạn phạm vi của sáng chế được xác định bởi yêu cầu bảo hộ kèm theo. Do đó, toàn bộ sáng chế đều nằm trong phạm vi và tinh thần của yêu cầu bảo hộ này.

YÊU CẦU BẢO HỘ

1. Chủng *Streptomyces fungicidicus* tái tổ hợp để tăng cường sản xuất enduracidin, bao gồm một hoặc nhiều gen biến đổi được chọn từ nhóm bao gồm khung đọc mở-24 (*orf24*) tăng mà mã hóa trình tự axit amin SEQ ID NO: 26, khung đọc mở-18 (*orf18*) giảm, và cả *orf24* tăng và *orf18* giảm; trong đó sự sản xuất tăng cường enduracidin thu được bằng chủng *Streptomyces fungicidicus* tái tổ hợp được so sánh với sự sản xuất tăng cường thu được bằng chủng *Streptomyces fungicidicus* kiểu dại đối chứng.
2. Chủng tái tổ hợp theo điểm 1, trong đó *orf18* giảm được giảm đi bởi vì nó đã bị bất hoạt.
3. Chủng tái tổ hợp theo điểm 2, trong đó *orf18* giảm đã bị bất hoạt bằng quy trình được chọn từ nhóm bao gồm xóa đoạn không làm lệch khung dịch mã, đột biến dịch khung, đột biến điểm, và sự kết hợp bất kỳ của chúng.
4. Chủng tái tổ hợp theo điểm 3, trong đó *orf18* giảm đã bị bất hoạt bằng cách xóa đoạn không làm lệch khung dịch mã.
5. Chủng tái tổ hợp theo điểm 4, trong đó sự xóa đoạn không làm lệch khung dịch mã là xóa đoạn các nucleotit từ 5 đến 660 của *orf18* (SEQ ID NO: 37).
6. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó *orf24* tăng được liên kết theo cách có điều khiển với gen khởi đầu khác loại.
7. Chủng tái tổ hợp theo điểm 6, trong đó gen khởi đầu khác loại là gen khởi đầu cơ định.
8. Chủng tái tổ hợp theo điểm 7, trong đó gen khởi đầu cơ định là *ermE**p.
9. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 8, trong đó *orf24* tăng được tăng lên bởi vì nó đã được biểu hiện quá mức.
10. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 9, trong đó *Orf24* chứa trình tự axit nucleic SEQ ID NO: 38.

11. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 10, trong đó *Orf18* chứa trình tự axit nucleic SEQ ID NO: 37.
12. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 11, trong đó *Streptomyces fungicidicus* kiểu đại là *Streptomyces fungicidicus* ATCC 21013.
13. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 12, trong đó việc sản xuất enduracidin bằng chủng tái tổ hợp tăng ít nhất là gấp 1,2 lần so với việc sản xuất enduracidin bằng chủng *Streptomyces fungicidicus* kiểu đại đối chứng.
14. Chủng tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 13, trong đó việc sản xuất enduracidin bằng chủng tái tổ hợp tăng gấp 1,2 đến 4,6 lần so với việc sản xuất enduracidin bằng chủng *Streptomyces fungicidicus* kiểu đại đối chứng.
15. Chủng tái tổ hợp theo điểm 1 là BM38-2.24/16 được chỉ định nộp lưu ATCC số PTA-124006.
16. Chủng tái tổ hợp theo điểm 1 là BM38-2.18pfrd-AmR được chỉ định nộp lưu ATCC số PTA-124007.
17. Phương pháp sản xuất enduracidin, trong đó phương pháp này bao gồm bước nuôi cấy chủng *Streptomyces fungicidicus* tái tổ hợp theo điểm bất kỳ trong số các điểm từ 1 đến 16 trong môi trường nuôi cấy trong các điều kiện sản xuất enduracidin.
18. Phương pháp theo điểm 17, trong đó phương pháp này còn bao gồm bước phân lập enduracidin từ môi trường nuôi cấy.

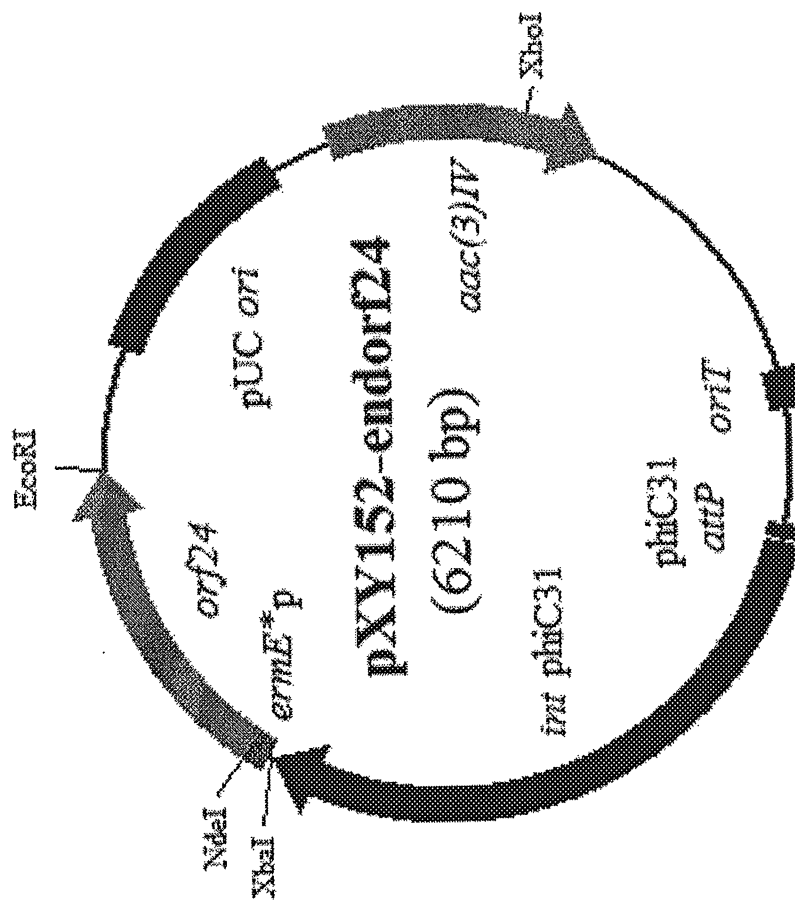


Fig. 2

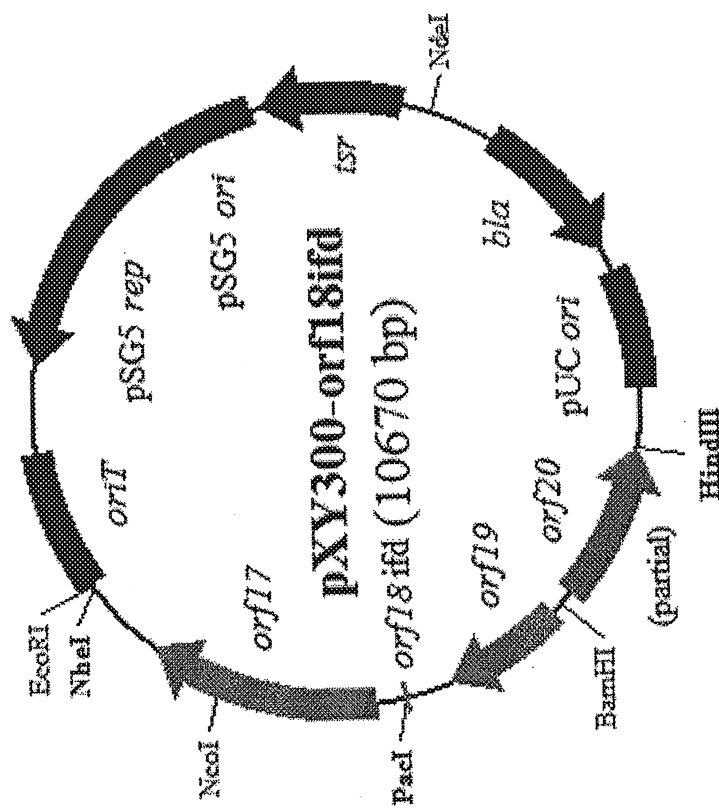


Fig. 3

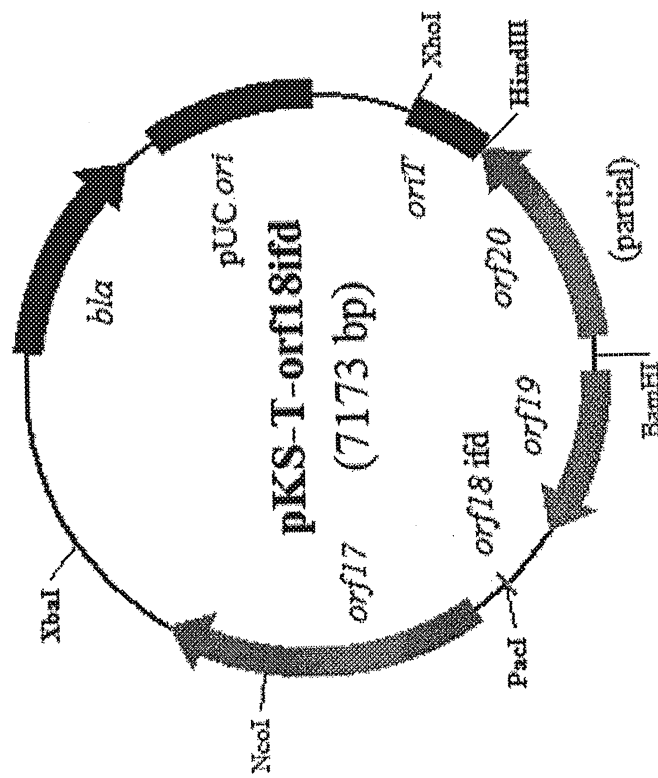


Fig. 4

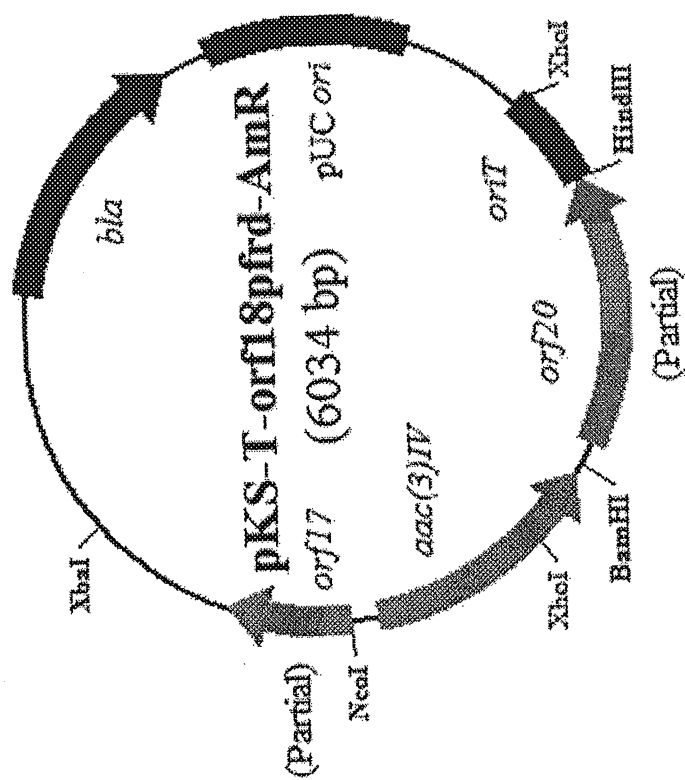


Fig. 5

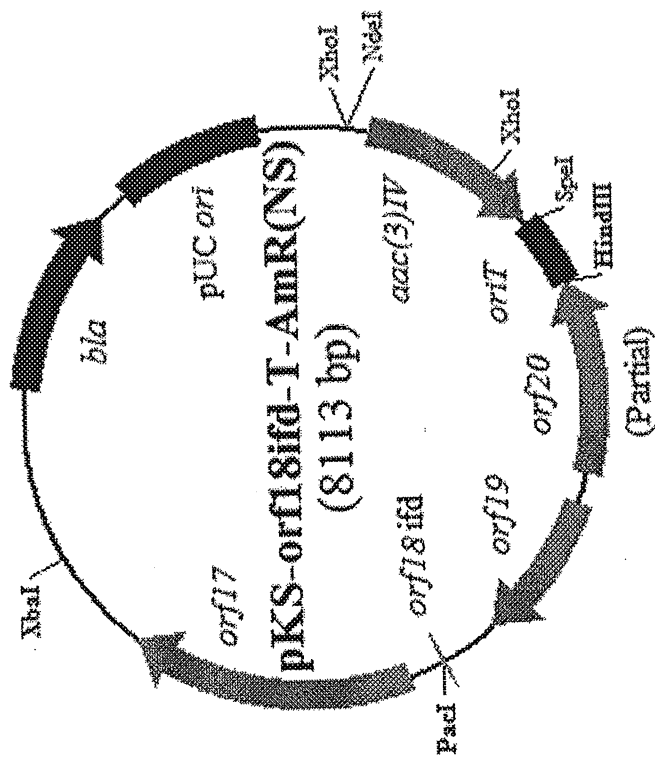


Fig. 6

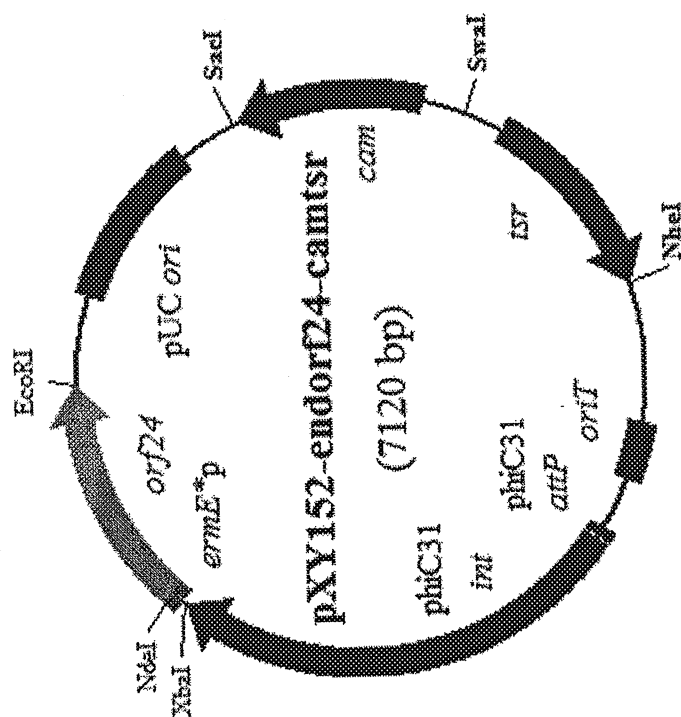


Fig. 7

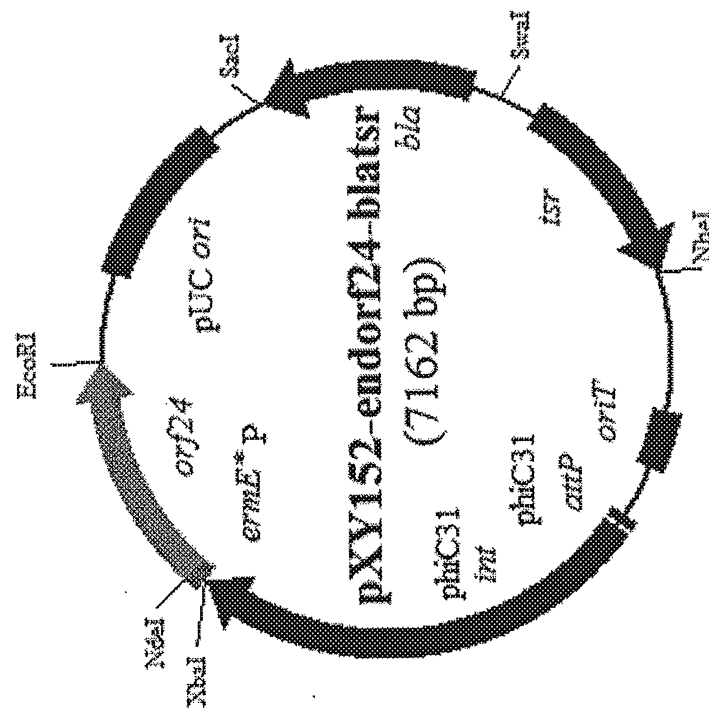


Fig. 8

Thang điểm	Mong muốn	Phương pháp	Độ đồng nhất	Dương tính	Khoảng trống
281 bits(720)	3e-97	Điều chỉnh chất nền tạo thành	167/311(54%)	212/311(68%)	8/311(2%)
StrR 16	IELSRLLSSASPTSGEDPEHVEITLLSAEGELPPILVHRPTMQVLDGLHRLKVARVRGDT +E+S LS+ SPR GE PEHVE L +A+ LPFI+VHR T +V+DG+HRL+ A + G T				75
Orf24 1	VEISSLSSTDGSPRIDGESPEHVEMLAAADTALPPIMVHRRRTGRVIDGMHRLRAAMLITGRT				60
StrR 76	KILARLVDATESDAFVLAVAEANIRHGLPLSLADRKRAAVQIIGTHPQWSDRRVASATGIS I R D TE DAFVLAV++NI HGLPLS ADK+FAA +I+ THP+WSDR +AS G S				135
Orf24 61	TIAVRFDTGTEEDAFVLAVKSNIAHGLPLSADRRPRAAGRIMATHPRWSDRMIAASVVGTS				120
StrR 136	AGTVADLRRRAGEDGT--EARIGRDCRVRPDSGSEPRRLAAELIRSDPGLSLRQVAKQVG A TVA++RR AG G RIGRDCRVRP D SE RRIA ++I DPGLSLRQVA+ G				193
Orf24 121	ARTVAEIRRDACAAGACEPTRICKDGRVVPDVSEGRRLAHDMIVRDEGLSLRQVABAAG				180
StrR 194	ISPETVRDVGRRLERGESPTPDGTERLPAKPHFLRLSEPDFGRA-----VDQDRLLALIER ISPETVRDVR R+ RGE P P R + R +EP G+A + + + + R				248
Orf24 181	ISPETVRDVHRMLRGEDPVPAEPRTLVERGADRRRAEP-AGKAAAPCGTEPPPPAVVMKR				239
StrR 249	LKSDPALRLNEVGRILLRMLTMHSMMDGQEWERILQVPPPHLHGVIAGFARDHARVWAEEFA L++DPALRLNE GR LLR+L +H++ ++W RI++ VPPH +A AR A W+E A				308
Orf24 240	LRADPALRLNENGRDLLRLLDIHTVLEDDWNRIIESVPPHRIETVAQLARSCADKWSEIFA				299
StrR 309	DHLESRAATELA 319 +ES A+ LA				
Orf24 300	SRIESNASHLA 310				

Fig. 9

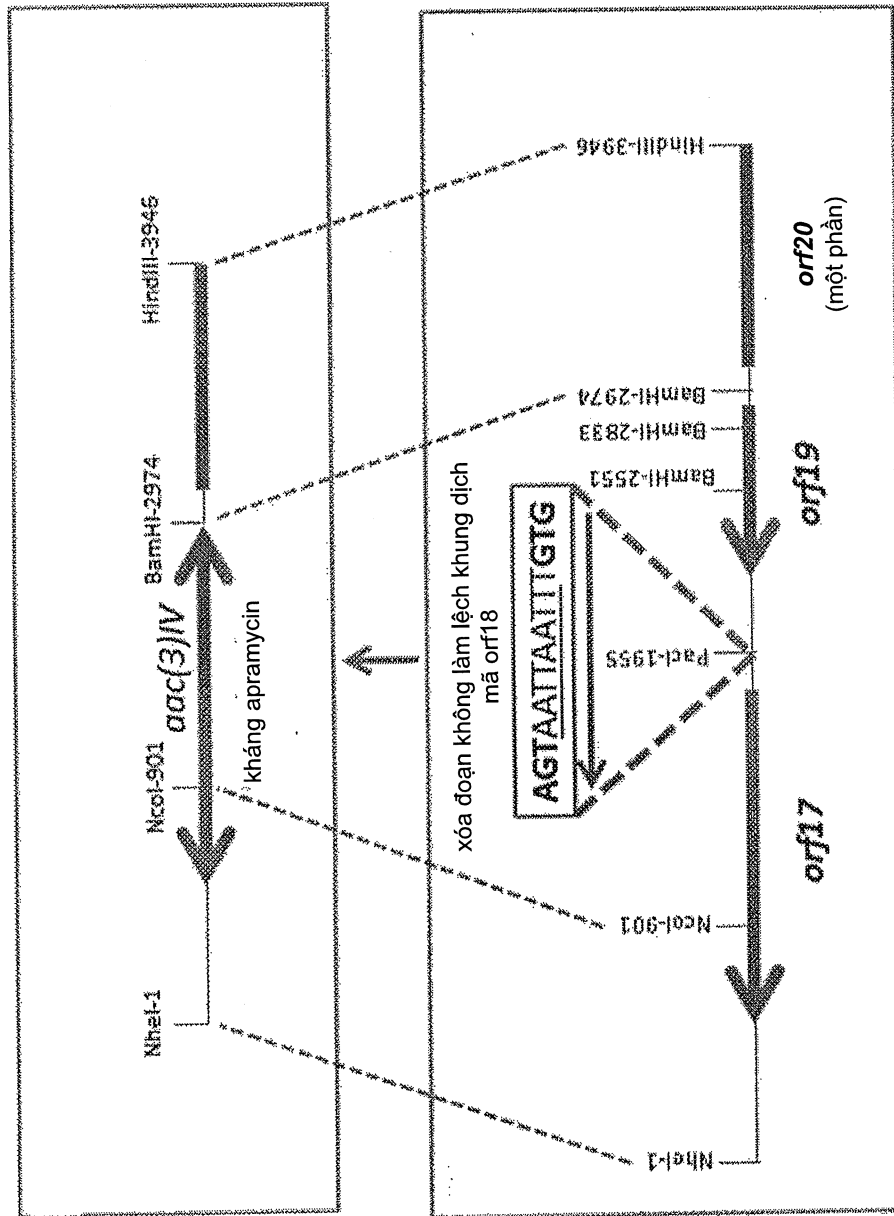


Fig. 10A

Fig. 10B

```

Bbr      MDPTR-----VDIFALPA-----VEIELSRLSSASSPRT
Kast     MAETVRADSPKSSYRNVPAAEVQGGSLVSGQRTTRIAISSLLAADSPRS
NovG     MTNSG-----DEBIT-P-----ASLKATRKGERVSIGSLPPSELVR
SgcR1    MKSDS-----AQRAVER-----SRVVRIDELIPADSPRL
StrR     MDPTR-----VDIFALPA-----VEIELSRLSSASSPRT
Tei15*   MTPDE-----EALNRQPI-----MEMEISSLSLGGSPRL
Orf24    -----VEISGLSTDGSPRI
          : : * ..

Bbr      SGEDPEHVETLLSAGELPPIVHRPTMQVLDGLHRLKVAARVGDTKILA
Kast     AGENAEHIRLLADSGARLPPIVVRSTMRVIDGMHRLKAAALRGTEIEV
NovG     SGESTEHIRVLAETDEDLPPIVVHRGTRRVVDGMHRLWAARFRGDESIEV
SgcR1    NGIDRSHVQRLATVYASLPVVLVHRPTMRVVVDGMHRIGAARLKGLDTEV
StrR     SGEDPEHVETLLSAGELPPIVHRPTMQVLDGLHRLKVAARVGDTKILA
Tei15*   AGGDEVHLEAMVAAQGELPPIVVRPTMRVIDGSHRIQAALRRGETTIAG
Orf24    DGESPEHVEMLAADALPPIMVHRRTGRVIDGMHRLKAAALRGTTIIV
          * . * : . :      * * : * * : * * : * * : * :

Bbr      RLVDATESDAFVLAVEANIRHGLPLSLADRKRAAVQIIGTHPQWSDRVA
Kast     RFFDGAEDSFLAVRSNIAHGLELSQEERAAAQRIIRSHAQWSNQAIG
NovG     VFVDGSPADVFVLAVELNRAHGLPLTLDERKSAAQINDSRPHWSDRRIA
SgcR1    TFFEGAEQVFLRSVAANITNGLPLSVADRKTAAARILASHPTLSDRAVA
StrR     RLVDATESDAFVLAVEANIRHGLPLSLADRKRAAVQIIGTHPQWSDRVA

Tei15*   RFFDGSDDAEFVMSVWLVNVSGLPLALADRKRAAERIAVSHHPQWSDRVA
Orf24    RFFDGTEDDAFVLAVKSNIAHGLPLSAADRKRAAGRIMATHPWSDRMIA
          : : : : : * : : * * : * * : * * : * * : * :

Bbr      SATGISAGTVADLRRA-GEDG-TEARIGRDGRVRPSDGSERRRLAELI
Kast     EVTGLDAKTIAALRRDAKDV-PQLDARIGRDGRVRVVDGQGRRLAGELM
NovG     RTTGLAASVSLRSSST-AG-TVGRRTGQDGRSRPNQDGTGQRAAALL
SgcR1    AHVGLDAKTIVAGVRTCSAAGSPLLNMRTGADGRVHFLDRTAERLHAAALL
StrR     SATGISAGTVADLRRA-GEDG-TEARIGRDGRVRPSDGSERRRLAELI
Tei15*   AVTGISPSTVADIRRRVAGTSAPEASKIGQDGRVRPLDCSAGRLLAGRLM
Orf24    SVVGTSARTVAEIRRDACAAGACEPTRIGRDGRVRVVDVSEGRRLAHDMI
          . * . * : * : *      * * * * : * * : * * : :

Bbr      RSDPGLSLRQVAKQVGLSPETVRDVRGRLERGESEPTDGTTRLP-AKPH-
Kast     AEQPDAPLRKIAHAAGVSLGTASDVRRRIENGQDPVPAGRQKAD-PQPF-
NovG     ARNPNASLREVTRAAGISVGTASDVRRRLRGEPAITARQQAVMKLRPA-
SgcR1    TQDPGLPLRSVVEQTGLSLGTAHDVRRRLRGEDEPVPQNRQSAM-LEPGL-
StrR     RSDPGLSLRQVAKQVGLSPETVRDVRGRLERGESEPTDGTTRLP-AKPH-
Tei15*   AENPALSRLQVAKAAAISETARDVRNRLSGAELVFNRRPRDA-A-PV-
Orf24    VRDPGLSLRQVAKAAISETARDVRNRLSGAELVFNRRPRDL-VERG-
          : * . * : . . . : * * . * * * : * .

Motip cấu trúc dạng xoắn-ngoặt-xoắn

Bbr      -PL--RLSEP-----DFGRA-----V--DQDLALLERL
Kast     -AR--YAASED----R-SGTTA-----PR--TG--EQNRKVLQKL
NovG     -AA--QRSGP-----FVGPAARFVKVPPAVAGRPVSPRSAPLEALRKL
SgcR1    APQKKATAKP----FVGPAARFVKVPPAVAGRPVSPRSAPLEALRKL
StrR     -PL--RLSEP-----DFGRA-----V--DQDLALLERL
Tei15*   -GV--KGGRRDRPLNLRSGD-----RP--EP--VFOHAVVINRL
Orf24    -AD--RRAEP----A-GKAAA-----PC--GT--EPPPAVVMKRL
          : : *

Bbr      DHARVWAEFADHLESKATEL-----AAG
Kast     GCSGVWQDFAAQLERRG--R-----ASA
NovG     ECAAAWQHLADQLADRD-----TA
SgcR1    HCSDAWHRFEEEMVRRHSAADGSGRLTQPTRR
StrR     DHARVWAEFADHLESKATEL-----AAG
Tei15*   QFADLWADFASRVGPPEE--R-----MAS
Orf24    SCADKWSEIASRIESNAS-H-----LAG
          : * : * : .

```

Fig. 11

```

SCO1745/Abra2 MT-----IRLLIVDDQELIRTGFRLFLQTQNDLEVVE
SCO3226/Abra2 M-----IRVLLADDETIIRAGVRSILTTEPGIEVVAE
SCO3818 MRE-----DGKIRVFLLDDHEVVRRGVHDLDSGEADIEVVE
Orf18 VSVLLEQPASLVAYRPNKPTAMVVV-ADPRVRSTVTRHLWA-LGVRDVIE

: :. : !* . * :. * *

SCO1745/Abra2 ADDGHGALAQAALRPDVVLMDIRMPMDGVEATSKLTASDSPPRVLILT
SCO3226/Abra2 ASDGREAVELARKHRRPDVALLDIRMPMDGLTAAGEMRTNPDVAVVLT
SCO3818 AGTAAEAQARVTATRPDVAVLVDVRLPDGSGVEVCRDIRSRDESVRCLMLT
Orf18 ASSVAEARPRIGNPR-DICVAEVHLPDGSGLTLLSETRAAGW-PNGLALS
* . * *: : : : * .*: : *:

SCO1745/Abra2 TYDLDEYVFGALRAGASGFLKKDASDRRLLEAIRVVHAGEALLSPSITRR
SCO3226/Abra2 TFGEEDRYIERALDQGVAGFLLKASPPRDLISGVRAVASGGSCLSPLVARR
SCO3818 SFADDSALFDALMAGASGYVLKDIRGAELLGAVREVAAGKSLDPAATAR
Orf18 AADDIGAVRNALAGGVKGYVVTGTRTNLGL-PTRPAA---PI-GAAAR
: : * : * .*: : : * : : : *

SCO1745/Abra2 LIEDYATRAAPV-RFR--EAVLAGLTPFREREILLVARGLSNPPIAARLV
SCO3226/Abra2 LMTELK--RAPSPRSEVSGERTLLTKREQEVLGMLGAGLSNAETIAQRLN
SCO3818 VLERLE--GGGA---RP-DDRLARLITEQERRILELIGEGLTNRAIGERLH
Orf18 LHRRPP--GAPS---HPG-G-YRELSGREVEVLRLVAEGQSNKAIGVSMG
: . * : * . : : * : * :

SCO1745/Abra2 VTEATVKSHVGSMAFLHLRDRAQAVVFAYENAIIVLP-GGTG
SCO3226/Abra2 LVEGTIKTYVSAIFTQLEVRNRVQAAIIAYEAGLVKDADLNR
SCO3818 LAEKTIKNYVSSLLGKLGMRQRRSQAAAFVAR---LE-AENR
Orf18 LSALTVKSHLARIARKLGTGDRAGMVAVALRTGII-----H
: * : * : : : * : . .

```

Fig. 12

DANH MỤC TRÌNH TỰ

<110> Oregon State University
Zabriskie, T. Mark
Yin, Xihou

<120> CHỦNG STREPTOMYCES FUNGICIDICUS TÁI TỔ HỢP ĐỂ TĂNG CƯỜNG SẢN XUẤT
ENDURACIDIN VÀ PHƯƠNG PHÁP SẢN XUẤT ENDURACIDIN

<130> 127789-236560-P005PCT

<150> 62/430,838

<151> 2016-12-06

<150> 62/479,087

<151> 2017-03-30

<160> 40

<170> PatentIn version 3.5

<210> 1

<211> 31

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Đoạn mồi oligonucleotit tổng hợp

<400> 1

ccaccacata tggaaataag ttcgctctcc a

31

<210> 2

<211> 32

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Đoạn mồi oligonucleotit tổng hợp

<400> 2

gtgtgtgaat tcctcgttca cccggccaga tg

32

<210> 3

<211> 6210

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Plasmid pXY152-endorf24 tổng hợp

<400> 3
 gaattcgtaa tcatgtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca 60
 cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaata agtgagctaa 120
 ctcacattaa ttgcgttgcg ctactgccc gctttccagt cgggaaacct gtcgtgccag 180
 ctgcattaat gaatcggcca acgcgcgggg agaggcgggt tgcgtattgg gcgctcttcc 240
 gcttcctcgc tctactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct 300
 cactcaaagg cggtaatagc gttatccaca gaatcagggg ataacgcagg aaagaacatg 360
 tgagcaaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc 420
 cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga 480
 aaccgcagag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct 540
 cctgtttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg 600
 gcgcttttctc atagctcacg ctgtaggtat ctgagttcgg tgtaggtcgt tcgctccaag 660
 ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggttaactat 720
 cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac 780
 aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac 840
 tacggctaca ctagaagaac agtatttggt atctgcgctc tgctgaagcc agttacctc 900
 ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt 960
 tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc 1020
 ttttctacgg ggtctgacgc tcagtgaac gaaaactcac gttaagggat tttggtcatg 1080
 agattatcaa aaaggatctt cacctagatc cttttggttc atgtgcagct ccatcagcaa 1140
 aaggggatga taagtttatc accaccgact atttgcaaca gtgccgttga tcgtgctatg 1200
 atcgactgat gtcacagcg gtggagtga atgtcgtga atacgaatgg cgaaaagccg 1260
 agctcatcgg tcagcttctc aaccttgggg ttacccccgg cggtgtgctg ctggtccaca 1320
 gtccttccg tagcgtccgg cccctcgaag atgggccact tggactgatc gaggccctgc 1380
 gtgctgcgct ggggtccggga gggacgctcg tcatgccctc gtggtcaggt ctggacgacg 1440
 agccgttcga tcctgccacg tcgcccgtta caccggacct tggagttgtc tctgacacat 1500
 tctggcgcct gccaaatgta aagcgcagcg cccatccatt tgcctttgcg gcagcggggc 1560

cacaggcaga gcagatcatc tctgatccat tgccccctgcc acctcactcg cctgcaagcc	1620
cggtcgcgccg tgtccatgaa ctcgatgggc aggtacttct cctcggcgtg ggacacgatg	1680
ccaacacgac gctgcatctt gccgagttga tggcaaaggt tccctatggg gtgccgagac	1740
actgcaccat tcttcaggat ggcaagttgg tacgcgtcga ttatctcgag aatgaccact	1800
gctgtgagcg ctttgccttg gcggacaggt ggctcaagga gaagagcctt cagaaggaag	1860
gtccagtcgg tcatgccttt gctcggttga tccgctcccg cgacattgtg gcgacagccc	1920
tgggtcaact gggccgagat ccgttgatct tcctgcatcc gccagaggcg ggatgcgaag	1980
aatgcgatgc cgctcgccag tcgattggct gagctcatga gcggagaacg agatgacgtt	2040
ggagggggcaa ggtcgcgctg attgctgggg caacacgtgg agcggatcgg ggattgtctt	2100
tcttcagctc gctgatgata tgctgacgct caatgccgtt tggcctccga ctaacgaaaa	2160
tcccgcatth ggacggctga tccgattggc acggcggacg gcgaatggcg gagcagacgc	2220
tcgtccgggg gcaatgagat atgaaaaagc ctgaactcac cgcgacgtat cgggccctgg	2280
ccagctagct agagtcgacc tgcagggtccc cggggatcgg tcttgccctg ctcgtcgggtg	2340
atgtacttca ccagctccgc gaagtcgctc ttcttgatgg agcgcacatggg gacgtgcttg	2400
gcaatcacgc gcaccccccg gccgttttag cggctaataaa agtcatggct ctgccctcgg	2460
gcggaccacg cccatcatga ccttgccaag ctcgtcctgc ttctcttcga tcttcgccag	2520
cagggcgagg atcgtggcat caccgaaccg cgccgtgcgc gggtcgtcgg tgagccagag	2580
tttcagcagg ccgcccaggc ggcccaggtc gccattgatg cgggccagct cgcgacgtg	2640
ctcatagtcc acgacgcccg tgattttgta gccctggccg acggccagca ggtaggccga	2700
caggctcatg ccggccgccg ccgccttttc ctcaatcgct cttcgttcgt ctggaaggca	2760
gtacaccttg ataggtgggc tgcccttcct ggttggcttg gtttcatcag ccatccgctt	2820
gccctcatct gttacgccgg cggtagccgg ccagcctcgc agagcaggat tcccgttgag	2880
caccgccagg tgcgaataag ggacagtga gaaggaacac ccgctcgcgg gtgggcctac	2940
ttcacctatc ctgcccggct gacgccgttg gatacaccaa ggaaagtcta cacgaaccct	3000
ttggcaaaat cctgtatatc gtgcgaaaaa ggatggatat accgaaaaaa tcgctataat	3060
gaccccggaag cagggttatg cagcggaataa gatccgtcga cctgcaggca tgcaagctct	3120
agcgattcca gacgtcccga aggcgtggcg cggcttcccc gtgccggagc aatcgccctg	3180

ggtgggttac acgacgcccc tctatggccc gtactgacgg acacaccgaa gccccggcgg 3240
 caaccctcag cggatgcccc ggggcttcac gttttcccag gtcagaagcg gttttcggga 3300
 gtagtgcccc aactggggta acctttgagt tctctcagtt gggggcgtag ggtcgccgac 3360
 atgacacaag gggttgtgac cggggtggac acgtacgcgg gtgcttacga ccgtcagtcg 3420
 cgcgagcgcg agaactcgag cgcagcaagc ccagcgacac agcgtagcgc caacgaagac 3480
 aaggcggccg accttcagcg cgaagtcgag cgcgacgggg gccggttcag gttcgtcggg 3540
 catttcagcg aagcgccggg cacgtcggcg ttcgggacgg cggagcgccc ggagttcgaa 3600
 cgcacacctga acgaatgccg cgccggggcg ctcaacatga tcattgtcta tgacgtgtcg 3660
 cgcttctcgc gcctgaaggt catggacgcg attccgattg tctcggaatt gtcgcccctg 3720
 ggcgtgacga ttgtttccac tcaggaaggc gtcttccggc agggaaacgt catggacctg 3780
 attcacctga ttatgaggct cgacgcgtcg cacaagaat cttcgctgaa gtcggcgaag 3840
 attctcgaca cgaagaacct tcagcgcgaa ttgggcgggt acgtcggcgg gaaggcgctt 3900
 tacggcttcg agcttgtttc ggagacgaag gagatcacgc gcaacggccg aatgggtcaat 3960
 gtcgtcatca acaagcttgc gactcgacc actccccta ccggaccctt cgagttcgag 4020
 cccgacgtaa tccggtggtg gtggcgtgag atcaagacgc acaaacacct tcccttcaag 4080
 ccgggcagtc aagccgccat tcacccgggc agcatcacgg ggctttgtaa gcgcatggac 4140
 gctgacgccg tgccgaccg gggcgagacg attgggaaga agaccgcttc aagcgcttg 4200
 gacccggcaa ccgttatgcg aatccttcgg gacccgcgta ttgcgggctt cgccgctgag 4260
 gtgatctaca agaagaagcc ggacggcacg ccgaccacga agattgaggg ttaccgcatt 4320
 cagcgcgacc cgatcacgct ccggccggtc gagcttgatt gcggaccgat catcgagccc 4380
 gctgagtggg atgagcttca ggcgtggtg gacggcaggg ggcgcggcaa ggggctttcc 4440
 cgggggcaag ccattctgtc cgccatggac aagctgtact gcgagtgtgg cgccgtcatg 4500
 acttcgaagc gcggggaaga atcgatcaag gactcttacc gctgccgtcg ccggaagggtg 4560
 gtcgaccctg ccgcacctgg gcagcacgaa ggcacgtgca acgtcagcat ggcggcactc 4620
 gacaagttcg ttgcggaacg catcttcaac aagatcaggc acgccgaagg cgacgaagag 4680
 acgttggcgc ttctgtggga agccgccga cgcttcggca agctcactga ggcgcctgag 4740

aagagcggcg aacgggcgaa cttgtttgcg gagcgcgccg acgccctgaa cgcccttgaa	4800
gagctgtacg aagaccgcgc ggcaggcgcg tacgacggac ccgttggcag gaagcacttc	4860
cggaagcaac aggcagcgct gacgctccgg cagcaagggg cggaagagcg gcttgccgaa	4920
cttgaagccg ccgaagcccc gaagcttccc cttgaccaat gggtccccga agacgccgac	4980
gctgacccga ccggccctaa gtcgtggtgg gggcgcgcgt cagtagacga caagcgcgtg	5040
ttcgtcgggc tcttcgtaga caagatcggt gtcacgaagt cgactacggg cagggggcag	5100
ggaacgcccc tcgagaagcg cgcttcgata acgtgggcga agccgccgac cgacgacgac	5160
gaagacgacg cccaggacgg cacggaagac gtagcggcgt aggtgagtga agatctagac	5220
gtggcaccgc gatgctgttg tgggcacaaat cgtgccggtt ggtaggatcc acatatggaa	5280
ataagttcgc tctccaccga cggctccccg cggatcgacg gggagagtcc cgagcacgtg	5340
gaaatgctgg ccgccgccga caccgcgctt ccaccgatca tggcgcaccg ccgcaccggg	5400
cgggtcatcg acggcatgca ccggctgcgc gccgcgatgc tgacgggccg tacgacgac	5460
gcggtgaggt tcttcgacgg caccgaggag gacgccttcg tcctcgccgt gaagtcgaac	5520
atcgcgcacg gactgccgct gtccgccgcc gaccgccggc gggccgccgg gcgcatcatg	5580
gccacccatc cccggtggtc ggaccggatg atgcctcgg tggcgcgcac ctccgccagg	5640
acggtcgccg agatccgccg cgacgccggc gccgccgggg cgggggagcc caccgcac	5700
ggccgggacg gcagggtacg gcccgtcgac gtgagcgagg gccgcagact ggcccacgac	5760
atgatcgctc gcgaccggg cctgtcgtg cgccaggctc cccgcgccgc cgggatctcg	5820
ccggagaccg tcagggacgt cagacaccgg atgctccgc gtgaggacct ggtgcccgcg	5880
ccgcggccgc ggaccctggt ggagcgcggc gcggaccgcc gggcggagcc ggccgggaag	5940
gccgccgcgc cgtgcgggac ggagccgccg cccgccgtcg tgatgaagcg gctgagggcc	6000
gatccggcgc tgcgtctcaa cgagaacgga cgcgacctgc tcgggcttct ggatatccac	6060
acggtccggc tggaggactg gaaccgcatt atcgaaagcg tgccgccgca ccgtctggag	6120
acggtggcgc agctggcacg ctctgcgcc gacaaatggt ccgagatcgc gtcacgcatc	6180
gaaagcaacg catcacatct ggccgggtga	6210

<210> 4

<211> 38

<212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

 <400> 4
 ttattgaagc ttgccggggc cgacgcggcg ggcggcct 38

 <210> 5
 <211> 34
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

 <400> 5
 gttgttttaa ttaaacacca ggcctcctgg ggtg 34

 <210> 6
 <211> 38
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

 <400> 6
 tttatattaa ttaatgaccc ttccgtcccg ccccgat 38

 <210> 7
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

 <400> 7
 tttggtgcta gctggtcgtg gcgctgttcc 30

 <210> 8
 <211> 10670
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> Plasmit pXY300-orf18ifd tổng hợp

<400> 8

ccgcggctcc tgcccgccga acgcgtcgtc gtcgacggcc tgggtgctcat cgacgagcac	60
ccggagccag gtgaaaagcg ccggcggacg ctcggactgg gcgcgggatt ccagcagtaa	120
cccaggtccg ccggccacct cacggcaggc agaccctcgg ctttcgcgcc gggaccgcca	180
tgagaccgcc acccggtatgt ccgggggtggc ggtctcatgg cggtcctca gcggccctac	240
gcgaccgctg tgtcgacgcg gaggcagtct ccggggagcc ggccccaggg ctggagtacc	300
ggggcgagct tgcccttgca gcggcggcac acgggcgacg cggcgccggg cgggggagtc	360
ttgacgtcga cggtgaccgg cttcggcttc aggcgcttcc gaaggctgat cgtcgggcgg	420
atgcctcct tcttcggctg acgcactcgg tcgagcgaag cggccagctc ctccgtgcgg	480
gcctcgttgg ccttccggcg cgcctcgcgg aaagcagctt cctcctcgga catgacctcg	540
aacctcatct ggtcagcgtc aaggtcgccc ggcgcgggccg gcgcttccgg ggcgggcggg	600
tccaggacgt ccttgcccca caccaggccc caggactcga cgagccgccg gacgcccggt	660
aggccgtacg tctcggcgac cttgatgagg tcgaggtcga ggcgacgtcc ggcgacgcgg	720
gcgatgtatc ggtaccagat gtaggccggg atgaccgcga tggcgaccag gccctcgggtg	780
tcgtcggtaga tctcctcctc ggtgcggacg tcctgctgga tgccgagttc cttgatcagc	840
cggttcaggt tctgcgaccg gtagtgcttg cggacctgga agacgccgaa ctcgcgctcg	900
cgggtacttct cgacgaacgg gccgggcccga cgaagccgct gcagctcggc ggccgccgcg	960
tcgcccaggt cgagcgggtcc catgcggtcg tcgccgcgac cggccttgaa gttctgtccg	1020
gccagctcca ggccgatctt ggcgacgccg cccttggtct tgtcgccgtc cttgtagagg	1080
tagcgggcct gcttgcccgc atcgccgtca gcggcgtccg cgccgttgag tgggcgcacg	1140
tcggtgccgt ggcccttgcc ctacaggag caaccgggcc ggtcgcacgt ctcgctgacg	1200
gtgtagccgc ccgcggattc gacccggcg gccaggctc cggcgagtgc gtcgcggaac	1260
gcctgggcgt ccgggcccag cacctcgcgg gtgaccaga gcgtgtgcca gtgcaggtgc	1320
cagccggagc cccagccgaa ggtgtcctcg aaggcccgt cgtagccgat gatcccgaag	1380
tcgtcgcgca tcgtgcgcac gcggcggccg gacgagccgt acgcgccctt ccagccgtcg	1440
tgcaagaccg cgaccaggcc gtgccgcatc cccttgcgga cggtgccgaa cgccatgcgc	1500
tcgaagtggc gcaacgtgtt cgtgccaaagg tgcagcccgt acccggcgtc cgcgagaccg	1560

tcggcggcga gctgcacgtt cgagccccgt acggccagga tgcggctcat gcaccacggg	1620
caggtgtgga cgttggttga gcggcacgtg ttgccccacg tcgcctcgcc cggcttccac	1680
atcagctcgc ggtccggcag tgagccgggt cccgcagccc ttgaacgcct cgttcagcga	1740
caccgtctgg tgccgggtccc gccgggcgaa ccgctcgtcg cggcgggtcc tcccggccgc	1800
tgtcgcggca ccctcgtttg gggtagaacc cgttccagtt acagcgctct gacctgcagt	1860
ggacggagat tttcccttac tactaaagcc cgcgtccgga ttacccgctg tagtcgtgct	1920
tgctacgctg cgtgactggg ccgcaatgag acgctttgcg cgctttcggc aggcgtccga	1980
gcagtagatt ttggggcgct tcccggggat gtggacgatc ggggtgccgc agtggcactt	2040
cgggtccggcg gggcgcggtg gtgtcgacgc gctgttctct cgtacgcctc gtcacagagc	2100
aaacgtcctc actcggcatg ctgcgccggt tcggggcgcg cgagccggga ggccaatccc	2160
gggctcgtgc catttctggg tcctgttgat catcactgac gaatcgaggt cgaggaaccg	2220
agcgtccgag gaacagaggc gcttatcggt tggccgcgag attcctgtcg atcctctcgt	2280
gcagcgcgat tccgagggaa acggaaacgt tgagagactc ggtctggctc atcatgggga	2340
tggaaaccga ggcggaagac gcctcctcga acaggctcga aggcccaccc ttttcgctgc	2400
cgaacagcaa ggccagccga tccggattgt ccccgagttc cttcacggaa atgtcgccat	2460
ccgccttgag cgtcatcagc tgcataccgc tgtcccgaat gaaggcgatg gcctcctcgc	2520
gaccggagag aacgacggga agggagaaga cgtaacctcg gctggccctt tggagacgcc	2580
ggtccgcgat gctggtgatg tctactgtcga ccaggatgat ccccgacgct ccgagcgcga	2640
gcgacgtcgc tactatcgcg ccgatgttcc cgacgatctt caccctcgtc agaacgacga	2700
cgtccccacg ccggctcgcg atatcgccga acctggccgg gcgaggggacg cgggcgatgc	2760
cgaatgtctt ggccttccgc tcccccttga acaactgggt gacgatcgag gagtcgatga	2820
ggcggaccgg tatgttctgc cgcccgaca gatccagcaa ctcatgatga aaaggactgc	2880
tgtcgtgcc gtagacctcg atgaactcca cccggccgc gatgctgtgc atgaggggct	2940
cgacgtcctc gatcaacgtt gtctttatgt tggatcgca cggttggtg acatcgatga	3000
tccgctgcac cgcgggatcg gacggatttg cgatggtgtc caactcagtc atggctgtcc	3060
taccggctgc tgtgttcagt gacgcgattc ctgggggtgtg acaccctacg cgacgatggc	3120

ggatggctgc cctgaccggc aatcaccaac gcaaggggaa gtcgtcgctc tctggcaaag	3180
ctccccgctc ttccccgtcc gggaccgcg cggtcgatcc ccgcatatgg tgcactctca	3240
gtacaatctg ctctgatgcc gcatagttaa gccagccccg acacccgcca acacccgctg	3300
acgcgccctg acgggcttgt ctgctcccg catccgctta cagacaagct gtgaccgtct	3360
ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg agacgaaagg	3420
gcctcgtgat acgcctatct ttataggcta atgtcatgat aataatgggt tcttagacgt	3480
cagggtggcac ttttcgggga aatgtgcgcg gaacccttat ttgtttatct ttctaaatac	3540
attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa taatattgaa	3600
aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt tttgcggcat	3660
tttgccttcc tgtttttgct caccagaaaa cgctggtgaa agtaaaagat gctgaagatc	3720
agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag atccttgaga	3780
gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg ctatgtggcg	3840
cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata cactattctc	3900
agaatgactt gggtgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag	3960
taagagaatt atgcagtgtc gccataacca tgagtataa cactgcggcc aacttacttc	4020
tgacaacgat cggaggaccg aaggagctaa ccgctttttt gcacaacatg ggggatcatg	4080
taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg	4140
acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact ggcgaactac	4200
ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac	4260
cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg	4320
agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg	4380
tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg	4440
agataggtgc ctactgatt aagcattggg aactgtcaga ccaagtttac tcatatatac	4500
tttagattga tttaaaactt cttttttaat ttaaaaggat ctaggtgaag atcctttttg	4560
ataatctcat gacaaaatc ccttaacgtg agttttcggt ccactgagcg tcagacccccg	4620
tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc	4680
aaacaaaaaa accaccgcta ccagcgggtg tttgtttgcc ggatcaagag ctaccaactc	4740

tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc cttctagtgt	4800
agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc	4860
taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact	4920
caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggggt tcgtgcacac	4980
agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag	5040
aaagcggcac gcttcccga gggagaaagg cggacaggta tccggtaagc ggcagggtcg	5100
gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg	5160
tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca ggggggcgga	5220
gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt	5280
ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct	5340
ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg	5400
aggaagcgga agagcgccca atacgcaaac cgctctccc cgcgcgttgg ccgattcatt	5460
aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc aacgcaatta	5520
atgtgagtta gctcactcat taggcacccc aggctttaca ctttatgctt ccggctcgta	5580
tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt	5640
acgccaagct tgccggggcc gacgcggcgg gcggcctcgt ccatcgggac gcgccagtcg	5700
acgccgacga tgtccgcgcc ggctcgccc atgagcccca gcagctcgcc ggtgccgacg	5760
ccgaagtgga tgcgcgggac gccgtggccg gccaccgcgc ggaacacctt cgccgaggcg	5820
ggcagcaccg aacgccggta gtcggagggg gcgagcgcgc cggcccagga gtcgaagagc	5880
tgcacggccg aggcgccggc ccggatctgg acgtcgagga aggccgccgt gatgtcggcg	5940
aggcggtcga gcaggtcggc ccagagctcg gggtcgccgt acatcatcgc cttggcgttc	6000
tcgtacgtac gggacgggccc gccctcgacg aggttaactcg caagggtgaa cggggcgccc	6060
gcgaaaccga tcagcgggggt ggacccgagc tcacgggtca gcatgccgat ggcctcggtg	6120
acgtaggaga cgtcctccgg ggtcaggctg cgcagccggg cgaggtcggc gcgggtgcgc	6180
accggctgct cgacgaccgg gccgatgccg ggcttgatgt cgaggtcgat gccgatggcc	6240
ttgagcggga cgacgatgtc gctgaagtag atcgccgcgt cgacgtggtg gcggcgcacc	6300

ggctggaggg tgatctcggg gaccagctcg ggccgcatgc aggagtcgag catcccgatg	6360
ccctcgcgca ccttgcggta ctccggcagt gagcgcccgg cctgccgcat gaaccacacc	6420
ggcgtgtgcg gcaccggctc gcgccggcac gccttgagga aggctgattc ccggacggcg	6480
tcgttcttga cggggctcggg ggtcgcggta ggcgtctggc ctttcgggct cgtgttggca	6540
ctcacaccgg ccagtctcgc acgaccgctc cgccgccccg gaccccaggg gcgtccggcg	6600
ggcggggccg cgggtgcagga tccggacagc gccggggcggc gcgggtgtcc ctccctgcgc	6660
cgggggcccc ttccgcttaa tcttcccggc atggctgcgg ctacgggacg actgtcggac	6720
ggcgctggcg gaatggacga accgaaggag ggcggggggg atcccgggca cggagggtgcg	6780
cctccgccgc ctttcgggc cgctgtcgag gcgtgcaga gcgccggct gcggccgcag	6840
atcgagggtg agacggtgcc cgcgccgaaa cggctcgccc cgtacgcgca cgcgtggag	6900
gcggcggtcg tcgacggcga ggaggatctg gccgacggcc ggctgggtgct gctgtgcgac	6960
ccggccggac acgacgcctg gcgggggacc ttccgtctgg tgacgctggg gcgcgccgag	7020
ctggagccgg agatggcggc ggatccgctg ctgccggacg tgtgctggtc ctggctgacc	7080
ggcgcgctgg cggcgcgcg cctgtcgtag ggcgagccga gcggcacggg gacgcgggcg	7140
agttcgcact acttcggcgg gctgtccgcg cggcccgcg cctcccagat cgagatccgt	7200
gcctcgtgga cgccgcgtga gggctctggc ggggttccgg acacggccgc ccatctggtc	7260
gcgtgggtccg atctgctggc gcaggctcgc gggctgccgc cggccgctcc gggggacgcg	7320
tccgtgggtga cgctgccgca gcggaggggg ccgcagtcgc gctgagcctc ccgctgcggg	7380
gtcccgcaga ccggcctcct gcattctctt ttgtcgatac ggccactttc ggaagcgttg	7440
cgacgcagac cgaaaccgtt cgatcttcga atgatcgatc gtgcgtccga attgcccga	7500
ttgttactca tcaattcgtg atcattcgtt aaaggacacc aggtttgctg ccgaagacga	7560
ctgtgacctt gaaagcacgg ttcgtcccgc cttaccccc acgagccggc ccgtcccga	7620
ccccaggagg cctgggtgtt aattaatgac ctttcgtcc cgccccgat ccgcccgtcc	7680
ccgatccgcc cgcccacggc caagcgaaca cgttctttca ctcttctgac cggaatacga	7740
cccaccggcg ccggtcacgg agcacaaccg tcgacgggcg ctttcgcggc acggataccc	7800
ttgacaggtg accgacgcc acgacaccgc agcagacagt tcaactgcgca ccaccggagg	7860
cgctcctccg gacgacggcg gatcttctgt tacggaggcg ccgaccccct tgctggaacc	7920

ccgcgagggc attccgcccg tgatagcgga cgaggccgcc ctcgccgagg cggctgccgc	7980
cttcgcggcc ggcagcggac ccgtcgccgt ggacgccgag cgcgcctccg ggtaccgcta	8040
cggccagcgc gcctacctcg tccagctgcg ccgcgaggggt gcgggtaccg cgctgatcga	8100
ccccgtggcc tgccccgacc tgtccgccct cggcgaggcg ctgtccggcg tcgagtgggt	8160
gctgcacgcc gccacccagg acctgccctg tctgcgcgag ataggcatgg tgccctcccg	8220
cctcttcgac accgagctgg ccggccgcct tgccgggttc ccccgcgtcg ggctcggcgc	8280
gatggtcgag aacgtgctcg gcttcgtcct ggagaagggc cactccgccg tcgactggtc	8340
cacccgtccg ctgcccgagc cctggctgcg gtacgccgcc ctcgacgtcg aactgctggt	8400
cgatctgcgg gacgccctgg agaaggagct ggaccgccag ggcaagctgg actgggcccg	8460
gcaggagttc gacgcgatcg cctcggcccc gccgccggag ccccgcaagg acccctggcg	8520
ccgcacctcc ggcatgcaca aggtgcgccg gcgccgccag atggcgggtg tgccggagct	8580
gtgggagacc cgcgaccgga tcgcccggcg ccgtgacgtc tccccggca aggtgctttc	8640
cgacgcggcg atcgtggagg ccgcgctcgc gctgcccgcc aacctgcacg ccatggccgc	8700
gctcaacggg ttcgggcagc ggggtggggcg gcgccagctg gagcagtggc aggcggccgt	8760
cgaccgcgcg aaggcgctga gcgaggccca gctgccgcag cccggccagc cggtgaccgg	8820
ccctccgccg ccgcgcgcct gggcggacaa ggaccccgtt gccgcggccc ggctgtcggc	8880
ggcccgcgcg ggggtcgccg aactcgccga gcggctgaac atgccgccgg agaacctgat	8940
caccccgga acgggtgcga gggctctgtg ggagccgccg gggcccagc agcggctccgt	9000
cgccgcggcg ctgacggcgc acggggcacg cgcgtggcag gtcgaccagg tcaactccgt	9060
gctgggtggc gcgctggcta cttcgtcgcc ccccgcatga acccgttgta gatgaagcgc	9120
tggaggaaca ggaagacgat caaggtgggc aggatgacca ggaccgcgcc cgccgagatc	9180
gtctcccagt gcgcgccgaa ggggcccttg aagcgggaca gggacgtcga gatgaccccc	9240
aggtcctcgg agggcatgta gaggaagggg atgtagaagt cgttgtagac gttgatcccc	9300
tttacgatca ccaccgtcg gatcgccggc ttgagcagcg ggaagatcac cttgcggtag	9360
acgggtgaac cgttggcgcc gtccaggcgc gccgcctcgt ccagggagac ggggatggag	9420
cggatgaact gcaggaagac gtagatcgag acgatgtccg tgcccatgta gagggcgatc	9480

ggcgccacaca ggctgtcgaa catgccgaag ctgttgacga tctggaaggt cgccacctgg 9540
 gtggtcaccc cggggaccag cgcggccagc aggaacagcg ccacgaccag ctagcgaatt 9600
 cctgcaggtc cccggggatc ggtcttgctt tgctcgtcgg tgatgtactt caccagctcc 9660
 gcgaagtgc tcttcttgat ggagcgcag gggacgtgct tggcaatcac gcgcaccccc 9720
 cggccgtttt agcggctaaa aaagtcatgg ctctgccctc gggcggacca cgcccatcat 9780
 gaccttgcca agctcgtcct gcttctcttc gatcttcgcc agcagggcga ggatcgtggc 9840
 atcaccgaac cgcgccgtgc gcgggtcgtc ggtgagccag agtttcagca ggccgcccag 9900
 gcggcccagg tcgccattga tgcgggccag ctgcgggacg tgctcatagt ccacgacgcc 9960
 cgtgattttg tagccctggc cgacggccag caggtaggcc gacaggctca tgccggccgc 10020
 cgccgccttt tcctcaatcg ctcttcgttc gtctggaagg cagtacacct tgataggtgg 10080
 gctgcccttc ctggttggct tggtttcacg agccatccgc ttgccctcat ctgttacgcc 10140
 ggcggtagcc ggccagcctc gcagagcagg attcccgttg agcaccgcca ggtgcgaata 10200
 agggacagtg aagaaggaa acccgctcgc ggggtggcct acttcaccta tcctgcccgg 10260
 ctgacgccgt tggatacacc aaggaaagtc tacacgaacc ctttggcaaa atcctgtata 10320
 tcgtgcgaaa aaggatggat ataccgaaaa aatcgctata atgaccccga agcagggtta 10380
 tgcagcggaa aagatccgtc gacctgcagc ccgggggatc cccgggtacc gagctcgggc 10440
 tggggctcgg ggccgccggt gagctggtag acgaaggcgc ccgagtctcc ttcgttcact 10500
 gcgtgccact cgtggtgcgg gtacttccgg cgcaacgtgc tgtcgtccat gggcggcatc 10560
 atggcagagg cggagacgcc gttccgcgcc tttcgtcggg gcccgtaggg tttcggacat 10620
 tcttgtgcgg ggtggggggg cgccggcgga cccggtgcgc ccggcgtcgc 10670

<210> 9

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Đoạn mồi oligonucleotit tổng hợp

<400> 9

agcacagcta gcttctagaa gcttcattca aaggccggca

40

<210> 10
 <211> 43
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mồi oligonucleotit tổng hợp

<400> 10
 gccagtgaat tctgcagctc gagcagagca ggattcccgt tga 43

<210> 11
 <211> 7173
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Plasmid pKS-T-orf18pfrd tổng hợp

<400> 11
 gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt 60
 caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 120
 ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt 180
 gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 240
 tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt 300
 ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg 360
 tattatcccc tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga 420
 atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 480
 gagaattatg cagtgtgtcc ataaccatga gtgataacac tgcggccaac ttacttctga 540
 caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa 600
 ctgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 660
 ccacgatgcc tntagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta 720
 ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 780
 ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc 840
 gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag 900
 ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga 960

taggtgcctc actgattaag catttgtaac tgtcagacca agtttactca tatatacttt	1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata	1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag	1140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa	1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt	1260
ttccgaaggt aactggcttc agcagagcgc agatacaaaa tactgtcctt ctagtgtagc	1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa	1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa	1440
gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc	1500
ccagcttggg gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa	1560
gcgccacgct tcccgaaggg agaaaggcgg acaggatatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgaggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc	1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccctttgc tggccttttg	1800
ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg	1860
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg	1920
aagcggaaga gcgccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat	1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg	2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt	2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct ggggtaccggg cccccctcg	2220
agcagagcag gattcccgtt gagcaccgcc aggtgcgaat aaggacagt gaagaaggaa	2280
cacccgctcg cgggtgggccc tacttcacct atcctgcccc gctgacgccg ttggatacac	2340
caaggaaagt ctacacgaac cttttggcaa aatcctgtat atcgtgcgaa aaaggatgga	2400
tataccgaaa aaatcgctat aatgaccccg aagcagggtt atgcagcgga aaatgcagct	2460
cacggtaact gatgccgtat ttgcagtacc agcgtacggc ccacagaatg atgtcacgct	2520
gaaaatgccg gcctttgaat gaagcttgcc ggggccgacg cggcgggcgg cctcgtccat	2580

cgggacgcgc cagtcgacgc cgacgatgtc cgcgccggcc tcgcccata ga gccccagcag	2640
ctcgccggtg ccgacgccga agtggatgcg cgggacgccg tggccggcca ccgcgcggaa	2700
caccttcgcc gaggcgggca gcaccgaacg ccggtagtcg gagggggcga gcgcgccggc	2760
ccaggagtcg aagagctgca cggccgaggc gccggcccgg atctggacgt cgaggaaggc	2820
cgccgtgatg tcggcgaggc ggtcgagcag gtcggcccag agctcggggg cgccgtacat	2880
catcgccctg gcgttctcgt acgtacggga cgggccgccc tcgacgaggt aactcgcaag	2940
ggtgaacggg gcgcccgcga aaccgatcag cgggggtggac ccgagctcac gggtcagcat	3000
gccgatggcc tcggtgacgt aggagacgtc ctccggggtc aggtcgcgca gccgggcgag	3060
gtcggcgcggt gtgcgcaccg gctgctcgac gaccggggcc atgccgggct tgatgtcgag	3120
gtcgatgccg atggccttga gcgggacgac gatgtcgctg aagtagatcg ccgcgtcgac	3180
gtggtggcgg cgcaccggct ggagggtgat ctcggtgacc agctcggggc gcatgcagga	3240
gtcgagcatc ccgatgccct cgcgcacctt gcggtactcc ggagtgagc gcccggcctg	3300
ccgcatgaac cacaccggcg tgtgcggcac cggctcgcgc cggcacgcct tgaggaaggc	3360
tgattcccgg acggcgctgt tcttgacggg gtcgggggtc gcggtaggcg tctggccctt	3420
cgggctcgtg ttggcactca caccggccag tctcgacga cccgtccgc gccccggacc	3480
ccaggggcgt ccggcgggcg gggccgcggt gcaggatccg gacagcgccg ggccggcgcg	3540
gtgtccctcc ctgcgccggg ggcccgttcc gcttaatctt cccggcatgg ctgcggctca	3600
gggacgactg tcggacggcg ctggcggaat ggacgaaccg aaggaggcg ggggggatcc	3660
cgggcacgga ggtgcgcctc cgccgccctt ccgggccgct gtcgaggcg tgacagcg	3720
ccggctgcgg ccgcagatcg aggtggagac ggtgcccgcg ccgaaacggc tcgccccgta	3780
cgcgacgcg ctggaggcgg cggctcgtga cggcgaggag gatctggccg acggccggct	3840
ggtgctgctg tgcgaccgg ccggacacga cgcctggcgg gggaccttcc gtctggtgac	3900
gctggtgcgc gccgagctgg agccggagat ggccggcgat ccgctgctgc cggacgtgtg	3960
ctggtcctgg ctgaccggcg cgctggcggc gcgcggcctg tcgtacggcg agccgagcgg	4020
cacggtgacg cgggcgagtt cgactactt cggcgggctg tccgcgcggc ccgccgcctc	4080
ccagatcgag atccgtgcct cgtggacgcc gcgtgagggt ctgggcgggg ttccggacac	4140

ggccgccc	at ctggtcgcgt	ggtccgatct	gctggcgcag	gtcgcggggc	tgccgccggc	4200
cgctccgggg	gacgcgtccg	tggtgacgct	gccgcagcgg	agggggccgc	agtcgcgctg	4260
agcctcccgc	tcggggggtcc	cgacgaccgg	cctcctgcat	ctctctttgt	cgatacggcc	4320
actttcggaa	gcgttgcgac	gcagaccgaa	accgttcgat	cttcgaatga	tcgatcgtgc	4380
gtccgaattg	cccggattgt	tactcatcac	ttcgtgatca	ttcgttaaag	gacaccaggt	4440
ttgctgccga	agacgactgt	gaccttgaaa	gcacggttcg	tcccgccttc	acccccacga	4500
gccggcccgt	cccgaccccc	aggaggcctg	gtgtttaatt	aatgaccctt	ccgtcccgcc	4560
cccgatccgc	ccgtccccga	tccgcccgcc	cacggccaag	cgaacacgtt	ctttcactct	4620
tctgaccgga	atacgaccca	ccggcgcccc	tcacggagca	caaccgtcga	cgggcgcctt	4680
cgcggcacgg	atacccttga	caggtgaccg	acgcccacga	caccgcagca	gacagttcac	4740
tgcgcaccac	cggaggcgct	cctccggacg	acggcggatc	ttctgttacg	gaggcgccga	4800
cccccttgct	ggaaccccgc	gagggcattc	cgcccgtgat	agcggacgag	gccgccctcg	4860
ccgaggcggg	cgccgccttc	gcggccggca	gcggacccgt	cgccgtggac	gccgagcgcg	4920
cctccgggta	ccgctacggc	cagcgcgctt	acctcgtcca	gctgcgccgc	gagggtgcgg	4980
gtaccgcgct	gatcgacccc	gtggcctgcc	ccgacctgtc	cgccctcggc	gaggcgctgt	5040
ccggcgctga	gtgggtgctg	cacgccgcca	cccaggacct	gccctgtctg	cgcgagatag	5100
gcatggtgcc	ctccgccttc	ttcgacaccg	agctggccgg	ccgccttgcc	gggttcccc	5160
gcgtcgggct	cggcgcgatg	gtcgagaacg	tgctcggctt	cgctcctggag	aagggccact	5220
ccgccgtcga	ctggtccacc	cgtccgctgc	ccgagccctg	gctgcggtac	gccgccctcg	5280
acgtcgaact	gctggctgat	ctgcgggacg	ccctggagaa	ggagctggac	cgccagggca	5340
agctggactg	ggcccggcag	gagttcgacg	cgatcgcctc	ggccccgccc	ccggagcccc	5400
gcaaggaccc	ctggcgccgc	acctccggca	tgacacaagg	gcgccggcgc	cgccagatgg	5460
cgggtggtgcg	ggagctgtgg	gagacccgcg	accggatcgc	ccggcgccgt	gacgtctccc	5520
ccggcaagg	gctttccgac	gcggcgatcg	tggaggccgc	gctcgcgctg	cccgccaaacc	5580
tgacagccat	ggccgcgctc	aacgggttcg	ggcagcgggt	ggggcgggcg	cagctggagc	5640
agtggcaggc	ggccgtcgac	cgcgcaagg	cgctgagcga	ggcccagctg	ccgcagcccc	5700
gccagccggg	gaccggccct	ccgccgccgc	gcgcctgggc	ggacaaggac	cccgttgccg	5760

cggccccggct gtcggcggcc cgcgcggggg tcgccgaact cgccgagcgg ctgaacatgc 5820
 cgccggagaa cctgatcacc ccggacacgg tgcgcagggt ctgctgggag ccgccggggc 5880
 ccgacgagcg gtccgtcgcc gcggcgctga cggcgcacgg ggcacgcgcg tggcaggctg 5940
 accaggtcac tcccgtgctg gtggccgcgc tggctacttc gtcgcccccc gcatgaaccc 6000
 gttgtagatg aagcgtgga ggaacaggaa gacgatcaag gtgggcagga tgaccaggac 6060
 cgcgcccgcc gagatcgtct ccagtgcgcc gccgaagggg cccttgaagc ggaacaggga 6120
 cgtcgagatg acccccaggt cctcggaggg catgtagagg aaggggatgt agaagtcgtt 6180
 gtagacgttg atccccctta cgatcaccac cgtcgcgac gccggcttga gcagcgggaa 6240
 gatcaccttg cggtagacgg tgaacgcgtt ggcgccgtcc aggcgcgccg cctcgtccag 6300
 ggagacgggg atggagcgga tgaactgcag gaagacgtag atcgagacga tgtccgtgcc 6360
 catgtagagg gcgatcggcg cccacaggct gtcgaacatg ccgaagctgt tgacgatctg 6420
 gaaggtcgcc acctgggtgg tcaccccggg gaccagcgcg gccagcagga acagcgccac 6480
 gaccagctag ttctagagcg gccgccaccg cgggtggagct ccaattcgcc ctatagttag 6540
 tcgtattacg cgcgctcact ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc 6600
 gttacccaac ttaatgcct tgcagcacat ccccctttcg ccagctggcg taatagcgaa 6660
 gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc tgaatggcga atggaaattg 6720
 taagcgtaa tattttgtta aaattcgct taaatttttg ttaaatacagc tcatttttta 6780
 accaataggc cgaaatcggc aaaatccctt ataaatcaaa agaatagacc gagatagggt 6840
 tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac tccaacgtca 6900
 aagggcgaaa aaccgtctat cagggcgatg gccactacg tgaaccatca ccctaataca 6960
 gttttttggg gtcgaggtgc cgtaaagcac taaatcgga ccctaaaggg agccccgat 7020
 ttagagcttg acgggggaaag ccggcgaacg tggcgagaaa ggaagggaag aaagcgaaag 7080
 gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct gcgcgtaacc accacacccg 7140
 ccgcgcttaa tgcgccgcta cagggcgctg cag 7173

<210> 12
 <211> 29
 <212> ADN

<213> Trình tự nhân tạo

<220>

<223> Đoạn mỗi oligonucleotit tổng hợp

<400> 12

gaatggccat ggttcattgt cagctccat

29

<210> 13

<211> 30

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Đoạn mỗi oligonucleotit tổng hợp

<400> 13

tctcgaggat ccgaatagga acttcggaat

30

<210> 14

<211> 6034

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Plasmid pKS-T-orf18pfrd-AmR tổng hợp

<400> 14

gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt

60

caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa

120

ggaagagtat gattattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt

180

gccttcctgt ttttgctcac ccagaaacgc tggatgaaagt aaaagatgct gaagatcagt

240

tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt

300

ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg

360

tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga

420

atgacttggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa

480

gagaattatg cagtgcctgcc ataaccatga gtgataacac tgcggccaac ttacttctga

540

caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa

600

ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca

660

ccacgatgcc tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta

720

ctctagcttc cggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac	780
ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc	840
gtgggtctcg cggatatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag	900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga	960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt	1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata	1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag	1140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa	1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt	1260
ttccgaaggc aactggcttc agcagagcgc agatacaaaa tactgtcctt ctagtgtagc	1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatactc gctctgctaa	1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa	1440
gacgatagtt accggataag gcgcagcggc cgggctgaac ggggggttcg tgcacacagc	1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa	1560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgaggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat tttgtgatg ctcgtcaggg gggcggagcc	1740
tatggaaaaa cgccagcaac gcggcctttt tacggttctt ggccttttgc tggccttttg	1800
ctcacatgtt ctttctgctg ttatcccctg attctgtgga taaccgtatt accgcctttg	1860
agtgagctga taccgctcgc cgagccgaa cgaccgagcg cagcgagtca gtgagcgagg	1920
aagcggaaga gcgccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat	1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg	2040
tgagttagct cactcattag gcacccagg ctttacactt tatgcttccg gctcgtatgt	2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	2160
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct gggtaccggg cccccctcg	2220
agcagagcag gattcccgtt gagcaccgcc aggtgcgaat aaggacagt gaagaaggaa	2280

cacccgctcg cgggtgggcc tacttcacct atcctgcccg gctgacgccg ttggatacac	2340
caaggaaagt ctacacgaac cttttggcaa aatcctgtat atcgtgcgaa aaaggatgga	2400
tataaccgaaa aaatcgctat aatgacccccg aagcagggtt atgcagcgga aaatgcagct	2460
cacggtaact gatgccgtat ttgcagtacc agcgtacggc ccacagaatg atgtcacgct	2520
gaaaatgccg gcctttgaat gaagcttgcc ggggccgacg cggcgggcgg cctcgtccat	2580
cgggacgcgc cagtcgacgc cgacgatgtc cgcgccggcc tcgcccataga gccccagcag	2640
ctcgccggtg ccgacgccga agtggatgcg cgggacgccg tggccggcca ccgcgcggaa	2700
caccttcgcc gaggcgggca gcaccgaacg ccggtagtcg gagggggcga gcgcgccggc	2760
ccaggagtcg aagagctgca cggccgaggc gccggccccg atctggacgt cgaggaaggc	2820
cgccgtgatg tcggcgaggc ggtcgagcag gtcggcccag agctcggggt cgccgtacat	2880
catcgccttg gcgttctcgt acgtacggga cgggccgccc tcgacgaggt aactcgcaag	2940
ggtgaacggg gcgcccgcga aaccgatcag cggggtggac ccgagctcac gggtcagcat	3000
gccgatggcc tcggtgacgt aggagacgtc ctccgggggtc aggtcgcgca gccgggcgag	3060
gtcggcgcggt gtgcgcaccg gctgctcgac gaccgggccg atgccgggct tgatgtcgag	3120
gtcgatgccg atggccttga gcgggacgac gatgtcgctg aagtagatcg ccgcgtcgac	3180
gtggtggcgg cgcaccggct ggagggtgat ctcggtgacc agctcgggcc gcatgcagga	3240
gtcgagcatc ccgatgccct cgcgcacctt gcggtactcc ggagtgagc gcccggcctg	3300
ccgcatgaac cacaccggcg tgtgcggcac cggctcgcgc cggcacgcct tgaggaaggc	3360
tgattcccgg acggcgtcgt tcttgacggg gtcgggggtc gcggtaggcg tctggccctt	3420
cgggctcgtg ttggcactca caccggccag tctcgcacga cccgtccgcc gccccggacc	3480
ccaggggcgt ccggcgggcg gggccgcggt gcaggatccg aataggaact tcggaatagg	3540
aacttcatga gctcagccaa tcgactggcg agcggcatcg cattcttcgc atccgcctc	3600
tggcggatgc aggaagatca acggatctcg gccagttga cccagggctg tcgccacaat	3660
gtcgcgggag cggatcaacc gagcaaaggc atgaccgact ggaccttcct tctgaaggct	3720
cttctccttg agccacctgt ccgccaaggc aaagcgctca cagcagtggc cattctcgag	3780
ataatcgacg cgtaccaact tgccatcctg aagaatggtg cagtgtctcg gcacccata	3840
gggaaccttt gccatcaact cggcaagatg cagcgtcgtg ttggcatcgt gtcccacgcc	3900

gaggagaagt acctgcccac cgagttcatg gacacgggcg accgggcttg caggcgagtg	3960
aggtggcagg ggcaatggat cagagatgat ctgctctgcc tgtggccccg ctgccgcaaa	4020
ggcaaatgga tgggcgctgc gctttacatt tggcaggcgc cagaatgtgt cagagacaac	4080
tccaaggtcc ggtgtaacgg gcgacgtggc aggatcgaac ggctcgtcgt ccagacctga	4140
ccacgagggc atgacgagcg tccctcccgg acccagcgca gcacgcaggg cctcgatcag	4200
tccaagtggc ccatcttcga ggggccggac gctacggaag gagctgtgga ccagcagcac	4260
accgccgggg gtaaccccaa ggttgagaag ctgaccgatg agctcggctt ttcgccattc	4320
gtattgcacg acattgcact ccaccgtga tgacatcagt cgatcatagc acgatcaacg	4380
gcactgttgc aaatagtcgg tggtgataaa cttatcatcc ccttttgctg atggagctgc	4440
acatgaacca tggccgcgct caacgggttc gggcagcggg tggggcggcg ccagctggag	4500
cagtggcagg cggccgtcga ccgcgcgaag gcgctgagcg aggcccagct gccgcagccc	4560
ggccagccgg tgaccggccc tccgccgccg cgcgcctggg cggacaagga ccccgttgcc	4620
gcggccccggc tgtcggcggc ccgcgcgggg gtcgccgaac tcgccgagcg gctgaacatg	4680
ccgccggaga acctgatcac cccggacacg gtgcgcaggg tctgctggga gccgccgggg	4740
cccgcagagc ggtccgtcgc cgcggcgctg acggcgcacg gggcacgcgc gtggcaggtc	4800
gaccagggtca ctcccgtgct ggtggccgcg ctggctactt cgtcgcccc cgcatgaacc	4860
cgtttagat gaagcgctgg aggaacagga agacgatcaa ggtgggcagg atgaccagga	4920
ccgcgccccg cgagatcgtc tcccagtgcg cgccgaaggg gcccttgaag cggaacaggg	4980
acgtcgagat gacccccagg tcctcggagg gcatgtagag gaaggggatg tagaagtcgt	5040
tgtagacgtt gatccccctt acgatcacca ccgtcgcgat cgccggcttg agcagcggga	5100
agatcacctt gcggtagacg gtgaacgcgt tggcgccgtc caggcgcgcc gcctcgtcca	5160
gggagacggg gatggagcgg atgaactgca ggaagacgta gatcgagacg atgtccgtgc	5220
ccatgtagag ggcgatcggc gccacaggc tgtcgaacat gccgaagctg ttgacgatct	5280
ggaaggtcgc cacctgggtg gtcaccccgg ggaccagcgc ggccagcagg aacagcgcca	5340
cgaccagcta gttctagagc ggccgccacc gcggtggagc tccaattcgc cctatagtga	5400
gtcgtattac gcgcgctcac tggccgtcgt tttaaacgt cgtgactggg aaaaccctgg	5460

cgttacccaa cttaatcgcc ttgcagcaca tccccctttc gccagctggc gtaatagcga 5520
 agaggcccgcc accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggaaatt 5580
 gtaagcgтта atattttgtt aaaattcgcg ttaaattttt gttaaatacag ctcatttttt 5640
 aaccaatagg cggaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg 5700
 ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc 5760
 aaaggcgaa aaaccgtcta tcaggcgat ggcccactac gtgaaccatc accctaata 5820
 agtttttttg ggtcgaggtg ccgtaaagca ctaaatacgga accctaaagg gagccccga 5880
 tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaaggga gaaagcgaaa 5940
 ggagcggcg ctagggcgct ggcaagtgtg gcggtcacgc tgcgcgtaac caccacaccc 6000
 gccgcgctta atgcgccgct acaggcgcg tcag 6034

<210> 15
 <211> 40
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

<400> 15
 agcacagcta gcttctagaa gcttcattca aaggccggca 40

<210> 16
 <211> 44
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

<400> 16
 aggcagctcg agcatatgac tagtcagagc aggattcccg ttga 44

<210> 17
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

<400> 17	
gaatggcata tggttcatgt gcagctccat	30
<210> 18	
<211> 30	
<212> ADN	
<213> Trình tự nhân tạo	
<220>	
<223> Đoạn mồi oligonucleotit tổng hợp	
<400> 18	
tctagaacta gtgaatagga acttcggaat	30
<210> 19	
<211> 8113	
<212> ADN	
<213> Trình tự nhân tạo	
<220>	
<223> Plasmid pKS-orf18ifd-T-AmR(NS) tổng hợp	
<400> 19	
gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt	60
caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa	120
ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt	180
gccttcctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt	240
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt	300
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg	360
tattatcccg tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga	420
atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa	480
gagaattatg cagtgtgcc ataaccatga gtgataacac tgcggccaac ttacttctga	540
caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa	600
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca	660
ccacgatgcc tntagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta	720
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac	780
ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc	840

gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag	900
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga	960
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt	1020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata	1080
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag	1140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa	1200
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt	1260
ttccgaaggt aactggcttc agcagagcgc agatacaaaa tactgtcctt ctagtgtagc	1320
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa	1380
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa	1440
gacgatagtt accggataag gcgcagcggg cgggctgaac ggggggttcg tgcacacagc	1500
ccagcttgga gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa	1560
gcgccacgct tcccgaaggg agaaaggcgg acaggatatcc ggtaagcggc agggtcggaa	1620
caggagagcg cacgaggagg cttccagggg gaaacgcctg gtatctttat agtcctgtcg	1680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc	1740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggcccttttg tggccttttg	1800
ctcacatgtt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg	1860
agtgagctga taccgctcgc gcgagccgaa cgaccgagcg cagcgagtca gtgagcgagg	1920
aagcgggaaga gcgccaata cgcaaaccgc ctctccccgc gcgttgggcg attcattaat	1980
gcagctggca cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg	2040
tgagttagct cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt	2100
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	2160
ccaagcgcgc aattaaccct cactaaaggg aaaaaagct ggggtaccggg cccccctcg	2220
agcatatgtt catgtgcagc tccatcagca aaaggggatg ataagtttat caccaccgac	2280
tatttgcaac agtgccgttg atcgtgctat gatcgactga tgtcatcagc ggtggagtgc	2340
aatgtcgtgc aatacgaatg gcgaaaagcc gagctcatcg gtcagcttct caaccttggg	2400
gttaccctcg gcggtgtgct gctgggtccac agctccttcc gtagcgtccg gccctcgaa	2460

gatgggccac ttggactgat cgaggccctg cgtgctgcgc tgggtccggg agggacgctc	2520
gtcatgccct cgtggtcagg tctggacgac gagccgttcg atcctgccac gtcgcccgtt	2580
acaccggacc ttggagtgtg ctctgacaca ttctggcgcc tgccaaatgt aaagcgcagc	2640
gcccattcat ttgcctttgc ggcagcgggg ccacaggcag agcagatcat ctctgatcca	2700
ttgcccctgc cacctcactc gcctgcaagc ccggtcgccc gtgtccatga actcgaatgg	2760
cagggtacttc tcctcggcgt gggacacgat gccaacacga cgctgcatct tgccgagttg	2820
atggcaaagg ttccctatgg ggtgccgaga cactgcacca ttcttcagga tggcaagtgt	2880
gtacgcgtcg attatctcga gaatgaccac tgctgtgagc gctttgcctt ggcggacagg	2940
tggctcaagg agaagagcct tcagaaggaa ggtccagtcg gtcatgcctt tgctcggttg	3000
atccgctccc gcgacattgt ggcgacagcc ctgggtcaac tgggccgaga tccgttgatc	3060
ttcctgcatc cgccagaggc gggatgcgaa gaatgcgatg ccgctcgcca gtcgattggc	3120
tgagctcatg aagttcctat tccgaagttc ctattcacta gtcagagcag gattcccgtt	3180
gagcaccgcc aggtgcgaat aaggacagc gaagaaggaa caccgctcg cgggtgggcc	3240
tacttcacct atcctgcccg gctgacgccg ttggatacac caaggaaagt ctacacgaac	3300
cctttggcaa aatcctgtat atcgtgcgaa aaaggatgga tataccgaaa aaatcgctat	3360
aatgacccccg aagcagggtt atgcagcgga aaatgcagct cacggtaact gatgccgtat	3420
ttgcagtacc agcgtacggc ccacagaatg atgtcacgct gaaaatgccg gcctttgaat	3480
gaagcttgcc ggggccgacg cggcggggcg cctcgtccat cgggacgcgc cagtcgacgc	3540
cgacgatgtc cgcgccggcc tcgcccata gcccagcag ctgcgccgtg ccgacgccga	3600
agtggatgcg cgggacgccg tggccggcca ccgcgcggaa caccttcgcc gaggcgggca	3660
gcaccgaacg ccggtagtcg gagggggcga gcgcgccggc ccaggagtcg aagagctgca	3720
cggccgaggc gccggcccgg atctggacgt cgaggaaggc cgccgtgatg tcggcgaggc	3780
ggtcgagcag gtcggcccag agctcggggc cgccgtacat catcgccttg gcgttctcgt	3840
acgtacggga cgggccgccc tcgacgaggt aactcgcaag ggtgaacggg gcgcccgcga	3900
aaccgatcag cgggggtggac ccgagctcac gggtcagcat gccgatggcc tcggtgacgt	3960
aggagacgtc ctccggggtc aggtcgcgca gccgggcgag gtcggcgccg gtgcgcaccg	4020

gctgctcgac gaccgggccc atgccgggct tgatgtcgag gtcgatgccg atggccttga	4080
gcgggacgac gatgtcgctg aagtagatcg ccgcgtcgac gtggtggcgg cgcaccggct	4140
ggagggtgat ctcggtgacc agctcgggcc gcatgcagga gtcgagcatc ccgatgccct	4200
cgcgcacctt gcggtactcc ggcagtgagc gcccggcctg ccgcatgaac cacaccggcg	4260
tgtgcggcac cggctcgcgc cggcacgcct tgaggaaggc tgattcccgg acggcgtcgt	4320
tcttgacggg gtcgggggtc gcggtaggcg tctggccctt cgggctcgtg ttggcactca	4380
caccggccag tctcgcacga cccgtccgcc gccccggacc ccaggggcgt ccggcgggcg	4440
gggccgcggt gcaggatccg gacagcgccg ggccggcgcgg gtgtccctcc ctgcgccggg	4500
ggcccgttcc gcttaatctt cccggcatgg ctgcggctca gggacgactg tcggacggcg	4560
ctggcggaat ggacgaaccg aaggagggcg ggggggatcc cgggcacgga ggtgcgcctc	4620
cgccgccctt ccgggccgct gtcgaggcgc tgcagagcgc ccggctgcgg ccgcagatcg	4680
aggtggagac ggtgcccgcg ccgaaacggc tcgccccgta cgcgcacgcg ctggaggcgg	4740
cggtcgtcga cggcgaggag gatctggccg acggccggct ggtgctgctg tgcgaccggg	4800
ccggacacga cgcctggcgg gggaccttcc gtctggtgac gctggtgcgc gccgagctgg	4860
agccggagat ggccggcgat ccgctgctgc cggacgtgtg ctggtcctgg ctgaccggcg	4920
cgctggcggc gcgcggcctg tcgtacggcg agccgagcgg cacggtgacg cgggcgagtt	4980
cgcactactt cggcgggctg tccgcgcggc ccgccgcctc ccagatcgag atccgtgcct	5040
cgtggacgcc gcgtgagggt ctgggcgggg ttccggacac ggccgcccac ctggtcgcgt	5100
ggtccgatct gctggcgcag gtcgcggggc tgccgccggc cgctccgggg gacgcgtccg	5160
tggtgacgct gccgcagcgg agggggccgc agtcgcgctg agcctcccgc tgcggggtcc	5220
cgacgaccgg cctcctgcat ctctctttgt cgatacggcc actttcggaa gcgttgcgac	5280
gcagaccgaa accgttcgat cttcgaatga tcgatcgtgc gtccgaattg cccggattgt	5340
tactcatcac ttcgtgatca ttcgttaaag gacaccaggt ttgctgccga agacgactgt	5400
gaccttgaaa gcacggttcg tcccgccttc acccccacga gccggcccgt cccgcacccc	5460
aggaggcctg gtgtttaatt aatgaccctt ccgtcccgcc ccgatccgc ccgtccccga	5520
tccgccgcc caccggccaag cgaacacgtt ctttactct tctgaccgga atacgacca	5580
ccggcgcccc tcacggagca caaccgtcga cgggcgcctt cgcggcacgg atacccttga	5640

caggtgaccg acgcccacga caccgcagca gacagttcac tgcgcaccac cggaggcgct	5700
cctccggacg acggcggatc ttctgttacg gaggcgccga ccccttgct ggaaccccg	5760
gagggcattc cgccgtgat agcggacgag gccgccctcg ccgaggcggt cgccgccttc	5820
gcggccggca gcggacccgt cgccgtggac gccgagcgcg cctccgggta ccgctacggc	5880
cagcgcgcct acctcgtcca gctgcgccg gaggggtgcgg gtaccgcgct gatcgacccc	5940
gtggcctgcc ccgacctgtc cgccctcggc gaggcgctgt ccggcgtcga gtgggtgctg	6000
cacgccgcca ccaggacct gccctgtctg cgcgagatag gcatggtgcc ctccgcctc	6060
ttcgacaccg agctggccgg ccgccttgcc gggttcccc gcgtcgggct cggcgcgatg	6120
gtcgagaacg tgctcggctt cgtcctggag aagggccact ccgccgtcga ctggtccacc	6180
cgtccgctgc ccgagccctg gctgcggtag gccgccctcg acgtcgaact gctggtcgat	6240
ctgcgggacg ccctggagaa ggagctggac cgccagggca agctggactg ggcccggcag	6300
gagttcgacg cgatcgctc ggccccgccg ccggagcccc gcaaggaccc ctggcgccgc	6360
acctccggca tgcacaaggt gcgccggcg cgccagatgg cggtggtgcg ggagctgtgg	6420
gagacccgcg accggatcg ccggcgccgt gacgtctccc ccggcaaggt gctttccgac	6480
gcggcgatcg tggaggccgc gctcgcgctg ccgccaacc tgcacgcat ggccgcgctc	6540
aacgggttcg ggcagcgggt ggggcggcg cagctggagc agtggcaggc ggccgtcgac	6600
cgcggaagg cgctgagcga ggcccagctg ccgagcccc gccagccggt gaccggccct	6660
ccgccgccgc gcgcctgggc ggacaaggac ccggttgccg cggcccggct gtcggcggcc	6720
cgcgcggggg tcgccgaact cgccgagcgg ctgaacatgc cgccggagaa cctgatcacc	6780
ccggacacgg tgcgcagggt ctgctgggag ccgccggggc ccgacgagcg gtccgtcgcc	6840
gcggcgctga cggcgcacgg ggcacgcgcg tggcaggtcg accaggtcac tcccgtgctg	6900
gtggccgcgc tggctacttc gtcgcccc gcatgaacct gttgtagatg aagcgctgga	6960
ggaacaggaa gacgatcaag gtgggcagga tgaccaggac cgcgccgcc gagatcgtct	7020
cccagtgcgc gccgaagggg cccttgaagc ggaacaggga cgtcgagatg acccccaggt	7080
cctcggaggg catgtagagg aaggggatgt agaagtcgtt gtagacgttg atccccctta	7140
cgatcaccac cgtcgcgatc gccggcttga gcagcgggaa gatcaccttg cggtagacgg	7200

tgaacgcgtt ggcgccgtcc aggcgcgccg cctcgtccag ggagacgggg atggagcgga	7260
tgaactgcag gaagacgtag atcgagacga tgtccgtgcc catgtagagg gcgatcggcg	7320
cccacaggct gtcgaacatg ccgaagctgt tgacgatctg gaaggtcgcc acctgggtgg	7380
tcaccccggg gaccagcgcg gccagcagga acagcgccac gaccagctag ttctagagcg	7440
gccgccaccg cgggtggagct ccaattcgcc ctatagttag tcgtattacg cgcgctcact	7500
ggccgtcgtt ttacaacgtc gtgactggga aaacctggc gttacccaac ttaatcgctt	7560
tgcagcacat ccccttttcg ccagctggcg taatagcgaa gaggcccgca ccgatcgccc	7620
ttccaacag ttgcgcagcc tgaatggcga atggaaattg taagcgtaa tattttgtta	7680
aaattcgcgt taaatTTTTg ttaaatcagc tcatttttta accaataggc cgaaatcggc	7740
aaaatccctt ataaatcaaa agaatagacc gagatagggt tgagtgttgt tccagtttgg	7800
aacaagagtc cactattaaa gaacgtggac tccaacgtca aaggcgaaa aaccgtctat	7860
cagggcgatg gcccactacg tgaaccatca ccctaataa gttttttggg gtcgaggtgc	7920
cgtaaagcac taaatcggaa ccctaaaggg agccccgat ttagagcttg acggggaaag	7980
ccggcgaacg tggcgagaaa ggaagggaag aaagcgaaag gagcgggcgc tagggcgctg	8040
gcaagtgtag cggtcacgct gcgcgtaacc accacacccg ccgcgcttaa tgcgccgcta	8100
cagggcgcgt cag	8113

<210> 20

<211> 7120

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Plasmid pXY152-endorf24-camtsr tổng hợp

<400> 20

gaattcgtaa tcatgtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca	60
cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaata agtgagctaa	120
ctcacattaa ttgcgttgcg ctactgccc gctttccagt cgggaaacct gtcgtgccag	180
ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtatttg gcgctcttcc	240
gcttcctcgc tcaactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct	300
cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg	360

tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc	420
cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga	480
aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct	540
cctgtttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg	600
gcgcttttctc atagctcacg ctgtaggtat ctgagttcgg tgtaggtcgt tcgctccaag	660
ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat	720
cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac	780
aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac	840
tacggctaca ctagaagaac agtatttggt atctgcgctc tgctgaagcc agttaccttc	900
ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtgggtttt	960
tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc	1020
ttttctacgg ggtctgacgc tcagtgaac gaaaactcac gttaagggtat tttggtcatg	1080
agattatcaa aaaggatctt cacctagatc cttttggttc atgtgcagct ccatcagcaa	1140
aaggggatga taagtttatc accaccgact atttgcaaca gtgccgttga tcgtgctatg	1200
atcgactgat gtcacagcg gtggagtga atgtcgtgca atacgaatgg cgaaaagccg	1260
agctcaataa ctgccttaaa aaaattacgc cccgccctgc cactcatcgc agtactgttg	1320
taattcatta agcattctgc cgacatggaa gccatcacia acggcatgat gaacctgaat	1380
cgccagcggc atcagcacct tgtcgccttg cgtataatat ttgcccacg tgaaaacggg	1440
ggcgaagaag ttgtccatat tggccacgtt taagtcaaaa ctggtgaaac tcaccaggg	1500
attggctgag acgaaaaaca tattctcaat aaacccttta gggaaatagg ccaggttttc	1560
accgtaacac gccacatctt gcgaatatat gtgtagaaac tgccggaaat cgtcgtggta	1620
ttcactccag agcgatgaaa acgtttcagt ttgctcatgg aaaacggtgt aacaagggtg	1680
aacactatcc catatcacca gtcaccgtc tttcattgcc atacgaaatt ccgcatgagc	1740
attcatcagg cgggcaagaa tgtgaataaa ggccggataa aacttgtgct tatttttctt	1800
tacggctctt aagaaggccg taatatccag ttgaacggtc tggttatagg tacattgagc	1860
aactgactga aatgcctcaa aatgttcttt acgatgcat tgggatatat caacgggtgt	1920

atatccagtg attttttttct ccatttttagc ttccttagct cctgaaaatc tcgataactc	1980
aaaaaatacg cccggtagtg atcttatttc attatggtga aagttggaac ctcttacgtg	2040
ccgatcaacg tctcattttc gccaaaagtt ggcccagggc ttcccgggtat caacagggac	2100
accaggattt atttattctg cgaagtgatc ttccgtcaca ggtattttatt cgcgataagc	2160
tcatggagcg gcgtaaccgt cgcacaggaa ggacagagaa agcgcggaag cttattttaa	2220
taccgcgcgg gtcccggacg gggaagagcg gggagctttg ccagagagcg acgacttccc	2280
cttgcgttgg tgattgccgg tcagggcagc catccgccat cgtcgcgtag ggtgtcacac	2340
cccaggaatc gcgtcactga acacagcagc cggtaggacg accatgactg agttggacac	2400
catcgcaa at ccgtccgatc cggcgggtgca gcggatcatc gatgtcacca agccgtcgcg	2460
atccaacata aagacaacgt tgatcgagga cgtcgagccc ctcatgcaca gcatcgcggc	2520
cgggggtggag ttcatcgagg tctacggcag cgacagcagt ccttttccat ctgagttgct	2580
ggatctgtgc gggcggcaga acataccggt ccgcctcatc gactcctcga tcgtcaacca	2640
gttgttcaag ggggagcggg aggccaaagac attcggcatc gccgcgtcc ctcgcccggc	2700
caggttcggc gacatcgcga gccggcgtgg ggacgtcgtc gttctcgacg ggggtgaagat	2760
cgtcgggaac atcggcgcga tagtacgcac gtcgctcgcg ctcggagcgt cggggatcat	2820
cctggtcgac agtgacatca ccagcatcgc ggaccggcgt ctcaaaggg ccagccgagg	2880
ttacgtcttc tcccttcccg tcgttctctc cggtcgcgag gaggccatcg ctttcattcg	2940
ggacagcggg atgcaactga tgacgtcaa ggcggtatggc gacatttccg tgaaggaact	3000
cggggacaat ccggatcggc tggccttgct gttcggcagc gaaaagggtg ggccttccga	3060
cctgttcgag gaggcgtctt ccgcctcggg ttccatcccc atgatgagcc agaccgagtc	3120
tctcaacgtt tccgtttccc tcggaatcgc gctgcacgag aggatcgaca ggaatctcgc	3180
ggccaaccga taagctagct agagtcgacc tgcaggtccc cggggatcgg tcttgccctg	3240
ctcgtcggtg atgtacttca ccagctccgc gaagtcgctc ttcttgatgg agcgcattggg	3300
gacgtgcttg gcaatcacgc gcacccccg gccgttttag cggctaaaaa agtcatggct	3360
ctgccctcgg gcggaccacg cccatcatga ccttgccaag ctcgtcctgc ttctcttcga	3420
tcttcgccag cagggcgagg atcgtggcat caccgaaccg gccgtgcgc gggtcgtcgg	3480
tgagccagag tttcagcagg ccgcccaggc ggcccaggtc gccattgatg cgggccagct	3540

cgcgacgtg ctcatagtcc acgacgccc tgattttgta gccctggccg acggccagca	3600
ggtaggccga caggctcatg ccggccgccg ccgccttttc ctcaatcgct cttcgttcgt	3660
ctggaaggca gtacaccttg ataggtgggc tgcccttcct ggttggcttg gtttcatcag	3720
ccatccgctt gccctcatct gttacgccgg cggtagccgg ccagcctcgc agagcaggat	3780
tcccgttgag caccgccagg tgcgaataag ggacagtga gaaggaacac ccgctcgcgg	3840
gtgggcctac ttcacctatc ctgcccggct gacgccgttg gatacaccaa ggaaagtcta	3900
cacgaaccct ttggcaaaat cctgtatatc gtgcgaaaaa ggatggatat accgaaaaaa	3960
tcgctataat gaccccgaag cagggttatg cagcgaaaaa gatccgtcga cctgcaggca	4020
tgcaagctct agcgattcca gacgtcccga aggcgtggcg cggcttcccc gtgccggagc	4080
aatcgccctg ggtgggttac acgacgcccc tctatggccc gtactgacgg acacaccgaa	4140
gccccggcgg caaccctcag cggatgcccc ggggcttcac gttttcccag gtcagaagcg	4200
gttttcggga gtagtgcccc aactggggta acctttgagt tctctcagtt gggggcgtag	4260
ggtcgccgac atgacacaag gggttgtgac cgggggtggac acgtacgcgg gtgcttacga	4320
ccgtcagtcg cgcgagcgcg agaactcgag cgcagcaagc ccagcgacac agcgtagcgc	4380
caacgaagac aaggcggccg accttcagcg cgaagtcgag cgcgacgggg gccggttcag	4440
gttcgtcggg catttcagcg aagcgccggg cacgtcggcg ttcgggacgg cggagcgccc	4500
ggagttcgaa cgcatacctga acgaatgccg cgccggggcg ctcaacatga tcattgtcta	4560
tgacgtgtcg cgcttctcgc gcctgaaggt catggacgcg attccgattg tctcggaatt	4620
gctcgccctg ggcgtgacga ttgtttccac tcaggaaggc gtcttcggc agggaaacgt	4680
catggacctg attcacctga ttatgaggct cgacgcgtcg cacaagaat cttcgctgaa	4740
gtcggcgaag attctcgaca cgaagaacct tcagcgcgaa ttgggcgggt acgtcggcgg	4800
gaaggcgcct tacggcttcg agcttgtttc ggagacgaag gagatcacgc gcaacggccg	4860
aatggatcaat gtcgtcatca acaagcttgc gcactcgacc actcccctta ccggaccctt	4920
cgagttcgag cccgacgtaa tccgggtgtg gtggcgtgag atcaagacgc acaaacacct	4980
tcccttcaag ccgggcagtc aagccgccat tcacccgggc agcatcacgg ggctttgtaa	5040
gcgcatggac gctgacgccg tgccgacccg gggcgagacg attgggaaga agaccgcttc	5100

aagcgcctgg gacccggcaa ccgttatgcg aatccttcgg gacccgcgta ttgcgggctt	5160
cgccgctgag gtgatctaca agaagaagcc ggacggcacg ccgaccacga agattgaggg	5220
ttaccgcatt cagcgcgacc cgatcacgct ccggccggtc gagcttgatt gcggaccgat	5280
catcgagccc gctgagtggg atgagcttca ggctgtggtg gacggcaggg ggcgcgggcaa	5340
ggggctttcc cgggggcaag ccattctgtc cgccatggac aagctgtact gcgagtgtgg	5400
cgccgtcatg acttcgaagc gcggggaaga atcgatcaag gactcttacc gctgccgtcg	5460
ccggaagggtg gtcgaccgt ccgcacctgg gcagcacgaa ggcacgtgca acgtcagcat	5520
ggcggcactc gacaagtctg ttgcggaacg catcttcaac aagatcaggc acgccgaagg	5580
cgacgaagag acgttggcgc ttctgtggga agccgcccga cgcttcggca agctcactga	5640
ggcgccctgag aagagcggcg aacgggcgaa ccttgttgcg gagcgcgccg acgccctgaa	5700
cgcccttgaa gagctgtacg aagaccgcgc ggcaggcgcg tacgacggac ccgttggcag	5760
gaagcacttc cggaagcaac aggcagcgct gacgctccgg cagcaagggg cggaagagcg	5820
gcttgccgaa cttgaagccg ccgaagcccc gaagcttccc cttgaccaat ggttccccga	5880
agacgccgac gctgaccga ccggccctaa gtcgtgggtg gggcgcgctg cagtagacga	5940
caagcgcgtg ttcgtcgggc tcttcgtaga caagatcgtt gtcacgaagt cgactacggg	6000
cagggggcag ggaacgcccc tcgagaagcg cgcttcgac acgtgggcga agccgccgac	6060
cgacgacgac gaagacgacg ccaggacgg cacggaagac gtagcggcgt aggtgagtga	6120
agatctagac gtggcaccgc gatgctgttg tgggcacaat cgtgccggtt ggtaggatcc	6180
acatatggaa ataagttcgc tctccaccga cggctccccg cggatcgacg gggagagtcc	6240
cgagcacgtg gaaatgctgg ccgccgccga caccgcgctt ccaccgatca tggcgcaccg	6300
ccgcaccggg cgggtcatcg acggcatgca ccggctgcgc gccgcgatgc tgacggggccg	6360
tacgacgatc gcggtgaggt tcttcgacgg caccgaggag gacgccttcg tcctcgccgt	6420
gaagtcgaac atcgcgcacg gactgccgct gtccgccgcc gaccgccggc gggccgccgg	6480
gcgcatcatg gccacccatc cccggtggtc ggaccggatg atcgctcgg tggcggcac	6540
ctccgccagg acggtcgccg agatccgccg cgacgccggc gccgccgggg cgggggagcc	6600
caccgcgacg ggccgggacg gcagggtacg gcccgtcgac gtgagcgagg gccgcagact	6660
ggcccacgac atgatcgtcc gcgaccggg cctgtcgctg cgccaggctg cccgcgccgc	6720

cgggatctcg ccggagaccg tcagggacgt cagacaccgg atgctccgcg gtgaggaccc 6780
 ggtgcccgcg ccgcggccgc ggaccctggt ggagcgcggc gcggaccgcc gggcggagcc 6840
 ggccgggaag gccgccgcgc cgtgcgggac ggagccgccg cccgccgtcg tgatgaagcg 6900
 gctgagggcc gatccggcgc tgcgtctcaa cgagaacgga cgcgacctgc tgcggcttct 6960
 ggatatccac acggtccggc tggaggactg gaaccgcatt atcgaaagcg tgccgccgca 7020
 ccgtctggag acggtggcgc agctggcacg ctcttcgcc gacaaatggt ccgagatcgc 7080
 gtcacgcacg gaaagcaacg catcacatct ggccgggtga 7120

<210> 21
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

<400> 21
 gtggcaattt aaatggaaat gtgcgcggaa 30

<210> 22
 <211> 31
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mỗi oligonucleotit tổng hợp

<400> 22
 tatatagagc tcaacttggc ctgacagtta c 31

<210> 23
 <211> 7162
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> pXY152-endorf24-blatsr tổng hợp

<400> 23
 gaattcgtaa tcatgtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca 60
 cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaata agtgagctaa 120

ctcacattaa ttgcgttgcg ctactgccc gctttccagt cgggaaacct gtcgtgccag	180
ctgcattaat gaatcgcca acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc	240
gcttcctcgc tactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct	300
cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg	360
tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc	420
cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga	480
aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct	540
cctgttccga cctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg	600
gcgcttttctc atagctcacg ctgtaggtat ctgagttcgg ttaggtcgt tcgctccaag	660
ctgggctgtg tgcacgaacc cccggttcag cccgaccgct gcgccttatac cggtaaactat	720
cgtcttgagt ccaaccgggt aagacacgac ttatcgccac tggcagcagc cactggtaac	780
aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac	840
tacggctaca ctagaagaac agtatttgggt atctgcgctc tgctgaagcc agttaccttc	900
ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cgggtggtttt	960
tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc	1020
ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg	1080
agattatcaa aaaggatctt cacctagatc cttttggttc atgtgcagct ccatcagcaa	1140
aaggggatga taagtttatc accaccgact atttgcaaca gtgccgttga tcgtgctatg	1200
atcgactgat gtcacagcg gtggagtgcg atgtcgtgca atacgaatgg cgaaaagccg	1260
agctcaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctacgcgac	1320
tgtctatttc gttcatccat agttgcctga ctccccgtcg ttagataac tacgatacgg	1380
gagggcttac catctggccc cagtgtgca atgataccgc gagaccacg ctacccggct	1440
ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca	1500
actttatccg cctccatcca gtctattaat tggtgccggg aagctagagt aagtagttcg	1560
ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg	1620
tcgtttggta tggcttcatt cagctccggt tccaacgat caaggcgagt tacatgatcc	1680
cccatgttgt gcaaaaaagc ggtagctcc ttcggtcctc cgatcgttgt cagaagtaag	1740

ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg	1800
ccatccgtaa gatgcttttc tgtgactggg gagtactcaa ccaagtcatt ctgagaatag	1860
tgtatgcggc gaccgagttg ctcttgcccgc gcgtaatac gggataatac cgcgccacat	1920
agcagaactt taaaagtgcct catcattgga aaacgttctt cggggcgaaa actctcaagg	1980
atcttaccgc tgttgagatc cagttcgatg taaccctctc gtgcacccaa ctgatcttca	2040
gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca	2100
aaaaaggga taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat	2160
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatattga atgtatttag	2220
aaaaataaac aaataggggt tccgcgcaca tttccattta aataccgcgc ggggtcccga	2280
cggggaagag cggggagctt tgccagagag cgacgacttc cccttgcgtt ggtgattgcc	2340
ggtcagggca gccatccgcc atcgctcgct aggggtgtcac accccaggaa tcgcgtcact	2400
gaacacagca gccggtagga cgaccatgac tgagttggac accatcgcaa atccgtccga	2460
tccggcgggtg cagcggatca tcgatgtcac caagccgtcg cgatccaaca taaagacaac	2520
gttgatcgag gacgtcgagc ccctcatgca cagcatcgcg gccgggggtg agttcatcga	2580
ggtctacggc agcgacagca gtccttttcc atctgagttg ctggatctgt gcgggcggca	2640
gaacataccg gtccgcctca tcgactcctc gatcgtcaac cagttgttca agggggagcg	2700
gaaggccaag acattcggca tcgcccgcgt ccctcgccc gccagggtcg gcgacatcg	2760
gagccggcgt ggggacgtcg tcgttctcga cggggtgaag atcgtcggga acatcggcgc	2820
gatagtacgc acgtcgctcg cgctcggagc gtcggggatc atcctggctg acagtacat	2880
caccagcatc gcggaccggc gtctccaaag ggccagccga ggttacgtct tctcccttcc	2940
cgctgttctc tccggtcgcg aggaggccat cgccttcatt cgggacagcg gtatgcaact	3000
gatgacgctc aaggcggatg gcgacatttc cgtgaaggaa ctcggggaca atccggatcg	3060
gctggccttg ctgttcggca gcgaaaaggg tgggccttcc gacctgttcg aggaggcgtc	3120
ttccgcctcg gtttccatcc ccatgatgag ccagaccgag tctctcaacg tttccgtttc	3180
cctcggaatc gcgctgcacg agaggatcga caggaatctc gcggccaacc gataagctag	3240
ctagagtcga cctgcaggtc cccggggatc ggtcttgctt tgctcgtcgg tgatgtactt	3300

caccagctcc gcgaagtcgc tcttcttgat ggagcgcgatg gggacgtgct tggcaatcac	3360
gcgcaccccc cggccgtttt agcggctaaa aaagtcattg ctctgccctc gggcggacca	3420
cgcccatcat gaccttgcca agctcgtcct gcttctcttc gatcttcgcc agcagggcga	3480
ggatcgtggc atcaccgaac cgcgccgtgc gcgggtcgtc ggtgagccag agtttcagca	3540
ggccgcccag gcggcccagg tcgccattga tgcgggccag ctgcggacg tgctcatagt	3600
ccacgacgcc cgtgattttg tagccctggc cgacggccag caggtaggcc gacaggctca	3660
tgccggccgc cgccgccttt tcctcaatcg ctcttcgttc gtctggaagg cagtacacct	3720
tgataggtgg gctgcccttc ctggttggct tggtttcatc agccatccgc ttgccctcat	3780
ctgttacgcc ggcggtagcc ggccagcctc gcagagcagg attcccgttg agcaccgcca	3840
ggtgcgaata agggacagtg aagaaggaaac acccgctcgc ggggtgggcct acttcaccta	3900
tcctgcccgg ctgacgccgt tggatacacc aaggaaagtc tacacgaacc ctttggcaaa	3960
atcctgtata tcgtgcgaaa aaggatggat ataccgaaaa aatcgctata atgaccccga	4020
agcagggtta tgcagcggaa aagatccgtc gacctgcagg catgcaagct ctagcgattc	4080
cagacgtccc gaaggcgtgg cgcggcttcc ccgtgccgga gcaatcgccc tgggtggggtt	4140
acacgacgcc cctctatggc ccgtactgac ggacacaccg aagccccggc ggcaaccctc	4200
agcggatgcc ccggggcttc acgttttccc aggtcagaag cggttttcgg gagtagtgcc	4260
ccaactgggg taacctttga gttctctcag ttgggggcgt agggtcgccg acatgacaca	4320
aggggttggt accgggggtg acacgtacgc gggtgcttac gaccgtcagt cgcgcgagcg	4380
cgagaactcg agcgcagcaa gccagcgac acagcgtagc gccaacgaag acaaggcggc	4440
cgaccttcag cgcgaagtcg agcgcgacgg gggccggttc aggttcgtcg ggcatttcag	4500
cgaagcgccg ggcacgtcgg cgttcgggac ggcggagcgc ccggagttcg aacgcatact	4560
gaacgaatgc cgcgccgggc ggctcaacat gatcattgtc tatgacgtgt cgcgttctc	4620
g'gcctgaag gtcatggacg cgattccgat tgtctcgga ttgctcgccc tgggcgtgac	4680
gattgtttcc actcaggaag gcgtcttccg gcagggaac gtcattggacc tgattcacct	4740
gattatgcgg ctcgacgcgt cgcacaaaga atcttcgctg aagtcggcga agattctcga	4800
cacgaagaac cttcagcgcg aattgggcgg gtacgtcggc gggaaggcgc cttacggctt	4860
cgagcttggt tcggagacga aggagatcac gcgcaacggc cgaatgggtca atgtcgtcat	4920

caacaagctt ggcactcga ccactcccct taccggaccc ttcgagttcg agcccgacgt	4980
aatccggtgg tggtagcggtg agatcaagac gcacaaacac cttcccttca agccgggcag	5040
tcaagccgcc attcaccggg gcagcatcac ggggctttgt aagcgcatgg acgctgacgc	5100
cgtgccgacc cggggcgaga cgattgggaa gaagaccgct tcaagcgctt gggacccggc	5160
aaccgttatg cgaatccttc gggaccgcg tattgcgggc ttcgccgctg aggtgatcta	5220
caagaagaag ccggacggca cgccgaccac gaagattgag ggttaccgca ttcagcgca	5280
cccgatcacg ctccggccgg tcgagcttga ttgcggaccg atcatcgagc ccgctgagtg	5340
gtatgagctt caggcgtggt tggacggcag ggggcgcggc aaggggcttt cccgggggca	5400
agccattctg tccgccatgg acaagctgta ctgcgagtgt ggcgccgtca tgacttcgaa	5460
gcgcggggaa gaatcgatca aggactctta ccgctgccgt cgccggaagg tggtcgaccc	5520
gtccgcacct gggcagcacg aaggcacgtg caacgtcagc atggcggcac tcgacaagtt	5580
cgttgccgaa cgcactttca acaagatcag gcacgccgaa ggcgacgaag agacgttggc	5640
gcttctgtgg gaagccgccc gacgcttcgg caagctcact gaggcgcctg agaagagcgg	5700
cgaacgggag aaccttggtg cggagcgcgc cgacgccctg aacgcccttg aagagctgta	5760
cgaagaccgc gcggcaggcg cgtacgacgg acccgttggc aggaagcact tccggaagca	5820
acaggcagcg ctgacgctcc ggcagcaagg ggcggaagag cggcttgccg aacttgaagc	5880
cgccgaagcc ccgaagcttc cccttgacca atggttcccc gaagacgccg acgctgaccc	5940
gaccggccct aagtcgtggt gggggcgcgc gtcagtagac gacaagcgcg tgttcgtcgg	6000
gctcttcgta gacaagatcg ttgtcacgaa gtcgactacg ggcagggggc agggaaacgcc	6060
catcgagaag cgcgcttcga tcacgtgggc gaagccgccg accgacgacg acgaagacga	6120
cgcccaggac ggcacggaag acgtacggc gtaggtgagt gaagatctag acgtggcacc	6180
gcgatgctgt tgtgggcaca atcgtgccgg ttggtaggat ccacatatgg aaataagttc	6240
gctctccacc gacggctccc cgcggatcga cggggagagt cccgagcacg tggaaatgct	6300
ggccgccgcc gacaccgcgc ttccaccgat catggtgcac cgccgcaccg ggcgggtcat	6360
cgacggcatg caccggctgc gcgccgat gctgacgggc cgtacgacga tcgcggtgag	6420
gttcttcgac ggcaccgagg aggacgcctt cgtcctcgcc gtgaagtcga acatcgcgca	6480

cggactgccg ctgtccgccg ccgaccgccg gcgggccgcc gggcgcatca tggccaccca 6540
 tccccggtgg tcggaccgga tgatgcctc ggtggctggc acctccgcca ggacggtcgc 6600
 cgagatccgc cgcgacgccg gcgccgccgg ggccgggggag cccacccgca tcggccggga 6660
 cggcagggtta cggcccgtcg acgtgagcga gggccgcaga ctggcccacg acatgatcgt 6720
 ccgcgacccg ggcctgtcgc tgcgccaggt cggccgcgcc gccgggatct cgccggagac 6780
 cgtcagggac gtcagacacc ggatgctccg cggtgaggac ccggtgcccg cgccgcggcc 6840
 gcggaccctg gtggagcgcg gcgcggaccg ccggggcggag ccggccggga aggccgccgc 6900
 gccgtgcggg acggagccgc cggccgccgt cgtgatgaag cggctgaggg ccgatccggc 6960
 gctgcgtctc aacgagaacg gacgcgacct gctgcggctt ctggatatcc acacggtcgc 7020
 gctggaggac tggaaccgca ttatcgaaag cgtgccgccg caccgtctgg agacgggtggc 7080
 gcagctggca cgctcctgcg ccgacaaatg gtccgagatc gcgtcacgca tcgaaagcaa 7140
 cgcatcacat ctggccgggt ga 7162

<210> 24
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mồi oligonucleotit tổng hợp

<400> 24
 aaggagaaga gccttcagaa ggaa

24

<210> 25
 <211> 321
 <212> PRT
 <213> S. griseus

<400> 25

Met Asp Pro Thr Arg Val Asp Ile Phe Ala Leu Pro Ala Val Glu Ile
 1 5 10 15

Glu Leu Ser Arg Leu Ser Ser Ala Ser Ser Pro Arg Thr Ser Gly Glu
 20 25 30

Asp Pro Glu His Val Glu Thr Leu Leu Ser Ala Glu Gly Glu Leu Pro

35

40

45

Pro Ile Leu Val His Arg Pro Thr Met Gln Val Leu Asp Gly Leu His
 50 55 60

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

Leu Val Asp Ala Thr Glu Ser Asp Ala Phe Val Leu Ala Val Glu Ala
 85 90 95

Asn Ile Arg His Gly Leu Pro Leu Ser Leu Ala Asp Arg Lys Arg Ala
 100 105 110

Ala Val Gln Ile Ile Gly Thr His Pro Gln Trp Ser Asp Arg Arg Val
 115 120 125

Ala Ser Ala Thr Gly Ile Ser Ala Gly Thr Val Ala Asp Leu Arg Arg
 130 135 140

Arg Ala Gly Glu Asp Gly Thr Glu Ala Arg Ile Gly Arg Asp Gly Arg
 145 150 155 160

Val Arg Pro Ser Asp Gly Ser Glu Arg Arg Arg Leu Ala Ala Glu Leu
 165 170 175

Ile Arg Ser Asp Pro Gly Leu Ser Leu Arg Gln Val Ala Lys Gln Val
 180 185 190

Gly Ile Ser Pro Glu Thr Val Arg Asp Val Arg Gly Arg Leu Glu Arg
 195 200 205

Gly Glu Ser Pro Thr Pro Asp Gly Thr Arg Arg Leu Pro Ala Lys Pro
 210 215 220

His Pro Leu Arg Leu Ser Glu Pro Asp Phe Gly Arg Ala Val Asp Gln
 225 230 235 240

Asp Arg Leu Ala Leu Leu Glu Arg Leu Lys Ser Asp Pro Ala Leu Arg
 245 250 255

Leu Asn Glu Val Gly Arg Ile Leu Leu Arg Met Leu Thr Met His Ser
 260 265 270

Met Asp Gly Gln Glu Trp Glu Arg Ile Leu Gln Gly Val Pro Pro His
 275 280 285

Leu His Gly Val Ile Ala Gly Phe Ala Arg Asp His Ala Arg Val Trp
 290 295 300

Ala Glu Phe Ala Asp His Leu Glu Ser Arg Ala Thr Glu Leu Ala Ala
 305 310 315 320

Gly

<210> 26
 <211> 311
 <212> PRT
 <213> Streptomyces fungicidicus

<400> 26

Val Glu Ile Ser Ser Leu Ser Thr Asp Gly Ser Pro Arg Ile Asp Gly
 1 5 10 15

Glu Ser Pro Glu His Val Glu Met Leu Ala Ala Ala Asp Thr Ala Leu
 20 25 30

Pro Pro Ile Met Val His Arg Arg Thr Gly Arg Val Ile Asp Gly Met
 35 40 45

His Arg Leu Arg Ala Ala Met Leu Thr Gly Arg Thr Thr Ile Ala Val
 50 55 60

Arg Phe Phe Asp Gly Thr Glu Glu Asp Ala Phe Val Leu Ala Val Lys
 65 70 75 80

Ser Asn Ile Ala His Gly Leu Pro Leu Ser Ala Ala Asp Arg Arg Arg
 85 90 95

Ala Ala Gly Arg Ile Met Ala Thr His Pro Arg Trp Ser Asp Arg Met
 100 105 110

Ile Ala Ser Val Val Gly Thr Ser Ala Arg Thr Val Ala Glu Ile Arg
 115 120 125

Arg Asp Ala Gly Ala Ala Gly Ala Gly Glu Pro Thr Arg Ile Gly Arg
 130 135 140

Asp Gly Arg Val Arg Pro Val Asp Val Ser Glu Gly Arg Arg Leu Ala
 145 150 155 160

His Asp Met Ile Val Arg Asp Pro Gly Leu Ser Leu Arg Gln Val Ala
 165 170 175

Arg Ala Ala Gly Ile Ser Pro Glu Thr Val Arg Asp Val Arg His Arg
 180 185 190

Met Leu Arg Gly Glu Asp Pro Val Pro Ala Pro Arg Pro Arg Thr Leu
 195 200 205

Val Glu Arg Gly Ala Asp Arg Arg Ala Glu Pro Ala Gly Lys Ala Ala
 210 215 220

Ala Pro Cys Gly Thr Glu Pro Pro Pro Ala Val Val Met Lys Arg Leu
 225 230 235 240

Arg Ala Asp Pro Ala Leu Arg Leu Asn Glu Asn Gly Arg Asp Leu Leu
 245 250 255

Arg Leu Leu Asp Ile His Thr Val Arg Leu Glu Asp Trp Asn Arg Ile
 260 265 270

Ile Glu Ser Val Pro Pro His Arg Leu Glu Thr Val Ala Gln Leu Ala
 275 280 285

Arg Ser Cys Ala Asp Lys Trp Ser Glu Ile Ala Ser Arg Ile Glu Ser
 290 295 300

Asn Ala Ser His Leu Ala Gly

305

310

<210> 27
 <211> 15
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Đoạn mồi oligonucleotit tổng hợp

<400> 27
 gtgtttaatt aatga

15

<210> 28
 <211> 321
 <212> PRT
 <213> Amycolatopsis

<400> 28

Met Asp Pro Thr Arg Val Asp Ile Phe Ala Leu Pro Ala Val Glu Ile
 1 5 10 15

Glu Leu Ser Arg Leu Ser Ser Ala Ser Ser Pro Arg Thr Ser Gly Glu
 20 25 30

Asp Pro Glu His Val Glu Thr Leu Leu Ser Ala Glu Gly Glu Leu Pro
 35 40 45

Pro Ile Leu Val His Arg Pro Thr Met Gln Val Leu Asp Gly Leu His
 50 55 60

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

Leu Val Asp Ala Thr Glu Ser Asp Ala Phe Val Leu Ala Val Glu Ala
 85 90 95

Asn Ile Arg His Gly Leu Pro Leu Ser Leu Ala Asp Arg Lys Arg Ala
 100 105 110

Ala Val Gln Ile Ile Gly Thr His Pro Gln Trp Ser Asp Arg Arg Val
 115 120 125

Ala Ser Ala Thr Gly Ile Ser Ala Gly Thr Val Ala Asp Leu Arg Arg
 130 135 140

Arg Ala Gly Glu Asp Gly Thr Glu Ala Arg Ile Gly Arg Asp Gly Arg
 145 150 155 160

Val Arg Pro Ser Asp Gly Ser Glu Arg Arg Arg Leu Ala Ala Glu Leu
 165 170 175

Ile Arg Ser Asp Pro Gly Leu Ser Leu Arg Gln Val Ala Lys Gln Val
 180 185 190

Gly Ile Ser Pro Glu Thr Val Arg Asp Val Arg Gly Arg Leu Glu Arg
 195 200 205

Gly Glu Ser Pro Thr Pro Asp Gly Thr Arg Arg Leu Pro Ala Lys Pro
 210 215 220

His Pro Leu Arg Leu Ser Glu Pro Asp Phe Gly Arg Ala Val Asp Gln
 225 230 235 240

Asp Arg Leu Ala Leu Leu Glu Arg Leu Lys Ser Asp Pro Ala Leu Arg
 245 250 255

Leu Asn Glu Val Gly Arg Ile Leu Leu Arg Met Leu Thr Met His Ser
 260 265 270

Met Asp Gly Gln Glu Trp Glu Arg Ile Leu Gln Gly Val Pro Pro His
 275 280 285

Leu His Gly Val Ile Ala Gly Phe Ala Arg Asp His Ala Arg Val Trp
 290 295 300

Ala Glu Phe Ala Asp His Leu Glu Ser Arg Ala Thr Glu Leu Ala Ala
 305 310 315 320

Gly

<210> 29
 <211> 395
 <212> PRT
 <213> *S. kasugaensis*

<400> 29

Met Ala Glu Thr Val Arg Ala Asp Ser Pro Leu Lys Ser Ser Tyr Arg
 1 5 10 15

Asn Val Pro Ala Ala Glu Val Gln Gly Ser Gly Leu Ser Val Gly Gln
 20 25 30

Arg Thr Thr Arg Ile Ala Ile Ser Ser Leu Leu Ala Ala Asp Ser Pro
 35 40 45

Arg Ser Ala Gly Glu Asn Ala Glu His Ile Arg Leu Leu Ala Asp Ser
 50 55 60

Gly Ala Arg Leu Pro Pro Ile Val Val Gln Arg Ser Thr Met Arg Val
 65 70 75 80

Ile Asp Gly Met His Arg Leu Arg Ala Ala Ala Leu Arg Gly Glu Thr
 85 90 95

Glu Ile Glu Val Arg Phe Phe Asp Gly Ala Glu Glu Asp Ser Phe Leu
 100 105 110

Leu Ala Val Arg Ser Asn Ile Ala His Gly Leu Pro Leu Ser Gln Glu
 115 120 125

Glu Arg Ala Ala Ala Ala Gln Arg Ile Ile Arg Ser His Ala Gln Trp
 130 135 140

Ser Asn Gln Ala Ile Gly Glu Val Thr Gly Leu Asp Ala Lys Thr Ile
 145 150 155 160

Ala Ala Leu Arg Arg Asp Ala Lys Asp Val Pro Gln Leu Asp Ala Arg
 165 170 175

Ile Gly Arg Asp Gly Arg Val Arg Pro Val Asp Gly Ala Gln Gly Arg
 180 185 190

Arg Leu Ala Gly Glu Leu Met Ala Glu Gln Pro Asp Ala Pro Leu Arg
 195 200 205

Lys Ile Ala His Ala Ala Gly Val Ser Leu Gly Thr Ala Ser Asp Val
 210 215 220

Arg Arg Arg Ile Arg Asn Gly Gln Asp Pro Val Pro Ala Gly Arg Gln
 225 230 235 240

Lys Ala Asp Pro Gln Pro Pro Ala Arg Tyr Ala Ala Ser Glu Asp Arg
 245 250 255

Ser Gly Thr Thr Ala Pro Arg Thr Gly Glu Gln Asn Arg Arg Val Leu
 260 265 270

Leu Gln Lys Leu Arg Lys Asp Pro Ser Leu Arg Cys Asn Glu Ala Gly
 275 280 285

Arg Ala Leu Leu Arg Trp Leu Glu Val Gln Ala Val Glu Gly Glu Asp
 290 295 300

Trp Glu Arg Leu Leu Asp Ser Val Pro Met His Cys Ala Ala Thr Ile
 305 310 315 320

Val Glu Leu Ala Arg Arg Lys Asp Pro Ser Leu Arg Cys Asn Glu Ala
 325 330 335

Gly Arg Ala Leu Leu Arg Trp Leu Glu Val Gln Ala Val Glu Gly Glu
 340 345 350

Asp Trp Glu Arg Leu Leu Asp Ser Val Pro Met His Cys Ala Ala Thr
 355 360 365

Ile Val Glu Leu Ala Arg Gly Cys Ser Gly Val Trp Gln Asp Phe Ala
 370 375 380

Ala Gln Leu Glu Arg Arg Gly Arg Ala Ser Ala
 385 390 395

<210> 30
 <211> 367
 <212> PRT
 <213> S. niveus

<400> 30

Met Thr Asn Ser Gly Asp Glu Glu Ile Thr Pro Ala Ser Leu Lys Ala
 1 5 10 15

Thr Arg Lys Gly Glu Arg Val Ser Ile Gly Ser Leu Leu Pro Pro Ser
 20 25 30

Glu Leu Val Arg Ser Gly Glu Ser Thr Glu His Ile Arg Val Leu Ala
 35 40 45

Glu Thr Asp Glu Asp Leu Pro Pro Ile Val Val His Arg Gly Thr Arg
 50 55 60

Arg Val Val Asp Gly Met His Arg Leu Trp Ala Ala Arg Phe Arg Gly
 65 70 75 80

Asp Glu Ser Ile Glu Val Val Phe Val Asp Gly Ser Pro Ala Asp Val
 85 90 95

Phe Val Leu Ala Val Glu Leu Asn Arg Ala His Gly Leu Pro Leu Thr
 100 105 110

Leu Asp Glu Arg Lys Ser Ala Ala Ala Gln Ile Met Asp Ser His Pro
 115 120 125

His Trp Ser Asp Arg Lys Ile Ala Arg Thr Thr Gly Leu Ala Ala Ser
 130 135 140

Thr Val Ala Ser Leu Arg Ser Ser Ser Thr Ala Gly Thr Val Gly Arg
 145 150 155 160

Arg Thr Gly Gln Asp Gly Arg Ser Arg Pro Asn Asp Gly Thr Asp Gly
 165 170 175

Arg Gln Arg Ala Ala Ala Leu Leu Ala Arg Asn Pro Asn Ala Ser Leu

180185190

Arg	Glu	Val	Thr	Arg	Ala	Ala	Gly	Ile	Ser	Val	Gly	Thr	Ala	Ser	Asp	
		195					200					205				
Val	Arg	Ala	Arg	Leu	Arg	Arg	Gly	Glu	Pro	Ala	Leu	Thr	Ala	Arg	Gln	
	210					215					220					
Gln	Ala	Val	Met	Lys	Leu	Arg	Pro	Ala	Ala	Arg	Asn	Pro	Asn	Ala	Ser	
225					230					235					240	
Leu	Arg	Glu	Val	Thr	Arg	Ala	Ala	Gly	Ile	Ser	Val	Gly	Thr	Ala	Ser	
				245					250					255		
Asp	Val	Arg	Ala	Arg	Leu	Arg	Arg	Gly	Glu	Pro	Ala	Leu	Thr	Ala	Arg	
			260					265					270			
Gln	Gln	Ala	Val	Met	Lys	Leu	Arg	Pro	Ala	Ala	Ala	Gln	Arg	Ser	Gly	
		275					280					285				
Pro	Asp	Tyr	Gly	Arg	Val	Leu	Glu	Asn	Leu	Arg	Lys	Asp	Pro	Ser	Leu	
	290					295					300					
Arg	Phe	Thr	Asp	Leu	Gly	Arg	Arg	Leu	Leu	Arg	Leu	Leu	Asp	Gly	Ser	
305					310					315					320	
Val	Pro	Gly	Ser	Val	Glu	Gln	Ile	Ala	Gln	Ile	Ala	Asp	Gly	Val	Pro	
				325					330					335		
Glu	His	Cys	Arg	Thr	Val	Val	Val	Asp	Met	Ala	Arg	Glu	Cys	Ala	Ala	
			340					345					350			
Ala	Trp	Gln	His	Leu	Ala	Asp	Gln	Leu	Ala	Asp	Arg	Asp	Thr	Ala		
		355					360					365				

```
<210> 31
<211> 359
<212> PRT
<213> S. globisporus
```

<400> 31

Met Lys Ser Asp Ser Ala Gln Arg Ala Val Glu Arg Ser Arg Arg Val
 1 5 10 15
 Val Arg Ile Asp Glu Leu Ile Pro Ala Asp Ser Pro Arg Leu Asn Gly
 20 25 30
 Ile Asp Arg Ser His Val Gln Arg Leu Ala Thr Val Tyr Ala Ser Leu
 35 40 45
 Pro Pro Val Leu Val His Arg Pro Thr Met Arg Val Val Asp Gly Met
 50 55 60
 His Arg Ile Gly Ala Ala Arg Leu Lys Gly Leu Asp Thr Val Glu Val
 65 70 75 80
 Thr Phe Phe Glu Gly Ala Glu Glu Gln Val Phe Leu Arg Ser Val Ala
 85 90 95
 Ala Asn Ile Thr Asn Gly Leu Pro Leu Ser Val Ala Asp Arg Lys Thr
 100 105 110
 Ala Ala Ala Arg Ile Leu Ala Ser His Pro Thr Leu Ser Asp Arg Ala
 115 120 125
 Val Ala Ala His Val Gly Leu Asp Ala Lys Thr Val Ala Gly Val Arg
 130 135 140
 Thr Cys Ser Ala Ala Gly Ser Pro Leu Leu Asn Met Arg Thr Gly Ala
 145 150 155 160
 Asp Gly Arg Val His Pro Leu Asp Arg Thr Ala Glu Arg Leu His Ala
 165 170 175
 Ala Ala Leu Leu Thr Gln Asp Pro Gly Leu Pro Leu Arg Ser Val Val
 180 185 190
 Glu Gln Thr Gly Leu Ser Leu Gly Thr Ala His Asp Val Arg Arg Arg
 195 200 205

Leu Leu Arg Gly Glu Asp Pro Val Pro Gln Asn Arg Gln Ser Ala Met
 210 215 220

Leu Glu Pro Gly Leu Ala Pro Gln Lys Lys Ala Thr Ala Lys Pro Pro
 225 230 235 240

Val Gly Pro Ala Ala Arg Pro Val Pro Lys Val Pro Pro Ala Val Ala
 245 250 255

Gly Arg Pro Pro Val Ser Pro Arg Ser Arg Ala Pro Leu Glu Ala Leu
 260 265 270

Arg Lys Leu Ser Asn Asp Pro Ser Leu Arg His Ser Asp Gln Gly Arg
 275 280 285

Glu Leu Met Arg Trp Leu His Asn Arg Phe Val Val Asp Glu Ala Trp
 290 295 300

Arg Arg Arg Ala Asp Ala Val Pro Ala His Cys Val Asp Ser Met Ala
 305 310 315 320

Glu Leu Ala Gln His Cys Ser Asp Ala Trp His Arg Phe Ala Glu Glu
 325 330 335

Met Val Arg Arg Arg His Ser Ala Ala Ala Asp Gly Ser Gly Leu Arg
 340 345 350

Thr Thr Gln Pro Thr Arg Arg
 355

<210> 32
 <211> 329
 <212> PRT
 <213> Actinoplanes teichomyceticus

<400> 32

Met Thr Pro Asp Glu Glu Ala Leu Asn Arg Gln Pro Ile Met Glu Met
 1 5 10 15

Glu Ile Ser Ser Leu Ser Leu Gly Gly Ser Pro Arg Leu Ala Gly Gly
 20 25 30

Asp Pro Val His Leu Glu Ala Met Val Ala Ala Gln Gly Glu Leu Pro
 35 40 45

Pro Ile Val Val His Arg Pro Thr Met Arg Val Ile Asp Gly Ser His
 50 55 60

Arg Ile Gln Ala Ala Leu Arg Arg Gly Glu Thr Thr Ile Ala Gly Arg
 65 70 75 80

Phe Phe Asp Gly Ser Asp Asp Glu Ala Phe Val Met Ser Val Trp Leu
 85 90 95

Asn Val Ser His Gly Leu Pro Leu Ala Leu Ala Asp Arg Lys Arg Ala
 100 105 110

Ala Glu Arg Ile Ala Val Ser His Pro Gln Trp Ser Asp Arg Arg Val
 115 120 125

Ala Ala Val Thr Gly Ile Ser Pro Ser Thr Val Ala Asp Ile Arg Arg
 130 135 140

Arg Val Ala Gly Thr Ser Ala Pro Glu Ala Ser Arg Ile Gly Gln Asp
 145 150 155 160

Gly Arg Val Arg Pro Leu Asp Cys Ser Ala Gly Arg Leu Leu Ala Gly
 165 170 175

Arg Leu Met Ala Glu Asn Pro Ala Leu Ser Leu Arg Gln Val Ala Lys
 180 185 190

Ala Ala Ala Ile Ser Pro Glu Thr Ala Arg Asp Val Arg Asn Arg Leu
 195 200 205

Leu Ser Gly Ala Glu Leu Val Pro Asn Arg Arg Pro Arg Asp Ala Ala
 210 215 220

Pro Val Gly Val Lys Gly Gly Arg Asp Arg Arg Pro Leu Asn Leu Ile
 225 230 235 240

Arg Ser Gly Asp Arg Pro Glu Pro Val Pro Asp His Ala Val Val Ile
 245 250 255

Asn Arg Leu Met Ser Asp Pro Ala Leu Arg Tyr Thr Asp Thr Gly Arg
 260 265 270

Asn Leu Leu Arg Leu Leu Ser Leu His Thr Arg Trp Ala Lys Glu Trp
 275 280 285

Glu Ala Ile Val Asp Asn Leu Pro Pro His Cys Ala Asp Ala Val Ala
 290 295 300

Asp Leu Ala Arg Gln Phe Ala Asp Leu Trp Ala Asp Phe Ala Ser Arg
 305 310 315 320

Val Gly Pro Glu Glu Arg Met Ala Ser
 325

<210> 33
 <211> 221
 <212> PRT
 <213> S. coelicolor

<400> 33

Met Thr Ile Arg Leu Leu Ile Val Asp Asp Gln Glu Leu Ile Arg Thr
 1 5 10 15

Gly Phe Arg Leu Phe Leu Gln Thr Gln Asn Asp Leu Glu Val Val Gly
 20 25 30

Glu Ala Asp Asp Gly His Gly Ala Leu Ala Gln Ala Ala Ala Leu Arg
 35 40 45

Pro Asp Val Val Leu Met Asp Ile Arg Met Pro Arg Met Asp Gly Val
 50 55 60

Glu Ala Thr Ser Arg Leu Thr Ala Ser Asp Ser Pro Pro Arg Val Leu
 65 70 75 80

Ile Leu Thr Thr Tyr Asp Leu Asp Glu Tyr Val Phe Gly Ala Leu Arg

85 90 95
 Ala Gly Ala Ser Gly Phe Leu Leu Lys Asp Ala Ser Arg Asp Arg Leu
 100 105 110
 Leu Glu Ala Ile Arg Val Val His Ala Gly Glu Ala Leu Leu Ser Pro
 115 120 125
 Ser Ile Thr Arg Arg Leu Ile Glu Asp Tyr Ala Thr Arg Ala Ala Pro
 130 135 140
 Val Arg Pro Arg Glu Ala Val Leu Ala Gly Leu Thr Pro Arg Glu Arg
 145 150 155 160
 Glu Ile Leu Leu Leu Val Ala Arg Gly Leu Ser Asn Pro Glu Ile Ala
 165 170 175
 Ala Arg Leu Val Val Thr Glu Ala Thr Val Lys Ser His Val Gly Ser
 180 185 190
 Met Phe Ala Lys Leu His Leu Arg Asp Arg Ala Gln Ala Val Val Phe
 195 200 205
 Ala Tyr Glu Asn Ala Ile Val Leu Pro Gly Gly Thr Gly
 210 215 220
 <210> 34
 <211> 222
 <212> PRT
 <213> S. coelicolor
 <400> 34
 Met Ile Arg Val Leu Leu Ala Asp Asp Glu Thr Ile Ile Arg Ala Gly
 1 5 10 15
 Val Arg Ser Ile Leu Thr Thr Glu Pro Gly Ile Glu Val Val Ala Glu
 20 25 30
 Ala Ser Asp Gly Arg Glu Ala Val Glu Leu Ala Arg Lys His Arg Pro
 35 40 45

Asp Val Ala Leu Leu Asp Ile Arg Met Pro Glu Met Asp Gly Leu Thr
 50 55 60

Ala Ala Gly Glu Met Arg Thr Thr Asn Pro Asp Thr Ala Val Val Val
 65 70 75 80

Leu Thr Thr Phe Gly Glu Asp Arg Tyr Ile Glu Arg Ala Leu Asp Gln
 85 90 95

Gly Val Ala Gly Phe Leu Leu Lys Ala Ser Asp Pro Arg Asp Leu Ile
 100 105 110

Ser Gly Val Arg Ala Val Ala Ser Gly Gly Ser Cys Leu Ser Pro Leu
 115 120 125

Val Ala Arg Arg Leu Met Thr Glu Leu Arg Arg Ala Pro Ser Pro Arg
 130 135 140

Ser Glu Val Ser Gly Glu Arg Thr Thr Leu Leu Thr Lys Arg Glu Gln
 145 150 155 160

Glu Val Leu Gly Met Leu Gly Ala Gly Leu Ser Asn Ala Glu Ile Ala
 165 170 175

Gln Arg Leu His Leu Val Glu Gly Thr Ile Lys Thr Tyr Val Ser Ala
 180 185 190

Ile Phe Thr Gln Leu Glu Val Arg Asn Arg Val Gln Ala Ala Ile Ile
 195 200 205

Ala Tyr Glu Ala Gly Leu Val Lys Asp Ala Asp Leu Asn Arg
 210 215 220

<210> 35
 <211> 218
 <212> PRT
 <213> S. coelicolor

<400> 35

Met Arg Glu Asp Gly Lys Ile Arg Val Phe Leu Leu Asp Asp His Glu

1	5	10	15
Val Val Arg Arg Gly Val His Asp Leu Leu Ser Gly Glu Ala Asp Ile	20	25	30
Glu Val Val Gly Glu Ala Gly Thr Ala Ala Glu Ala Gln Ala Arg Val	35	40	45
Thr Ala Thr Arg Pro Asp Val Ala Val Leu Asp Val Arg Leu Pro Asp	50	55	60
Gly Ser Gly Val Glu Val Cys Arg Asp Ile Arg Ser Arg Asp Glu Ser	65	70	75
Val Arg Cys Leu Met Leu Thr Ser Phe Ala Asp Asp Glu Ala Leu Phe	85	90	95
Asp Ala Ile Met Ala Gly Ala Ser Gly Tyr Val Leu Lys Asp Ile Arg	100	105	110
Gly Ala Glu Leu Leu Gly Ala Val Arg Glu Val Ala Ala Gly Lys Ser	115	120	125
Leu Leu Asp Pro Ala Ala Thr Ala Arg Val Leu Glu Arg Leu Arg Gly	130	135	140
Gly Gly Ala Arg Pro Asp Asp Arg Leu Ala Arg Leu Thr Glu Gln Glu	145	150	155
Arg Arg Ile Leu Glu Leu Ile Gly Glu Gly Leu Thr Asn Arg Ala Ile	165	170	175
Gly Glu Arg Leu His Leu Ala Glu Lys Thr Ile Lys Asn Tyr Val Ser	180	185	190
Ser Leu Leu Gly Lys Leu Gly Met Gln Arg Arg Ser Gln Ala Ala Ala	195	200	205
Phe Val Ala Arg Leu Glu Ala Glu Asn Arg	210	215	

<210> 36
 <211> 220
 <212> PRT
 <213> Streptomyces fungicidicus

<400> 36

Val Ser Val Leu Leu Glu Gln Pro Ala Ser Leu Val Ala Tyr Arg Pro
 1 5 10 15

Asn Lys Pro Thr Ala Met Val Val Val Ala Asp Pro Arg Val Arg Ser
 20 25 30

Thr Val Thr Arg His Leu Trp Ala Leu Gly Val Arg Asp Val Ile Glu
 35 40 45

Ala Ser Ser Val Ala Glu Ala Arg Pro Arg Ile Gly Asn Pro Arg Asp
 50 55 60

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

Leu Ser Glu Thr Arg Ala Ala Gly Trp Pro Asn Gly Leu Ala Leu Ser
 85 90 95

Ala Ala Asp Asp Ile Gly Ala Val Arg Asn Ala Leu Ala Gly Gly Val
 100 105 110

Lys Gly Tyr Val Val Thr Gly Thr Arg Thr Asn Leu Gly Leu Pro Thr
 115 120 125

Arg Pro Gly Ala Ala Pro Ile Gly Ala Ala Ala Arg Leu His Arg
 130 135 140

Arg Pro Pro Gly Ala Pro Ser His Pro Gly Gly Tyr Arg Glu Leu Ser
 145 150 155 160

Gly Arg Glu Val Glu Val Leu Arg Leu Val Ala Glu Gly Gln Ser Asn
 165 170 175

Lys Ala Ile Gly Val Ser Met Gly Leu Ser Ala Leu Thr Val Lys Ser
 180 185 190

His Leu Ala Arg Ile Ala Arg Lys Leu Gly Thr Gly Asp Arg Ala Gly
 195 200 205

Met Val Ala Val Ala Leu Arg Thr Gly Ile Ile His
 210 215 220

<210> 37
 <211> 663
 <212> ADN
 <213> Streptomyces fungicidicus

<400> 37
 gtgtccgttc tcctcgagca gcccgaagc ctggtcgcct accgcccga caagccgacc 60
 gccatggtgg tcgtggccga ccccgcgctc cgttcgaccg tcacccgcca cctgtgggcg 120
 ctcggcgtac gcgacgtcat cgaggcctcg tccgtcgcgg aggctcgtcc ccgcatcggc 180
 aacccccgcg acatctgcgt cgccgaagtc catctgccgg atggttccgg cctcaccctc 240
 ctctccgaga cccgcgccgc gggctggccc aacggcctcg ccctctccgc ggcggacgac 300
 atcggcgccg tgcgcaacgc cctcgcgggc ggagtcaagg gctacgtcgt caccggcacc 360
 cgcaccaacc tcgggctccc caccggcccg ggtgccgctc ccatcggcgc cgccgccgcg 420
 cgcttcgacc gccgcccccc gggtgccccg agccaccgga gcggctaccg cgagctgtcc 480
 ggccgcgagg tggaggtgct gcggctggtg gcggaaggcc agtcgaacaa ggcgatcggc 540
 gtctcgaatg gcctgtccgc actgaccgtc aagagccacc tggcccggat cggccgcaag 600
 ctcggcacgg gcgaccgcgc cggcatggtg gccgtggccc tgcgcaccgg catcatccac 660
 tga 663

<210> 38
 <211> 936
 <212> ADN
 <213> Streptomyces fungicidicus

<400> 38
 gtggaaataa gttcgtcttc caccgacggc tccccgcgga tcgacgggga gaggccccgag 60
 cacgtggaaa tgctggccgc cgccgacacc gcgcttccac cgatcatggt gcaccgccgc 120

accgggcggg tcacgcacgg catgcaccgg ctgcgcgccg cgatgctgac gggccgtacg 180
 acgatcgagg tgaggttctt cgacggcacc gaggaggacg ctttcgtcct cgccgtgaag 240
 tcgaacatcg cgcacggact gccgctgtcc gccgccgacc gccggcgggc cgccgggcgc 300
 atcatggcca cccatccccg gtggtcggac cggatgatcg cctcgggtgtt cggcacctcc 360
 gccaggacgg tcgccgagat ccgccgcgac gccggcgccg ccggggcggg ggagcccacc 420
 cgcacgcggc gggacggcag ggtacggccc gtcgacgtga gcgaggggcg cagactggcc 480
 cacgacatga tcgtccgcga cccgggcctg tcgtgcgcc aggtcgcccg cgccgccggg 540
 atctcgccgg agaccgtcag ggacgtcaga caccggatgc tccgcggtga ggacccggtg 600
 cccgcgccgc ggccgcggac cctggtggag cgcggcgccg accgccgggc ggagccggcc 660
 gggaaggccg ccgcgccgtg cgggacggag ccgccgccg ccgtcgtgat gaagcggctg 720
 agggccgacg cggcgctgcg tctcaacgag aacggacgcg acctgctgcg gcttctggat 780
 atccacacgg tccggctgga ggactggaac cgcattatcg aaagcgtgcc gccgcaccgt 840
 ctggagacgg tggcgagctt ggcacgtcc tgcgccgaca aatggtccga gatcgcgctc 900
 cgcacgaaa gcaacgcac acatctggcc ggggtga 936

<210> 39
 <211> 39331
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Tổng hợp fosmid pXYF148

<400> 39
 atcgaatata acttcgtata atgtatgcta tacgaagtta ttagcgatga gctcggactt 60
 ccattgttca ttccacggac aaaaacagag aaaggaaacg acagaggcca aaaagctcgc 120
 tttcagcacc tgtcgtttcc tttcttttca gagggtattt taaataaaaa cattaagtta 180
 tgacgaagaa gaacggaaac gccttaaacc ggaaaatttt cataaatagc gaaaacccgc 240
 gaggtcgccg ccccgtaacc tgtcggatca ccggaaagga cccgtaaagt gataatgatt 300
 atcatctaca tatcacaacg tgcgtggagg ccatcaaacc acgtcaaata atcaattatg 360
 acgcaggat cgtattaatt gatctgcatc aacttaacgt aaaaacaact tcagacaata 420
 caaatcagcg aactgaata cggggcaacc tcatgtccga gtcgcgagc tcgtcgacag 480

cgacacactt gcatcggatg cagcccggtt aacgtgccgg cacggcctgg gtaaccaggt	540
at ttgtcca cataaccgtg cgcaaaatgt tgtggataag caggacacag cagcaatcca	600
cagcaggcat acaaccgcac accgaggtta ctccgttcta caggttacga cgacatgtca	660
atacttgccc ttgacaggca ttgatggaat cgtagtctca cgctgatagt ctgatcgaca	720
atacaagtgg gaccgtggtc ccagaccgat aatcagaccg acaacacgag tgggatcgtg	780
gtcccagact aataatcaga ccgacgatac gagtgggacc gtggtcccag actaataatc	840
agaccgacga tacgagtggg accgtggttc cagactaata atcagaccga cgatacgagt	900
gggaccgtgg tcccagacta ataatacagac cgacgatacg agtgggacca tgggtcccaga	960
ctaataatca gaccgacgat acgagtggga ccgtgggtccc agtctgatta tcagaccgac	1020
gatacgagtg ggaccgtggt cccagactaa taatcagacc gacgatacga gtgggaccgt	1080
ggtcccagac taataatcag accgacgata cgagtgggac cgtgggtccca gtctgattat	1140
cagaccgacg atacaagtgg aacagtgggc ccagagagaa tattcaggcc agttatgctt	1200
tctggcctgt aacaaaggac attaatgaaa gacagataaa cgtagactaa aacgtggctg	1260
catcagggtg ctggcttttc aagttcctta agaatggcct caat tttctc tatacactca	1320
gttggaacac gagacctgtc caggttaagc accat tttat cgcccttata caatactgtc	1380
gctccaggag caaactgatg tcgtgagctt aaactagttc ttgatgcaga tgacgtttta	1440
agcacagaag ttaaaagagt gataacttct tcagcttcaa atatcacccc agct tttttc	1500
tgctcatgaa ggtagatgc ctgctgctta agtaattcct ctttatctgt aaaggctttt	1560
tgaagtgcac cacctgaccg ggcagatagt tcaccggggt gagaaaaaag agcaacaact	1620
gatttaggca atttggcggg gttgatacag cgggtaataa tcttacgtga aatattttcc	1680
gcatcagcca gcgcagaaat atttccagca aattcattct gcaatcggct tgcataacgc	1740
tgaccacgtt cataagcact tgttgggcga taatcgttac ccaatctgga taatgcagcc	1800
atctgctcat catccagctc gcccaaccaga acacgataat cactttcggg aagtgcagca	1860
gctttacgac ggcgactccc atcggcaatt tctatgacac cagatactct tcgaccgaac	1920
gccggtgtct gttgaccagt cagtagaaaa gaagggatga gatcatccag tgcgtcctca	1980
gtaagcagct cctgggtcacg ttcattacct gaccataccc gagaggtctt ctcaacacta	2040

tcaccccgga gcacttcaag agtaaacttc acatcccgcac cacatacagg caaagtaatg	2100
gcattaccgc gagccattac tcctacgcgc gcaattaacg aatccaccat cggggcagct	2160
ggtgtcgata acgaagtatc ttcaaccggt tgagtattga gcgtatgttt tggaataaca	2220
ggcgcacgct tcattatcta atctcccagc gtggtttaat cagacgatcg aaaatttcat	2280
tgcagacagg ttcccaaata gaaagagcat ttctccaggc accagttgaa gagcgttgat	2340
caatggcctg ttcaaaaaca gttctcatcc ggatctgacc ttaccaact tcatccgttt	2400
cacgtacaac attttttaga accatgcttc cccaggcatc ccgaatttgc tcctccatcc	2460
acggggactg agagccatta ctattgctgt atttggttaag caaaatacgt acatcaggct	2520
cgaacccttt aagatcaacg ttcttgagca gatcacgaag catatcgaaa aactgcagtg	2580
cggaggtgta gtcaaacac tcagcaggcg tgggaacaat cagcacatca gcagcacata	2640
cgacattaat cgtgccgata cccaggttag gcgcgctgtc aataactatg acatcatagt	2700
catgagcaac agtttcaatg gccagtcgga gcatcagggtg tggatcgggtg ggcagtttac	2760
cttcatcaaa ttgtcccatt aactcagttt caatacgggtg cagagccaga caggaaggaa	2820
taatgtcaag ccccggccag caagtgggct ttattgcata agtgacatcg tccttttccc	2880
caagatagaa aggcaggaga gtgtcttctg catgaatatg aagatctggt acccatccgt	2940
gatacattga ggctgtttcc tgggggtcgt taccttccac gagcaaaaaca cgtagcccct	3000
tcagagccag atcctgagca agatgaacag aaactgaggt tttgtaaacg ccacctttat	3060
gggcagcaac cccgatcacc ggtggaaata cgtcttcagc acgtcgcaat cgcgtaccaa	3120
acacatcacg catatgatta atttgttcaa ttgtataacc aacacgttgc tcaaccgctc	3180
ctcgaatttc catatccggg tgcggtagtc gccctgcttt ctcggcatct ctgatagcct	3240
gagaagaaac cccaactaaa tccgtgctt cacctattct ccagcgccgg gttattttcc	3300
tcgcttccgg gctgtcatca ttaactgtg caatggcgat agccttcgtc atttcatgac	3360
cagcgtttat gactgggtta agtgtttcca tgagtttcat tctgaacatc ctttaatcat	3420
tgctttgcgt ttttttatta aatcttgcaa ttactgcaa agcaacaaca aaatcgcaaa	3480
gtcatcaaaa aaccgcaaag ttgtttaaaa taagagcaac actacaaaag gagataagaa	3540
gagcacatac ctcagtcact tattatcact agcgctcgcc gcagccgtgt aaccgagcat	3600
agcgagcgaa ctggcgagga agcaaagaag aactgttctg tcagatagct cttacgctca	3660

gcgcaagaag aaatatccac cgtgggaaaa actccaggta gaggtacaca cgcggatagc	3720
caattcagag taataaaactg tgataatcaa ccctcatcaa tgatgacgaa ctaacccccg	3780
atatcaggtc acatgacgaa gggaaagaga aggaaatcaa ctgtgacaaa ctgccctcaa	3840
atttggcttc cttaaaaatt acagttcaaa aagtatgaga aaatccatgc aggctgaagg	3900
aaacagcaaa actgtgacaa attaccctca gtaggtcaga acaaattgtga cgaaccaccc	3960
tcaaatctgt gacagataac cctcagacta tcctgtcgtc atggaagtga tatcgcggaa	4020
ggaaaatacg atatgagtcg tctggcggcc tttctttttc tcaatgtatg agaggcgc	4080
tggagtcttg ctgttgatct cattaacaca gacctgcagg aagcggcggc ggaagtcagg	4140
catacgctgg taactttgag gcagctggta acgctctatg atccagtcga ttttcagaga	4200
gacgatgcct gagccatccg gcttacgata ctgacacagg gattcgtata aacgcatggc	4260
atacggattg gtgatttctt ttgtttcact aagccgaaac tgcgtaaacc ggttctgtaa	4320
cccgataaag aagggaatga gatatgggtt gatatgtaca ctgtaaagcc ctctggatgg	4380
actgtgcgca cgtttgataa accaaggaaa agattcatag cttttttcat cgccggcatc	4440
ctcttcaggg cgataaaaaa ccacttcctt ccccgcgaaa ctcttcaatg cctgccgtat	4500
atccttactg gcttccgcag aggtcaatcc gaatatttca gcatatttag caacatggat	4560
ctcgcagata ccgtcatgtt cctgtagggt gccatcagat tttctgatct ggtcaacgaa	4620
cagatacagc atacgttttt gatcccgga gagactatat gccgcctcag tgaggctggt	4680
tgactggacg attcgcgggc tattttttacg tttcttgtga ttgataaccg ctgtttccgc	4740
catgacagat ccatgtgaag tgtgacaagt ttttagattg tcacactaaa taaaaaagag	4800
tcaataagca gggataactt tgtgaaaaa cagcttcttc tgagggaat ttgtcacagg	4860
gttaagggca atttgtcaca gacaggactg tcatttgagg gtgatttgtc aactgaaag	4920
ggcaatttgt cacaacacct tctctagaac cagcatggat aaaggcctac aaggcgtct	4980
aaaaaagaag atctaaaaac tataaaaaaa ataattataa aaatatcccc gtggataagt	5040
ggataacccc aagggaagtt ttttcaggca tcgtgtgtaa gcagaatata taagtgtgt	5100
tccctggtgc ttctcgctc actcgaccgg gagggttcga gaaggggggg caccctt	5160
cggcgtgcgc ggtcacgcgc acagggcgca gccctgggta aaaacaaggt ttataaatat	5220

tggtttaaaa gcaggttaaa agacaggtta gcggtggccg aaaaacgggc ggaaaccctt	5280
gcaaagtctg gatcttctgc ctgtggacag cccctcaaat gtcaataggt gcgcccctca	5340
tctgtcagca ctctgcccct caagtgtcaa ggatcgcgcc cctcatctgt cagtagtcgc	5400
gcccctcaag tgtcaatacc gcagggcact tatccccagg cttgtccaca tcattctgtg	5460
gaaactcgcg taaaatcagg cgttttcgcc gatttgcgag gctggccagc tccacgtcgc	5520
cggccgaaat cgagcctgcc cctcatctgt caacgccgcg ccgggtgagt cggcccctca	5580
agtgtcaacg tccgcccctc atctgtcagt gagggccaag ttttccgcga ggtatccaca	5640
acgccggcgg ccggccgcgg tgtctcgcac acggcttcga cggcgtttct ggcgcgtttg	5700
cagggccata gacggccgcc agcccagcgg cgagggcaac cagccgaggg cttcgccctg	5760
tcgctcgact gcggcgagca ctactggctg taaaaggaca gaccacatca tggttctgtg	5820
ttcattaggt tgttctgtcc attgctgaca taatccgctc cacttcaacg taacaccgca	5880
cgaagatttc tattgttcct gaaggcatat tcaaatcggt ttcgttaccg cttgcaggca	5940
tcattgacaga acactacttc ctataaacgc tacacaggct cctgagatta ataatgcgga	6000
tctctacgat aatgggagat tttcccgact gtttcgttcg cttctcagtg gataacagcc	6060
agcttctctg ttttaacagac aaaaacagca tatccactca gttccacatt tccatataaa	6120
ggccaaggca tttattctca ggataattgt ttcagcatcg caaccgcatc agactccggc	6180
atcgcaaact gcacccggtg ccgggcagcc acatccagcg caaaaacctt cgtgtagact	6240
tccgttgaac tgatggactt atgtcccatc aggctttgca gaactttcag cggatataccg	6300
gcatacagca tgtgcatcgc ataggaatgg cggaacgtat gtgggtgtgac cggaacagag	6360
aacgtcacac cgtcagcagc agcggcggca accgcctccc caatccaggt cctgaccgtt	6420
ctgtccgtca cttcccagat ccgcgctttc tctgtccttc ctgtgcgacg gttacgccgc	6480
tccatgagct tatcgcgaaat aaatacctgt gacggaagat cacttcgcag aataaataaa	6540
tcctgggtgtc cctgttgata ccgggaagcc ctgggccaac ttttggcgaa aatgagacgt	6600
tgatcggcac gtaagagggt ccaactttca ccataatgaa ataagatcac taccgggcgt	6660
atTTTTTgag ttatcgagat tttcaggagc taaggaagct aaaatggaga aaaaaatcac	6720
tggatatacc accgttgata tatcccaatg gcatcgtaaa gaacattttg aggcatTTTca	6780
gtcagttgct caatgtacct ataaccagac cgttcagctg gatattacgg cTTTTTaaa	6840

gaccgtaaag aaaaataagc acaagtttta tccggccttt attcacattc ttgccgcct	6900
gatgaatgct catccggaat ttcgtatggc aatgaaagac ggtgagctgg tgatatggga	6960
tagtgttcac ccttgttaca ccgttttcca tgagcaaact gaaacgtttt catcgctctg	7020
gagtgaatac cacgacgatt tccggcagtt tctacacata tattcgcaag atgtggcgtg	7080
ttacggtgaa aacctggcct atttcctaa agggtttatt gagaatatgt ttttcgtctc	7140
agccaatccc tgggtgagtt tcaccagttt tgatttaaac gtggccaata tggacaactt	7200
cttcgcccc gttttcacca tgggcaaata ttatacgcaa ggcgacaagg tgctgatgcc	7260
gctggcgatt caggttcac atgccgtttg tgatggcttc catgtcggca gaatgcttaa	7320
tgaattacaa cagtactgcg atgagtggca gggcggggcg taatTTTTTT aaggcagtta	7380
ttggtgccct taaacgcctg gttgctacgc ctgaataagt gataataagc ggatgaatgg	7440
cagaaattcg atgataagct gtcaaacatg agaattggtc gacggcccg gcgccgcaa	7500
ggggttcgcg ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag	7560
cgggcagtg gcgcaacgca attaatgtga gttagctcac tcattaggca cccaggtt	7620
tacactttat gttccggct cgtatgttgt gtggaattgt gagcggataa caatttcaca	7680
caggaaacag ctatgaccat gattacgcca agctatttag gtgagactat agaatactca	7740
agcttgcatg cctgcaggtc gactctagag gatccacca ccgccggacg ggggtcgccg	7800
gcccggaagt ccaccaggat cgagccgccg atctgccgcc cgcgccacca cgactccatc	7860
agcccgacct cctcctccgg gctgcgcacc agccgcagca gcagcgagca ggaccgctcc	7920
accaggacgc tctccacccc ggagatgaac tccatgtaga acgggtccag gccgagcagc	7980
cgggcgggcc ggcagaccgc gagaccacc acgtccaccc gcgaccggc cagcgtgcgc	8040
gcggttcggc tcggcgccca cccagctcc cgcgccgcc ggaagatgcg gtcccgggtc	8100
gcctccgaca gcccgggctt ccggttgaag gcgagggaca cggcgccctt ggacacgccg	8160
gcgcgcgcgg cgacgtccct gatggtgacg cgaggggtcg gcgttgccgt catcgagtgg	8220
gctccacgca gtacaggcg gaacgggcgg tgtccgggtc cggagtcttc cccccgca	8280
ccccgatggc cacctgttcc ccggggagca gggtcaccag ccccggtcg gccgcgccc	8340
cggggtccag ccggtcggcc tggagcagca ggtcccgtac gaggggtgcgg gccgtgaccg	8400

tgatcccgtc cggcgcgagg gcgacctga actccggcgg ggggtagggg atctcccgtt	8460
ccggcgccgg gaagtgccac gcccgcaccc cgtccgcgtc ggcgaccagg aactccccgg	8520
ggccgtccgg cagcagttcg accgggacct cgaccacggc caccgtccgc cccccggcgt	8580
ccagcgccgg ggccgcctcc gcgatcgggg cgccgtcgac ggacatccgg cgcagccgca	8640
gcgttccccg ccagtcctcc gcggactggt tgaccgccgc caccaccaga ccgtcaccgt	8700
ccgcgcgcac ggtcagcagc cggtcgcgt acagccggcg cagctcgtgg tagagcggct	8760
tctcccgccc gtccccgtcg atcgcgggcc acgacgtcac cggccagcag tcgttgagct	8820
gccagaccac cgtgcccgcg cacaccggcc agtgcgagcg ccagtgtctg acaccggccg	8880
ccaccgcacg cgcctggttg acctgcgtca gatagtcca gcggtcgaag tcgccctccg	8940
gcacggcgaa gtggcggggc aggccgcgt ccagcttgcc gttgccgtcc tccgccttct	9000
ggtggtgcag catgccgggg gagtccggcg cggggtcctc cccgggcagc gcccgccgca	9060
gcgtggcgtg cgcgggaggc gcctgccagc cgaactcggc cacgaagcgc gggacgtcgc	9120
gccggtagtc ggcgtagtcg gcgcggttcc acacctcca ggagtgggtg gtgccgtgcg	9180
ccggatcgtt ggggtggtgc cgccaggaac cggaccaggg actgcccgcg gtgtacggcc	9240
gcgtcgggtc cagctccgcg accaccgcg gcaggacgcc gaggtagtag ccctcgcccc	9300
aggagtcccc ggcgagcccc tgctcccagt cccagtcccc gaacccccac aggttctcgt	9360
tgttgccgtt ccacagcacc agggaggggt gcggcatcag ccgtacgacg ttctcccggg	9420
cctccgcctc cacctccccg cgcagcggct gctcctcggg gtaggcggcg cacgcgaacg	9480
ggaagtcctg ccagaccagc agccccaact cgtcgcaggc gtcgtagaag tcctcgtcct	9540
cgtagatccc gccgccccag acccggacca ggtccacccc cgcgccggcc gcctgctcca	9600
gccggtgccg gtagcgctcc cgggtgatcc gggacgggaa cacgtcgtcc gggatccagt	9660
tgacgccccg cgcgaacagc cgctcaccgt tgacgaccag ggtgaacccg gtgccgtgcg	9720
cgtcggccga ggtgtccagc tcaaccgtcc ggaaccgggt cctgcgccgc caggcgtcca	9780
gcgcctcgtc accgtgggac aacgtcagct cgacgtcgta cagcggctgt tcgccgtatc	9840
cgcgcgggcca ccacaggcgg acgtccggca cccggagccg cacgggtccc gccgtcccat	9900
cgacccgcgc ccgggcgcgc acgccccgg cgctgcctc cagggtgagc ggtgcctcga	9960
cccgggagcg ctccacgtcg accgccagct cgatctgccc caccctgtcc tcgacggtga	10020

ccagcgggcg caccgggcg atccgcgccg tcgaccagcg ctccagccgc accggccgcc	10080
agatcccggc cgtcaccagc gtcggccccc agtcccagcc gaacgagcag gccatcttcc	10140
gcaggtactg gtacggctcg gcgtacgctc cggggcgctc gccagcctg ccgcgcaccg	10200
cctccgcctc ggcgtagcg gaggcgaacc gcaccgtgag ccggccgctc agtcccgtca	10260
cgtcgaagcg gtacgagcgg tgcatgttcc gcgtccggcc cagtggccgg ccgtcgagca	10320
ggatctcggc gacggtgtcg agaccgtcga agacgaggtc cgtctgctcg tgcgggcccg	10380
tcccggcggt cagctccgtc tcgtacgtcc actcccgcg gccacccag gccacctcgg	10440
tctcgttgcg gccgaggaac ggatcgggga tcagcccggc cgccagcaga tcggtgtgca	10500
cacaccccgg caccgaggcg gggagggcgt ccccggtgcc gtccgggtgt cgcaggatcc	10560
atccctcggg gagcgggtgtg acctgacgca tgcacactcc ctaaaccggg tgagccttct	10620
ctgaagagtg gtctggcatc gttggcgcgga ttgcgacttt accggttcag ttcagggtctg	10680
ccagagtgcc gaatcagcca tcccactcgt gctcgtccgt cccgtgaacg gagccgtgat	10740
gcatctgaac cgccgtacga cactcaccgg atcgctcgcc ctgctcgccc tcctggcctc	10800
cgcctgcacg ggcacggggg gttcctcgaa gggcgcgggac gccaaggctc ccgacgaccc	10860
gtcaaaggctc aaggggtccc tcacggctct caccacccgg accgatctgg tgcaggacgg	10920
gacgatgaag aagtacccg ccgagttcaa cgagacctat cccggggtga aggtggagtt	10980
cgacggcctc accgactacg agggcgagggt caagatccgt atgaacacgg agaactacgg	11040
cgacgtcctc atgatcccgg cggtcgtcga gaagaaggac tacccgaagt tcttcgcctc	11100
cctgggcacc aaggccgaac gcgccgcaa gtaccgggtc accgactact ccaccgtcga	11160
cggcaaggctc tacgggcaga gccccgtcgg cgtcgtcccc gggttcattc acaacaagcg	11220
ggtgtggagc gaggccggcg tcaccgactg gccaccacc cccgccgagt tcctggacga	11280
cctgaaggcg atccggtcga agaccgacgc ggtgccgtac tacaccaact tcaaggacat	11340
gtggccgctg acccagtgga ccaacgtcaa cggtccgctc ggctgcgacc cgcacgccac	11400
cacgaagctc gccgagggcg acccggtggc cgagggggcc gacctgcgcg tgggcgacac	11460
cctgctccac gacatcgtgc gcggcggact cgccgagaag gaccgacca ccaccaactg	11520
ggagggtcc aagcccaagc tggccaaggg cgagatcgcc accatgtggc tgggctcctg	11580

ggccgtcgtg cagatgcggg acgcggcgaa gcaggccggc gccgaccccg ccgacatcgg	11640
cttcatgccc ttccccgcac agcgggacgg cacgttctgc gcggtgacct ccccgacta	11700
ccagcaggcg gtcaacgtca actccgacaa caaggaggcc gcccgcgctt ggatcgactg	11760
gttcaccgac aagtccggct acgccgaggc caacctcgcc ctatcccccc tgaaggacgc	11820
cccgtgccc gccgtcctcg agccctacga gaaggccggc gtgaagctcc tggacctcga	11880
ggacagcaag ggcgccgagg tgaagtccct cgacaaccgc tccgaggctg gcatctacaa	11940
gcccgactac cgccaggaac tcgtcgacct cgcccgcggc gcccgcaagg gcggcctgga	12000
cgactacctc ggcggcctcg gcgagcgctg ggccgaggcg cgcagcgcg cgggggcctg	12060
atgacggaca ccacccgcaa ggcggcgcg gcggttcccc cggcccgcg ccggggccg	12120
ggcccgcgcg cgccggcccc gcgccgcacc cggctgtcgc gccgcctcac cccgtggctg	12180
ttcctggccg caccgtggc cctgctcctg accttcacct acgcgcccga tcgccaacat	12240
ggtcgcgtac agcttcaccg actgggacgg cgtgagcccg gagctgaact ggacgggcac	12300
cgggaactac accgaactcc tcacccgctc cgagctgttc gaggtcttct tcgtcagcgg	12360
ctactacctc gtcgcctccg cggcgcagat cgtgctcgcc ctctacttcg ccacggctct	12420
cagcttcgac gtccgcttcc ggaacttctt caaggcgctg ctgttcttcc cgtacctcat	12480
caacgggggtg gccatcggct tcgtcttcct ctacttcttc caggacggcg gcaccctcga	12540
ctccgtactg ggctgtctcg gcgtcgagac cgaccacgcc tggctgggca cgccgttctc	12600
cgcgaacacc tcgtggccg gcgtctccgt ctggcgctac ctcggactga acttcgtcct	12660
cttcctcggc gcgatccagt ccatcccggg cgagctgtac gaggcggccg agatcgacgg	12720
cgcgaaccgc tggcagcagt tccggcacat catcgcgccc ggcatcagac ccgtgctgag	12780
cctgagcgtg atcctctcgg tctccggctc gctgtcggtc ttcgagatcc cgtacatcat	12840
gaccggcggc gccaccggca cggagacctt cgtgatccag accgtgaagc tggcgttcca	12900
gttcaacaag acgggactcg cctcggccgc cgccgtcgtc ctgctgctga tcgtcctggc	12960
ggtcacctgg gtgcagcggc gcatcgctcc cgacgagaag gtggacctcg tatgaccgcg	13020
cgtaccgcgg cacgcgccct ggtcctgacg tccctgatcc tggcgacgct ggtggtgctg	13080
ctgccgctcg ccgtggtctt cctgacctcg ctgaagtcct ccgaggagat ggcgaacggc	13140
agcggagcgc tgacgccgcc cgacgacccg ctgaacttcg gcaactacgt gacggcgctc	13200

cgggacggcc agatgctgtc cgcgttcggg aacacggccg tcatcctggg cgtggccgtc	13260
ggcggaacga tcctgatcgg ctcgatgacg gcgtacgcga tcgaccgctt ccggttcgcg	13320
ttcaagaagc tggtcgtggc gctgttcctg ctggccgcgc tgggtccccg ggtgaccacc	13380
caggtggcga cttccagat cgtcaacagc ttcggcatgt tcgacagcct gtgggcgcgc	13440
atcgccctct acatgggcac ggacatcgtc tcgatctacg tcttcctgca gttcatccgc	13500
tccatccccg tctccctgga cgaggcggcg cgcctggacg gcgccaacgc gttcaccgtc	13560
taccgcaagg tgatcttccc gctgctcaag ccggcgatcg cgacgggtggg gatcgtaaag	13620
gggatcaacg tctacaacga cttctacatc cccttcctct acatgccctc cgaggacctg	13680
gggggtcatct cgacgtccct gttccgcttc aaggggcccct tcggcgcgca ctgggagacg	13740
atctcggcgg gcgcggctct ggtcatcctg cccaccttga tcgtcttcct gttcctccag	13800
cgcttcatct acaacgggtt catgcggggg gcgacgaagt agccagcgcg gccaccagca	13860
cgggagtgac ctggtcgacc tgccacgcgc gtgccccgtg cgccgtcagc gccgcggcga	13920
cggaccgctc gtcggggccc ggcggtctcc agcagaccct gcgcaccgtg tccgggggtga	13980
tcaggttctc cggcggcatg ttcagccgct cggcgagttc ggcgaccccc gcgcggggccg	14040
ccgacagccg ggccgcggca acggggctct tgtccgcca ggcgcgcggc ggcgaggggc	14100
cggtcaccgg ctggccgggc tgcggcagct gggcctcgct cagcgccttc gcgcggtcga	14160
cggccgcctg cactgctcc agctggcgcc gccccaccg ctgcccgaac ccgttgagcg	14220
cggccatggc gtgcaggttg gcgggcagcg cgagcgcggc ctccacgac gccgcgtcgg	14280
aaagcacctt gccgggggag acgtcacggc gccgggcgat ccggtcgcgg gtctcccaca	14340
gctccgcac caccgccatc tggcggcgcc ggcgcacctt gtgcatgccg gaggtgcggc	14400
gccaggggtc cttgcggggc tccggcggcg gggccgaggc gatcgcgctc aactcctgcc	14460
gggcccagtc cagcttgccc tggcgggtcca gtccttctc cagggcgtcc cgagatcga	14520
ccagcagttc gacgtcgagg gcggcgtagc gcagccaggg ctcgggcagc ggacgggtgg	14580
accagtcgac ggcgaggagg cccttctcca ggacgaagcc gagcacgttc tcgaccatcg	14640
cgccgagccc gacgcggggg aaccgggcaa ggcgccggc cagctcggtg tcgaagaggc	14700
gggagggcac catgcctatc tcgcgcagac agggcaggtc ctgggtggcg gcgtgcagca	14760

cccactcgac gccggacagc gcctcgccga gggcggacag gtcggggcag gccacggggt	14820
cgatcagcgc ggtacccgca ccctcgcggc gcagctggac gaggtaggcg cgctggccgt	14880
agcggtagcc ggaggcgcg ctcggcggtcca cggcgacggg tccgctgccg gccgcgaagg	14940
cggcgaccgc ctcggcgagg gcggcctcgt ccgctatcac gggcggaatg ccctcgcggg	15000
gttccagcaa gggggtcggc gcctccgtaa cagaagatcc gccgtcgtcc ggaggagcgc	15060
ctccgggtggt gcgcagtga ctgtctgctg cgggtgctgt ggcgtcggtc acctgtcaag	15120
ggtatccgtg ccgcgaaggc gcccgtcgac ggttgtgctc cgtgacgggc gccgggtgggt	15180
cgtattccgg tcagaagagt gaaagaacgt gttcgcttgg ccgtgggcgg gcggatcggg	15240
gacgggcgga tcgggggcgg gacggaaggg tcagtggatg atgccggtgc gcagggccac	15300
ggccaccatg ccggcgcggt cggccgtgcc gagcttgctg gcgatccggg ccagggtggct	15360
cttgacggtc agtgcggaca ggcccatcga gacgccgatc gccttgttcg actggccttc	15420
cgccaccagc cgcagcacct ccacctcgcg gccggacagc tcgcggtagc cgcccggtg	15480
gctcggggca cccggggggc ggcggtgcag gcgcgcggcg gcggcgccga tgggagcggc	15540
acccggccgg gtggggagcc cgaggttggt gcgggtgccg gtgacgacgt agcccttgac	15600
tccgcccgcg agggcggtgc gcacggcgcc gatgtcgtcc gccgcggaga gggcgaggcc	15660
gttggggccag cccgcggcg gcgtctcgga gaggagggtg aggccggaac catccggcag	15720
atggacttcg gcgacgcaga tgtcgcgggg gttgccgatg cggggacgag cctccgcgac	15780
ggacgaggcc tcgatgacgt cgcgtacgcc gagcgccac aggtggcggg tgacggtcga	15840
acggacgcgg gggtcggcca cgaccaccat ggcggtcggc ttgttcgggc ggtaggcgac	15900
caggcttgcg ggctgctcga ggagaacgga caccaggcct cctgggggtgc gggacgggcc	15960
ggctcgtggg ggtgaaggcg ggacgaaccg tgctttcaag gtcacagtcg tcttcggcag	16020
caaacctggt gtcctttaac gaatgatcac gaagtgatga gtaacaatcc gggcaattcg	16080
gacgcacgat cgatcattcg aagatcgaac ggtttcggtc tgcgtcga cgcttcgaa	16140
agtggccgta tcgacaaaga gagatgcagg aggccggtcg tcgggacccc gcagcgggag	16200
gctcagcgcg actgcggccc cctccgctgc ggacgcgtca ccacggacgc gtccccgga	16260
gcggccggcg gcagccccgc gacctgcgc agcagatcgg accacgcgac cagatgggcg	16320
gccgtgtccg gaaccccgcc cagaccctca cgcggcggtc acgaggcacg gatctcgatc	16380

tgggaggcgg	cgggccgcgc	ggacagcccc	ccgaagtagt	gcgaactcgc	ccgcgtcacc	16440
gtgccgctcg	gctcgccgta	cgacaggccg	cgcgccgcca	gcgcgccggt	cagccaggac	16500
cagcacacgt	ccggcagcag	cggatccgcc	gccatctccg	gctccagctc	ggcgcgcacc	16560
agcgtcacca	gacggaaggt	ccccgccag	gcgtcgtgtc	cggccgggtc	gcacagcagc	16620
accagccggc	cgtcggccag	atcctcctcg	ccgtcgacga	ccgccgcctc	cagcgcgtgc	16680
gcgtacgggg	cgagccgttt	cggcgcgggc	accgtctcca	cctcgatctg	cggccgcagc	16740
cgggcgctct	gcagcgcctc	gacagcggcc	cgggaagggcg	gcggaggcgc	acctccgtgc	16800
ccgggatccc	ccccgccctc	cttcggttcg	tccattccgc	cagcgcctgc	cgacagtcgt	16860
ccctgagccg	cagccatgcc	gggaagatta	agcggaacgg	gccccggcg	caggaggagg	16920
caccgcgcgc	gcccggcgct	gtccggatcc	tgcaccgcgg	ccccgcccgc	cggacgcccc	16980
tgggggtccgg	ggcggcggac	gggtcgtgcg	agactggccg	gtgtgagtgc	caacacgagc	17040
ccgaagggcc	agacgcctac	cgcgaccccc	gaccccgtca	agaacgacgc	cgtccgggaa	17100
tcagccttcc	tcaaggcgtg	ccggcgcgag	ccggtgccgc	acacgccggt	gtggttcattg	17160
cggcaggccg	ggcgctcact	gccggagtac	cgcaagggtc	gcgaggggcat	cgggatgctc	17220
gactcctgca	tgcgggcccga	gctggtcacc	gagatcacc	tccagccggt	gcgccgccac	17280
cacgtcgacg	cggcgatcta	cttcagcgac	atcgctgtcc	cgctcaaggc	catcggcatc	17340
gacctcgaca	tcaagcccgg	catcggcccc	gtcgtcgagc	agccggtgcg	caccgcgcgc	17400
gacctgccc	ggctgcgcga	cctgaccccc	gaggacgtct	cctacgtcac	cgaggccatc	17460
ggcatgctga	cccgtgagct	cgggtccacc	ccgtgatcgc	gtttcgcggg	cgccccgttc	17520
acccttgcca	gttacctcgt	cgagggcggc	ccgtcccgtg	cgtacgagaa	cgccaaggcg	17580
atgatgtacg	gcgaccccga	gctctgggcc	gacctgctcg	accgcctcgc	cgacatcacg	17640
gcggccttcc	tcgacgtcca	gatccgggcc	ggcgccctcg	ccgtgcagct	cttcgactcc	17700
tgggccggcg	cgctcgcccc	ctccgactac	cggcgttcgg	tgctgcccgc	ctcggcgaag	17760
gtgttccgcg	cgggtggccgg	ccacggcgtc	ccgcgcatcc	acttcggcgt	cggcaccggc	17820
gagctgctgg	ggctcatggg	cgaggccggc	gcggacatcg	tcggcgtcga	ctggcgcgtc	17880
ccgatggacg	aggccgcccc	ccgcgtcggc	cccggcaagg	cgctccaggg	caacctggac	17940

ccgaccgtgc tgttcgccgg ccgggaggcc gtcgagacga aggcgcgcga ggtcctggac	18000
accgccgcgg gcctggaggg ccacatcttc aacctcggtc acggagtgat gccctccacc	18060
gacccggacg ccctcaccg tctcgtggag tacgtccaca cgcagacggc gcgctgaccc	18120
accgctcacg cgccggacgc gagtcggaat ccggggggcg ggtactgggc acgggtgccc	18180
accacgttca cccccgggta cgggcaggtg gaggcccat gaggctcgag atgttcgacc	18240
ccgccccgat cggcgtcgtg ttcacccagg ggccggagca ccggctcgcg tacaccaacg	18300
ccgtctaccg ggagaccttc ggcgaccgcc cgctggggcg gacgatccgc gaggccttcc	18360
ccgacctcgc gcagtcggc tacttcgaca tcttcgaccg ggtcctcacc acgggcgcgg	18420
ccgaggtggt caccgcggtg cccctcgacc tgatctaccc cggctccacg ggcgagggca	18480
ggcgctactt cacgttcagc atctccgcg ccacgatgag cgacggccgg ccgggagtg	18540
tcggcgtgat cgtggaggtg accgcgcagg tgaccgccgc ggaacggatc cgtgtgctgg	18600
ccgaggagcg ccgccgcgcg ctgcagcgct accgcagcct ggtgaacgcc ggaacgcaga	18660
tggtgtgggt ggccggacgcc aagggccgga tcaccgagcc gagccccggc tgggaacgcg	18720
tgaccgggca gacctgggag gagttccgcg gcgagggctg gatgaacgcc gtccaccccg	18780
acgaccgcgc cgctcggtc gaggcgtggc ggccggcgac gaccgaacag gtgccgcgct	18840
ggatccacac ctaccggctg cggctggccg ccggcgggta ccggcacttc gtcgtcgacg	18900
ccgcgcccgt gcgcgacggg aacacggtga tcgaatgggt gggcacctgc acggacatcg	18960
agcgggaatg gcaggagggc cgccgtacgg aactgctggc gcgggcccgc accgccacgt	19020
ccggcatcgc gcggctggac gagatgctcg ccgccctggc cgatgtgatc gtgcccgaca	19080
tcgccgacaa ctgcaccatc cacctcctgc cgcaggccct gcaccgtctg ccgggcaccc	19140
cgctgaccac cgaacgcgtc gccgcggtca ccgccccggg gctcccggac ctgccccgc	19200
accacgagga gcacctgcgg cccggcagcc cgctggcccc cgccgccgac cgccgcagcc	19260
cgctccactt cgtcttcccc cccggcgagc cgccggccga cctcgtccg ctcgacggcg	19320
agccctggat ggccgaggac gtcaacagcg tcgtgctgct gcccgtcgtc gtcgacggca	19380
ccaccgccgc cctggtcgcc gtctccacca gcggcgcccc cccgcccctc ggccaggcgg	19440
agatcggcct gctgcagaca ctcttggaac gcgcccacac cccctcagc aacgccctgg	19500
agtaccagcg caccggcag gtggccctgg ccctgcagaa cagcctgctc accgaccgcg	19560

cggacgcgcc cggcctggac atcgccgtcc gctaccggcc cagcaccgcc gccgccgagg	19620
tcggcgggga ctggtacgac gcgttcgtgc tgcgcgacgg cgccaccgtc ctcaccatcg	19680
gcgacgtctc cggccacgac ctgccggccg ccgtcaccat gagccagctg cgcaacatgc	19740
tgcgcgggct cacgctggac cgccaggaac cgaccggcac catcctgcgc cggctggaca	19800
tcgccgtgca gaccctctat acggagtgca ccgccacctg cgtgctggcc cgggtggaac	19860
gcccggactc cggcggcgtc cggctgcact actccgtcgc cggtcacccg ccgccgctgc	19920
tcgtcgaggc ggacggctcc gcgcgcttcc tgaccggggc gcggtccccg atgtcgggc	19980
tcgtccccgc gccggagtac tcgagcgcca tggaaccgct gccgccggc tccaccctgc	20040
tgctgtacac cgacgggctg gtggagcgcc gcgacgagga tctcaccgtg ggcctggagc	20100
ggctgcggca ccacgcctcg gaggcggtca gccgcccgct gcaggacttc tgcgacacac	20160
tgctcaccgg ccagctcacc gtcgacaacg acgacgacgt ggcgatgctg gtcctgcgcc	20220
ggtaggagcg tgccgaggag cgccactctg gccgatttta cccttgcttt tccatcggga	20280
ttcgttctcc ggatttcccg atccggcgcc gacggcgaga ccgttgggat caccaatacc	20340
ccggaattcc cgcctccgcc accgttgggc agcgacggat cctgtgatat ttcgactacg	20400
cgcggtgatg aattggctcg gtgccggtcg cgcccggtg tagcagttct ggagcgcgtc	20460
tggacatcgt cacgagcgct tgtgattctt ggtcctgtac acgcaagccg gcgcaacgtc	20520
cacgttgccc atcagcggtt atcggcggtc caccggcgcg acggtgaccg cgggcgggta	20580
ctcatagggg gaactgcaat gaattactca aaagcagcga gaggaatgcc gacagccgga	20640
caaggtgccg ttcgggcggc gcgcgtcgtc cgtgaaagtc cggcggaatc agaaacggtc	20700
acagttcaga tagcgtcggt attaccgggt gagtcgctgc gtcgaaagg gatcgagcag	20760
aaccacgtcg cggcactcgc ggaggtagac gcgccgcttc cgccatact ggtggaccgg	20820
aagacgatgc gggtcgtcga cgggatgcac cggctcctcg cggctctgct caacggacgg	20880
cagacgatcg aggccgaact gttcgacgga accgcggatg agggattcct gcgcgccgtc	20940
cgggagaacg tggcgcacgg actcccgtg tcgcaggcgg accgccgggc cgccgctgcg	21000
cgcatcatcg tgtcccaccc gcatctgtcg gacagggcga tcgccgggc gtccgggctc	21060
ggggcgaaga ccgtcgcggc cgtgcggcgc agttcaactg ccgtcgtgcc gcagttgaac	21120

acccgggtgg gccaggacgg cagggtccgg ccgctgaacg ggggcgaggg gcggcgaggg	21180
gccatggcgg tactggccga acaccccgac gcgtccctgc gcgaggtcgc ccgtctgtcc	21240
ggggtgtcgc ccgcgacggt cagcgacgta cgccggcggc tggccgccgg cgagtcgccc	21300
ctgccgtcga gacgggaacc ggccgaaccg cggacgggcg ccgactccca ccgcaaccag	21360
agcttcgtgg atccccgtccc ggtgctggag aagctgctgc gcgaccctc tctgcggcac	21420
aaggagggcg gccgccagct gctccagctg ctccgccaga acgcggtcgg cgtgcaggac	21480
ctgatggagc tgtccgacgc cgtgccgtcc cactgcaggt ccctggtgat ccatctcgcg	21540
cagcagtacc gggacgcctg gcagtccttc gcggagaagc tggacgagcc cgcctgcgcc	21600
tgtcccgggt gacgaacggg cggcacggac ccgttcaccg gacatgaccg gcgccgcgcc	21660
gcgttcacgg cgcgccgccg gcactccac ggaccccgga ccaccgccgc gtatccggcg	21720
gacccgggccc cgggcgggccc ggattcagcg ggcggggggccc cagggtgccac ccgattccag	21780
ccaccgggag agctccgccg ccgagtcctt gcgcacgacc agttcgacga ggccgcgggt	21840
ctggtcctga ccgtgctcga tgcggacgtc ctcgatgttg acgccaagt cgccgatcga	21900
cgtgaacagt tcggccaggg cgccgggctt gtcggagatg gtcaccgaga cggtcgcgag	21960
ctccgtccgg cgcgtaccgg gtttgcgcac gatcctggcg ccccccggt tcccctcccg	22020
caacagctcc tcgagctcct cctgcgcgcg gcggcggacc agcgggtcgg cgtcggagac	22080
ggcgcgcagc gcgccgacgg cccggcccag gccggcggcg agggagtcga gaacgtccgc	22140
cacggccgtg gcgttggaaac gcaggatgtc cccccagagc cgggcgtcac cggccgcgat	22200
ccgggtgacg tcggcgacgc cctgccccgc cagccggacg ctgtcctccg ccgcgtgctc	22260
cagccgcgcg gcgagcaggg aggagagccg atggggcgcg tgcgagacga gggccaccgc	22320
gtggtcgtgc acaccggcgt ccatgaccac cggcatgccg tcgcacaacg acaccatctc	22380
cagggcggtg ttcagcacgt cctgcccgggt cagctccgac ggggtgagca cccaggggcg	22440
cccctcgaag aggtccgccc gggcggcgag cggcccggaa cgctcggtgc cggccagcgg	22500
atggcttcct atgtagctgg ccgggtcggc ccgcatcgcg cgcacgtcgt cgtgcgggac	22560
cttcttgacg ctggcgacat cgaggtaggc tcgggccagc ccgctctcct gtgcgcgcgc	22620
gagcacgcgt ccgacctgtg ccggggggcac ggccagcacc gccaggtcga cctgacggtc	22680
cggtctctcc agggatcccg cgcccatcgc ctccgccgtc ctggcggcgt tccggtcgac	22740

gtcctccagg tgcacgccga ccccgcggcg ggtcagcgcg agagcgacgg acgtgccgat	22800
ggccccggtg ccgatgactg tgggtggtcct caacgcgcgc cccaggtgc ggtgatccga	22860
aatcggtctg gacaagtgcc gtgcccggca cgggaaaagg gaattcccat ggcgccgtgc	22920
gccgccaatt taacgcttcg gcgcgcatgt tcaactgcgg cgtcgcagcg gtcgaacaca	22980
gtagcggtac accggaccat tgaggcatcg tgctcagttg gcgacaccgg gtcggataaa	23040
cgccggaatc cgaggagttg acgttgacgt cagcgctgag acacgacgac ctgcatccga	23100
tagaagaagt ggaaataagt tcgctctcca ccgacggctc ccccgggatc gacggggaga	23160
gtcccagaca cgtggaaatg ctggccgccc ccgacaccgc gcttccaccg atcatggtgc	23220
accgccgcac cgggcgggtc atcgacggca tgcaccggct gcgcgccgcg atgtgacgg	23280
gccgtacgac gatcgcggtg aggttcttcg acggcaccga ggaggacgcc ttcgtcctcg	23340
ccgtgaagtc gaacatcgcg cacggactgc cgctgtccgc cgccgaccgc cggcggggccg	23400
ccgggcgcat catggccacc catccccggt ggtcggaccg gatgatcgcc tcggtggtcg	23460
gcacctccgc caggacggtc gccgagatcc gccgcgacgc cggcgccgcc ggggcggggg	23520
agcccccccg catcggccgg gacggcaggg tacggcccgt cgacgtgagc gagggccgca	23580
gactggccca cgacatgac gtccgcgacc cgggcctgtc gctgcgccag gtcgcccgcg	23640
ccgccgggat ctgccggag accgtcaggg acgtcagaca ccggatgctc cgcggtgagg	23700
acccggtgcc cgcgccgcgg ccgcggaccc tgggtggagcg cggcgccggac cgccgggcgg	23760
agccggccgg gaaggccgcc gcgccgtgcg ggacggagcc gccgcccgc gtcgtgatga	23820
agcggctgag ggccgatccg gcgctgcgtc tcaacgagaa cggacgcgac ctgctgcggc	23880
ttctggatat ccacacggtc cggctggagg actggaaccg cattatcgaa agcgtgccgc	23940
cgcaccgtct ggagacggtg gcgcagctgg cacgctcctg cgccgacaaa tgggtccgaga	24000
tcggtcacg catcgaaagc aacgcatcac atctggccgg gtgaacgagg aaacacacga	24060
atccttcgag gagccgtcgg agaaagcggg acggcccgtc ggaacaccct tgtggagggg	24120
caatggagat acggtcgatc gatcacgtcg aattgttcgt cgaggacgcc caggacacgg	24180
ccggcaggct gtgcgactcc ttcggcttcg tccgcgtggg ccgcggcgcc gggaccaccg	24240
gactgcgcgg ctgcgagtcc gtcctgctgc gccagaacga catcgccctg ctgctgacca	24300

cggccaccga cgccgaccac cgtgccgccg agtacgtgaa gcagcacggg gacgggggtcg	24360
cggtgatcgg catcgggggtg gacgacgcgc gcgccgccta cgccgaggcc gtgcggcgcg	24420
gagccgtccc ggtcgccgcg cccgaggagt tcgggcccgc cggcgcccgt gtcgtcttcg	24480
cctcgggtggc gggattcggc gacgtggagc accgcttcgt ctcccgggag gaccccggag	24540
cgccgttcgc gcccttcacg gaggagaccg gcgcccacgg ctccggggggc atgctgaagc	24600
gggtcgacca cttcgcggtc tgcgtcccgg ccggcgaact cgacgggacc gtccgccgct	24660
accaggaggt gttcggcctc agccagacct tcgaggagcg gatcgtcgtc ggctcgcagg	24720
ccatggactc caaggctctg cagagcgacc gcggcgcggt gacgttcacc gtcatcgagc	24780
cggacaccac ccgcgcaccc ggccagatcg acgcgttcgt ggcctccac ggcgggggccg	24840
gtgtgcagca cgtcgcgttc ctactgagg acatcaccac cgcggtgcgc acctgcaccg	24900
ggcgcggggt ccgcttcctc accacgccgc cgagctacta cgagatgctg ccggggcggc	24960
tgggcccggg cggcgtagcc gtggaggagc tcagcgcgct caacatcctg gccgaccgcg	25020
acccgtccgg gatcatgctg cagatcttca ccgagtcgac gcacccgagg cggaccctgt	25080
tctgggaact gatcgaccgc cgcgggcgcg agaccttcgg cagcaacaac atccaggccc	25140
tgtacgaggc cgtggagcgc cagcaggcgg cggaggcggc cgaccaggaa tgaggaagct	25200
ccccgcagac gcgtgtggac ccggaggaca cgccctccgg gtccacacgc gtctgcgggg	25260
ccagcgtcgg ctacgccccg aggagccggc cgccgtgcag cgtctccccg tacgcgaaca	25320
catggccggc cccacacctc acgggcgcga gcagtgccag atccgtgccc ggacggccgg	25380
agagggcgag ccggccggga ccgctccaga ccggtgagaa ctcgacgccg ctacacctcg	25440
aggcgaccag gcggggccgt ggcgggggcgt cggacggcac ccaccgcgga aggacgagac	25500
tgtgcgccag cgggacgtcg tgcagcgcg gcggccgctc ggagcggcgc tcgaccgtga	25560
ccgacgcctc ggcggtgagc cggccgtgga cggagagcgc gccgtcgaag cggcccccg	25620
gagcgagccg tgagcccgc cggccgacgg tcaccggcct ggtctggtgg atggcgccga	25680
actgcttggg catgccctgg accagccgc gcaccatcgg cacgggctgg tcgaccagg	25740
cgaacgggca gcgcgccatc ggccggccct cgaacgcga cccgaggagg atcaggaact	25800
ccgagaaccg gcagacggcc gggtcggcca gtcgcgcgc gtcctcgag caccactgcc	25860
aggtggcgaa cacggcggcc gccgcacccg gatccgctcc cgcgtccagg cccggcgga	25920

ggaaacgccg tgcggcgctc gggtcgacac ggtagtcgac catgaggatc tcgccggaga	25980
agtgccacgg cggaggcgtg agcatcgacg cctgccccga aggggacagg ggaaggctgt	26040
agccgatggg cccggcggcc ccggccgctt ccggatccgt cggatgtgtg tgcccgggtg	26100
tggccgtcat gggttccctc cgatctgccg gtccggcggg ccgccggacc atgcctgggt	26160
cagccgtcga gcggcgctt cgagcagcgg gggcggattg aagctgtagg ccaggcgcac	26220
gctcggctcc cgtgtgccgc cgaagcgca acccgcggtg acgcgcacac ccgccgctc	26280
cgcacgggag agcagttcgt cctcgccgag cccgggtgcc caccggagcc agaggaagaa	26340
cccgccctcc ggacggctga tccgcaccgg gaggtccgcc gcctcccgca gcgcgtcag	26400
gagggcgctc cgccgcgccc tcagaccgc ccgcaacatt tccagatgcc ggtcgtagcc	26460
gccgtcggac agcagccctg cgacggcgag cgaggtgatg tggttgagcg acccgccgct	26520
gcggaacagc ccgtgcgacg cgatccgtt gcccagtgcc ggctccgtca ccagccagcc	26580
cagccggagc cccggcccca gggctttgga gaagctgccc agccgcacca cgccccgggt	26640
tccggcgagg gccgccagt gcggcggggc cgggggaccg tccgtcaggc ccagttcgcc	26700
gtaggcgctc tcctcgacga ccaggacgcc gtgctccgcc gccgcctcca gcagccgag	26760
ccggcgctcc agcggcatgg tggcgcccgt cggattgtgg tgggtcgggg tgaggtacac	26820
gaacgcggtg cggccggtgc cgccttcgcc gccccgcgcg gtcccggcga gggcgcgccg	26880
gagcgctcc ggcaccatgc ccgacgcgtc gagggcgacc cgcctcaggc gcagcgcgca	26940
gtccccgagg atgcgtgcc cgaggtcgta gccgaggccc tccacgagca ccgtgtcgcc	27000
gggcctcgcg aggggtggtc ccagcaggtg gagcgcttg gacgtgccc ccgtgacgac	27060
cacgtgctcc ggcccgcacg gggaccgccc ccgcacgggt gcccgggcgg ccagctcggc	27120
gcgcaggggc agggcgcccc gatcgtgtcc gtagccagt gccgccgtc cgtactcctc	27180
cagcgcgct gcgtaggcgt cccgcaccag cccaccggc agcagcgccg gttcgaggt	27240
gccgggcccc aggtcgagga cggcgcggg ggcgacctc tgcaccacac cgcgacgcca	27300
ccgccgctg tgcgacaac ggcgggcccgt gccgtacggc aggggtcccct cgccggcagc	27360
ggtcatcagc ggggtgtcag cacatggcgc aacgcccgt cgcactgggc cagcggggcg	27420
gacggcggg cgagcgccag gcgcagcgt cggtcgccgc gggcggggtc agcccagtag	27480

aaggcacggc agggcagggc gtacacatgg tgctcgcgca ggcctccca gacctcggtc	27540
ccggtcagat gcctgatcag caccgctcc acactggccc ggctgtccgg gtcgggcacc	27600
ccggtggtcg acaggtccgc cagcccggcg cgcaccaccg accgctgggc ggcgatgaac	27660
tcgtgcagct ccgtcagccc gccggcggcg gcgtcctcgg agaagcggcg gaccatcccg	27720
aggatcagcg gggagacgcc cagcaggatg tcggagtaga tcttctccac cggcaggccc	27780
aggttctcgg agtggaccag catgccgacc ttgaggtcga gggtcggcca gagcttgccc	27840
gtgtcctcga tgacgacca gcgcacatcg ctggcgtcga ggatctcgta gtggtcgtac	27900
tgggcgcggg tgtcgaagcc gcggaaggac gtgtcgaggg cgaggatcac gccgtgccgt	27960
gcgcactgcc cggccagccg gcgcagccgc tccgccgaca cgaccggcc cgtcgggttg	28020
ttcggcgtgg tgacgaagac acagcccacg gactcgagca gctccgcggg caggtcgtcg	28080
gcgtgcagcg gatcctctc caggggcacc agaccaggc ggttgccgcg caacaggtcg	28140
gcgatgttgt cgaaggtggg gtggaccagc gccacggagt ccgtgaccga cgccaggcg	28200
cgggagagga tctccatggc caccgacgag gagtagcagc tcagcacacg gccgggtgcg	28260
gacgggtagc ggtgctggcc gagggccttg aagaaggccg cgtgggcctc gcgttcgagc	28320
tgctcgacgg ggcgcttctc gccgtcctcg aaaagcagcg ggagatcatt gacgatcttg	28380
ctctggccgg gagtgagcgg ctgccgggca tgcccgtcgg cgatgttgaa ctgctgttg	28440
agtgcgagga attccagttg ggtgaggttc tccgcgcctg atccggcgtg cgcagcgtgg	28500
gccttgcttt gcagtgttcc ggacacaggc atgcctctct gggatgtgag ggtttccaga	28560
agcggagcgg acgtaaatga gcggcccact ctacggcctt cgccctccgg ctgaaatgcc	28620
tcttcttttc ggcaccgtgt tcaactgcgg tggcgcggca gtcgaacgag ccgtctcgcc	28680
cgccgtatcg gccggacatc gcgttccgac ggtgacgcgc gtgcgggttc cgtgtccaac	28740
tgacctgagg gcgcagttgg acgggccacc ggcacacggc cgccgatcc ttgtcggacg	28800
ggcccgggca cgcgaaagtg gacgtgcggg atctgtgttc cgccgccggg gtctcttcgt	28860
aagccgtgaa gtggggcctt gatggaattg tcgctcgatg aattcgctc gctcgcccgg	28920
gaacggctgg acccgccgt ctgggatttc atcgaaggcg gcgccggaga ggaacgcacg	28980
ctcgccgca acaccgcggc attcgaccgc gtcccgtgc ggccgtcggg gctgcgcggc	29040
gcgggcagcc cgcacaccgg caccacgatc ctcgggcgga cgtgggacgc gccctcgcg	29100

gtcgccccgg tggcctacca cacgctcgcg gacccggccg gtgaggctgc caccgtccgg	29160
ggaacggcgg ccgccgccgg actcccggtc gtcgtcagca ctttcgcggg ccgcacgttc	29220
gaggacatcg ccgccgaggc caccgtcccc ctctggctcc aggtgtactg cctgcgggac	29280
cgctccctca cccgaggcct catcgaacgc gccgagaacg cgggcttcga ggccctggtc	29340
ctcacggtcg acgcgccgca cctcggccgc cggctgcggg acctgcgcaa cggcttcagg	29400
ctgcccgcgg gcacgggtccc cgccaacctc ccggtggacg gattcgcgga ccccgcgggc	29460
cactccccgc ccgacttcga ccccggcctg gactggtcgg tggaggagtg gctgcgctcg	29520
gtctccgaac tgccgttgct cgtcaagggg atcctcaccg gcgccgacgc ggtgcgcgcg	29580
gccgaggccg ggggtggacgg cgtcatggtc tccaaccacg ggggccgcca gtcgacgga	29640
gtgccggcca ccctcgacgt cctgcccag gtcgccgagg cggtcggcgg acgcctcccc	29700
gtcctcctcg acggcggggg ccgccggggg cgggacatcc tggcggcgct cgcgctcggc	29760
gccgacgcgg ccctcgtcgg ccgccgggtg ctgcacggcc tcgcggccgg cggggccggc	29820
ggggtgaccg gcgtcctctc cgtcctcctg gaggagctga cggacgcgat gtcccttgcg	29880
ggcctgagga ccctcgccga catcggtccc tctcgtcg gccgggctcc tgaccacccc	29940
cgccgaagca ccgtggacgc cgggaagggc gcggggagcg accggcgcac cgccgccggg	30000
ggaggggccc ggctgcgcct cgcggacctg caccgagtg tcgccgacct ggtcatggac	30060
accatgaact tcctcaacga ggtgacactg cgctaccccg aggcgggtgtc cttcgcccc	30120
ggacggccct acgcggagtt cttcgagacc gagcaggctt tccgccatct gcgccgtac	30180
ctcgaccacc tggccgagca gggccgttcg cccgcgcagg tgcgcgacgc gctgttcag	30240
tacggtccgt ccgccggtgt gatccgcgag ctgatcgccc actcgtgcg ggtggacgag	30300
ggcatcgacg tgtcgcccga gtcgatcgtg gtgacggtcg gctgccagga ggcgatgttc	30360
ctgacgctgc gcgcgtcat gtccggcccc gacgacgtgc tgctcgtctc cagcccctgc	30420
tacgtgggga tcaccggggc cgcccggctg ctggacgtcg cggtgaccgc cgtcgaggag	30480
ggcgaggacg gcctgtcgtg cgacgccctc gaggccgccg tctcggcgga gcgggcgcgc	30540
ggcaggcggc cgcgggccgt ctacgtggtc ccggaccact cgaacccgtc cggcgcgacc	30600
atgccgctcg aggcccggaa gtccctcctg gagctggcgc agcggctcga cgtcctcgtc	30660

ctggaggaca gcccgtaccg gcacgtcagc ccgggcacgc aggtggcgctc cctgaaggcc	30720
ctcgaccgga cacggcgagt gatccacctc ggttcctacg ccaagaccgt cttccccggg	30780
gcacgcctcg ggttcgcggt cgccgaccag ccggtgctgg cgccggacgg cggcacgagt	30840
ctgctggcgg acgaactcgc caagatcaag agcatgggtca cgggtcaacac ctgcccgtc	30900
agccaggccg cgggtggcggg cgcgtgctg gagtcgggcg gccgtgtctc ggagctcaac	30960
gcccgcacg ccgcccacta cggggaggcc atgcgcttca ccctgcagtg cctggagcgg	31020
gagttcccgg ccgcgcggcg gacccggctc ggcgctccgt ggaacgcgcc cagcggcggg	31080
tttttctca ccctccaggt gccgttccgc gcggacaact ccgcgtggc ccggtccgcg	31140
caggacttcg ggggtcatctg gacgccgatg tcgtacttct atccgcaggg cggcggcctg	31200
cacaccctca ggctctccac cagctacctg acccacgccg acatcgagaa gggcatctcc	31260
cggctggccg gggtcatcga gttcgagtgc ggggacccgg tggcctgaac cgccgcgacg	31320
acgaagggcc ccggccgcgc cggcggggcc cttcgtcggg cgcgacgctc aggacggatg	31380
cggctcctcc cagaacatgc tgtcggacga ggcgacgaga ccgccgtcga agagcgggac	31440
gtcctcgccg gcgtactctc ccagctgcgc ccgggtctgc acctgcgagc cctcctgcat	31500
cgcccggctc accgccgacg accggaacag cggggccatg ttctcctcct gccgccggc	31560
catgtcgctc acggcgctcg cgaactccgc cgactgctgc ttgagccgca ggacgctcga	31620
ctcggcgctc gagaggctga agtccgtgct ggacatgccc gccaccagct ccacgaacga	31680
ctccagctcg gcgtggctgc tccgggtgac cttcttcgcc gtccagaagt aggagtctc	31740
gtcgacgtgc atgtcgtaga acgaggctcag gaactcgtag aagacgccgt actcccgccg	31800
gtaccgggcc tcgaactcgt cgaaggcccc ccgctcgtcg atccgccccg ccagcacgct	31860
gttgagggag cgggcggcca gcagcgcgct gtaggtggcc agatggaccc cggaggagaa	31920
gacggggctc acgaagcacg ccgcgtcacc gacgagcacc atccccggcc gccagaaggt	31980
cgtgtgggtg tacgagtagt ctttgcgcac ccgcagctgc ccgtactgac cggtcgtgac	32040
ccgggtggcg tccgccaggt actccttgat catcgggcat tcgtcgatga ggccgcgcag	32100
cgcgctctcc gggtcgccct gcaccttggc cgcgctctcc cggcggacga ccgcgccgac	32160
gctggctcagc gtggagctga gcgggatgta ccagaaccag ccgctgccga acgccacaca	32220
gaggatgttg ccggcgtagg gcgccggcat ccgcttgccg ttctcgaagt agccgaacag	32280

cgccaggctc ttgaagaagt ccgagtacgt gcgcgagccg ccgacccgct tgtggatgcg	32340
gctggtgttg cccgaggcgt ccaccacgta ccgggcccgc acctcgtgct ccgtgccgtc	32400
cgggtcgggtg taccgcagcc cgcgggccccg cccgtcggcg tcgtccacga catcggtgac	32460
cgtgcgggcc tggcggacca cgacaccctt gcgggcccgc ttgtccagca ggatcttgct	32520
gaatttgctc cgctccacct gataggcgaa cgaggctcggc ccggagacct tggacgagac	32580
ggagaaggag aaattccacg gcttggggct cgcacccac cggaacgtcc caccgcgctt	32640
gtgcggaaaa ccggcggcgg cgagttcgtc ggtgacaccg agcagatggc agatgccgtg	32700
aatggtcgac ggcagaagcg actcgcctat ctggtacctg gggaaggctt ctttttcgag	32760
cagcagcaca ctgtgcccct gcatggccac cagggtcgag agcgtcgacc ccgaagggcc	32820
gccgcccacg accacgacgt cgaattcctc gtgctgtcct gtactcattc ggcctcccgg	32880
cacgcactga tgcggtcatc gcgctggtga ctcttttgct agggttccac aggactcaaa	32940
ggggccaaag ggtggacgac attccgtgaa tggacagccg ggcccgggcg cgcgcacgca	33000
aaggcgccgc cggggaattc ccggcggcgc ccggaccggc gtcgggtgag acgggggtca	33060
actcaccggc cgccggatcc gccacacccc gaacaccgtg aggattccgc tgagtgccac	33120
atagatcgcg gtctccagcc actggaacgc ccagtagcgg ctgctggggc ggtacaggac	33180
gtcgacgtgc aggtcgtgtt cggcgaggca caccgcgggt tcgccgaacg tgcccccg	33240
gccggtcttg ggcgggtcgt cgaggcagcc gttgaactcg ctggaggcga gggctctgcc	33300
gtccgcgggtg cgcagcggac tggctcggc gatccacgcg tccggcgcgt ccgggatccg	33360
caccccgccg atgacggatc cgccgccgat actgcccagg ttctgcgccg agttgatcgc	33420
ctcggccgtc atcgccagcg tcgtcctgtc cggcggcatc aggctgggcc gcaccacgtt	33480
cgggaagaag aactggaagg cgatgaagac caccagcgtc accgccatcg cgggcagggt	33540
ccgccgcagc aggagcccga cgacggtgcc gaacgtgaag gccagcgcgg cgtagccgat	33600
cggggcgatg ttgcgcgcac cgaacacgaa ggtgtcgaac tgctccttga cgacgtcgtc	33660
gaagggccgg gccgcccagg tgagcagggc cgcggccgca ccggtcacga tcaccgaggc	33720
cgcgccgatg agcaggatct tgctgagcag ccagcgcggc cgggtgacgc tctggttcca	33780
caccagccga tgggtgccgt tctcgagttc cctggcgatc aggggagcgc ccagaagg	33840

gccgatgagc gcggggatca gggccaggcc ggtcgccagg aacagcaggg tgttctggaa	33900
ggtgctgcgg aactggctcc tggcctgggc gcagttggcc gagttgtcgc agttggcctg	33960
gtagacgtca tgggcgtcac ggatgtcccc gcccaggtag agcaggtaga cggcgatcac	34020
ggccagcgcg cggcgccga acagggcctg gacgcggaac tgccgccagc tgagccacat	34080
catcgggtgg cccccaggc tgcggcctcg gtgcgggcgg cgggaacggc ggccgcccgg	34140
gtcatgtagg cgagcacgag ttcttcgagg gtgaccggct cggaccggta gggcagtgcc	34200
tcggtcgcgg cgccggtgcg gacgaccgcg ctgctgtgct tgccgctgtg ctcgaccgag	34260
atcacctcga tcccggcggg cggctggtcg aactcgccgc gggccgcgac cagccgggcg	34320
tgcccggcca gcagctcccg ggtgtcgccg gcgacctgca cccgggcgtc gcacagcacg	34380
atgagatagt cgagacctg ctccacgtca ccgaggaggt gcgaggagag gacggcgctg	34440
gcgccgagct ccagcacgaa ctccatcagg ttctgcagga acccccggcg cgccaggggg	34500
tccagggccg ccgccggctc gtcgaagatc agcagctccg gccgcttggc cgccgcgatg	34560
gtcagcgcaa gctgcgcgcg ctggccaccc gagagctgcc cggccttctg cccggcgctg	34620
agccccacct ggctgatgcg ccgctctgcc aggaccgggt cccagcccgg gttcatcttc	34680
gcgccgaact tcaggtgctc cgccacgggtg aacgcgccgt acaccggcgt gttctgcgcg	34740
acgaacccca cccgggcccag gtgcgacgcg ttggccgccg gacgcgagcc gaggacgctc	34800
agtgagccgg acgtcggttc ggtcagcccc caggccaggt gcaggagggt cgatttgccg	34860
gccccgttcg ggccgaccag tccgatgaca cggccggcgg ggacgctgag gtgcacgtcg	34920
ctcagggcga gcttgccgcg gcggccgtac ttcttcgtca agccctccgc gtgaagcacg	34980
ggaggggagt ctgcgtgtgg catgactcca tcctcgaatt ccgccccgtt cacggcatca	35040
gtccaaagca cggttcccgt tgccggcggc cgtactttcg gccggtcggc cacggccctg	35100
ctggtgggcg aactgggtgc ggtacagctc ggagtagaga ccgccgccg ccagcagctg	35160
gtcgtgggtg ccccgctcct ggatccgccc gtcgtcgatg acgaggatct ggtcggcgtc	35220
ctggatggtg gacagccggt gcgcgatgac gagcgagggt cgcccgggtca gggcggtctt	35280
gagggcccg cggatggcca gctcggactc ggagtccagg tgcgccgtcg cctcgtccag	35340
gacgacgatc ggaggcgact tgagcaggag ccgggcgatg gccagccgct gcttctcacc	35400
gccggacagc cggtagccgc ggtcgccgac gaccgtgtcg agaccgtccg ggagctggga	35460

gatcgtcggc	cagatccgcg	ccgcctcgca	cgcctggacg	atctcgggct	cggaggcgtc	35520
cgggcgggcg	tacagcaggt	tggcccggat	ggtgtcgtgg	aacaggtgcg	cgtcctgggt	35580
gaccacgccg	accgtgttct	gcagcgagcc	gagggtcagg	tcgcggacgt	cgtggccgcc	35640
gatccgcacc	gttcccaggg	tggcgtcgta	gagccgtggc	accaggtggg	tgatcgtggt	35700
cttgcccgcg	ccggacgggc	cgaccagcgc	cgtgagccgg	ccggccgggg	cgtggaagct	35760
cacgtcgttg	aggaccagcg	cgccggggcc	ctgctcgctc	ttgcgctgcg	gcatcaactc	35820
cagtgagggc	agggacactt	cctcggcgct	ggggtagcgg	aaggcgacct	ggtcgaactc	35880
gacgggggga	gcggtgccgt	cgccgttcgc	cgaggcgagg	gccggcaggg	ggcgggcgcc	35940
gggacgctcg	gtgatcagcg	gcttcaggtc	cagcacctcg	aagacgcggt	cgaagctgac	36000
cagcgcggtc	atgacgtcgc	tctggatgtt	cgtcagctgg	ttgacggggc	cgtacagcat	36060
cagcagcagg	gcgaccatgg	ccaccagcgt	gccgatctgc	agcgagccgt	cgatgacgaa	36120
ccagccgccg	aagccgtaca	ccatcgccgt	ggtgacggtg	gtgagcaggg	tgacgaggat	36180
gaacagcagc	cgtgcgtgca	cgtccatcga	gatcgcgatg	tcccggacga	ggcccgcctt	36240
cttggagaac	tcggcggact	cgtcctccgg	acggccgtag	agcttgacga	gcatcgcgcc	36300
ggagatgttg	aaccgctcgg	tcatcatcga	gcccagcttg	gcgtcgttct	gcatgccggc	36360
gcgggccagc	ttctccagcc	gctgggcgat	gatcttcccg	gggatgaaga	acagcgggat	36420
catgatcagc	gccgccacgg	tgatcggcca	cgagaggtag	agcatcgccg	cgagcaccag	36480
gaccagcgtc	agcagcgctg	acagcgactg	cgacagcagc	gaggtagagg	cctgttgggc	36540
gcccacgatg	tcggtgttga	tccggctgac	cagcgacccg	gtctgggtgc	gggtgaagaa	36600
cgccaccggc	tgccgctgga	tgtgggagaa	caccgcggtc	cgcaggtcga	agatgaggcc	36660
ctggccgacc	cttccggaga	accacgtctg	cgtgtagacc	gccacgacgt	tcagcagggc	36720
cagtccggcg	acgagcccgg	cgaggccgaa	cacgacggac	gtcttcccgg	ggatgatgcc	36780
gtcatcgatg	atcattttga	gggtcagcgg	gatcgacacc	gtgatcaggg	agtcgacgat	36840
cgtcgccacc	atgaccatcg	ccatggcccc	gcggtagcgc	atggcgtagg	gaatgatccg	36900
cttgaacgtg	ccggacctga	ccggctgcgg	gtccaccagt	ccttcgaccc	gcagtccgat	36960
cgtgcccata	gtcgggtcgt	gtcccacggg	cacggagact	ctcctcagtg	tgtgtcgcgt	37020

cgctatgtgt cgcgtcgtg tgtggcgcgt cgctgtgtgt cgtcggggcg tcgctcagt	37080
gcttgtcgag gaactcgacg atcgccgcgt tgaccgcgtc gggccgctcg aagtacccga	37140
ggtgccccga gtccgggatc tccacgagat cgcagtcggg aacggcctcg gcgacctcca	37200
cgcccaggtg cggcgggggtg atgaggtcgt cggcgaaggt cagcagcgga cagggggcgg	37260
ccaccggcg cagcgccggg cggcggtcgt ccatgatgtc ggcccaggcg tgccgggcct	37320
gcgcctcccc gccccggac agctcgaaga cgtccagcca ggcggtcacc gcctggtcgt	37380
cgttgagcgt cgcgggcgag aacatccgga acaccgtcga cgcggcgtcg tacgcggccg	37440
gcagccgcac cccgctctcc accagtgccg tctccggccg cgtctgggcc cgccgcgcgg	37500
cgtccgcacg ggcccgggtg gcgatgagca ccgcgcaccg cagcagttcg ggatgccga	37560
tcgccagctc ctgcgcgatc atgccccca gggaggtgcc cagcatccgg cagggcgcca	37620
gatccagggc ctgatcagg cccttggcgt cggcggtcac gtccagcagc gtagtacctgc	37680
ccggcggcgc gtcggacggc gggacacccc ggtggtcgaa gacgaccgtg gtagtagccg	37740
ccgtgtgcag cgccggcgtc tgggtgcaggg tccaggcatg gccggccgag cccgagccca	37800
tgatcatgag caccggttcg ccccgggccg cagcgtggta ggcgatgcgg acgccccca	37860
cggtgacgaa gtgcggggcg cgccggcccc cgggtgtggtc catgccgcct ctccctctc	37920
gtcgtcgcgg gggccgcccc gtggtaccgg ccggccccgg gggcggctca ctatcgacg	37980
cggccacggg gcggggcagt gtgcgcgggg cagctccatg gacaccccc ggccgcgtc	38040
caactgcggt gatgccctca gttggacacc ggccggccgc gtccaagcag gcccggccga	38100
cggttgatcc gctgtgtgga gctgagccat attgggcccgc cgtgagccac tgacggccac	38160
caagtccccg cgcttcttcc gcaccggccg cctggcgcgc ccgtcgcgcc gagggaggga	38220
ccaccttgtc agcttccgac cttccagcca cccggctgac acccgagaag atccggtcct	38280
ggctcgtcga ccgggtcgcc tactacgcca ggctgccccg cgaggagatc ggcgccgacg	38340
tcccgctcgc gcactacgga ctggactcgg tgtacgcctt cgccctgtgc ggagacatcg	38400
aggacggcct cggcctcgtc gtcgagcccc tctgtctctg ggacgtcgac accatcaccg	38460
agctcaccga ccatctcgcc gaactgacag ccgactgagg gccttcgagg gggaggacga	38520
tgcgtcgaaa ggacctggag aggctgacgt ccggtcagct cggcgtctgg tacgcgcagc	38580
agctcgaacc cctgagcccc gtgtacaaca tcgccgagta cgtggagatc cgcggcgacg	38640

tggacgtcgg gcttctggtg tcggcgctgc ggtctgccct cgacgaggcc cagacctacc 38700
 ggctccgctt ccggcaggag gacgccggcc ccggacagta cgtcgacgac tcgctggagc 38760
 ttcccgctca cgtcgccgac ctccggtccg caggggaccc gcgcgccgcg gccgtggagt 38820
 ggatgaccgc cgacctggac cgccccgcgg accccctcac cggcccgctg gccgcccacg 38880
 ccgtgttccg gctgggaccc ggccatgtcc tctggtacca gcgtgcccac cacctcgtcc 38940
 tcgacgggac cagcctctcc gtgttcgccg gtgggatccc cgggtaccga gctcgaattc 39000
 gccctatagt gagtcgtatt acaattcact ggccgtcgtt ttacaacgtc gtgactggga 39060
 aaaccctggc gttaccaaac ttaatcgctt tgcagcacat ccccttttcg ccagctggcg 39120
 taatagcgaa gaggcccgca ccgatcgccc ttccaacag ttgcgcagct gaatggcgaa 39180
 tggcgcctga tgcgggtattt tctccttacg catctgtgcg gtatttcaca ccgcatatgg 39240
 tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagccccg acaccgcca 39300
 acaccgctg acgcgaaccc cttgcggccg c 39331

<210> 40
 <211> 40551
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Fosmid pXYF24 tổng hợp

<400> 40
 atcgaatata acttcgtata atgtatgcta tacgaagtta ttagcgatga gctcggactt 60
 ccattgttca ttccacggac aaaaacagag aaaggaaacg acagaggcca aaaagctcgc 120
 tttcagcacc tgtcgtttcc tttcttttca gaggggtattt taaataaaaa catlaagtta 180
 tgacgaagaa gaacggaaac gccttaaacc ggaaaatttt cataaatagc gaaaaccgcg 240
 gaggtcgccg ccccgtaacc tgtcggatca ccggaaagga cccgtaaagt gataatgatt 300
 atcatctaca tatcacaacg tgcgtggagg ccatcaaacc acgtcaaata atcaattatg 360
 acgcaggat cgtattaatt gatctgcatc aacttaacgt aaaaacaact tcagacaata 420
 caaatcagcg aactgaata cggggcaacc tcatgtccga gtcgcgagc tcgtcgacag 480
 cgacacactt gcatcggatg cagcccgggt aacgtgccgg cacggcctgg gtaaccaggt 540

attttgtcca cataaccgtg cgcaaaatgt tgtggataag caggacacag cagcaatcca	600
cagcaggcat acaaccgcac accgaggtta ctccgttcta caggttacga cgacatgtca	660
atacttgccc ttgacaggca ttgatggaat cgtagtctca cgctgatagt ctgatcgaca	720
atacaagtgg gaccgtggtc ccagaccgat aatcagaccg acaacacgag tgggatcgtg	780
gtcccagact aataatcaga ccgacgatac gagtgggacc gtgggtcccag actaataatc	840
agaccgacga tacgagtggg accgtggttc cagactaata atcagaccga cgatacgagt	900
gggaccgtgg tcccagacta ataatcagac cgacgatacg agtgggacca tgggtcccaga	960
ctaataatca gaccgacgat acgagtggga ccgtgggtccc agtctgatta tcagaccgac	1020
gatacgagtg ggaccgtggt cccagactaa taatcagacc gacgatacga gtgggaccgt	1080
ggtcccagac taataatcag accgacgata cgagtgggac cgtgggtccc gtctgattat	1140
cagaccgacg atacaagtgg aacagtgggc ccagagagaa tattcaggcc agttatgctt	1200
tctggcctgt aacaaaggac attaagtaaa gacagataaa cgtagactaa aacgtgggtcg	1260
catcaggggtg ctggcttttc aagttcctta agaatggcct caattttctc tatacactca	1320
gttggaaacac gagacctgtc cagggttaagc accattttat cgcccttata caatactgtc	1380
gctccaggag caaactgatg tcgtgagctt aaactagttc ttgatgcaga tgacgtttta	1440
agcacagaag ttaaaagagt gataacttct tcagcttcaa atatcacccc agcttttttc	1500
tgctcatgaa ggtagatgc ctgctgctta agtaattcct ctttatctgt aaaggctttt	1560
tgaagtgcac cacctgaccg ggcagatagt tcaccggggt gagaaaaaag agcaacaact	1620
gatttaggca atttggcggg gttgatacag cgggtaataa tcttacgtga aatattttcc	1680
gcatcagcca gcgcagaaat atttccagca aattcattct gcaatcggct tgcataacgc	1740
tgaccacgtt cataagcact tgttgggcga taatcgttac ccaatctgga taatgcagcc	1800
atctgctcat catccagctc gcccaaccaga acacgataat cactttcggg aagtgcagca	1860
gctttacgac ggcgactccc atcggcaatt tctatgacac cagatactct tcgaccgaac	1920
gccggtgtct gttgaccagt cagtagaaaa gaagggatga gatcatccag tgcgtcctca	1980
gtaagcagct cctggtcacg ttcattacct gaccataccc gagaggtctt ctcaacacta	2040
tcaccccgga gcacttcaag agtaaaactt acatcccgcac cacatacagg caaagtaatg	2100
gcattaccgc gagccattac tcctacgcgc gcaattaacg aatccaccat cggggcagct	2160

ggtgtcgata acgaagtatc ttcaaccggt tgagtattga gcgtatgttt tggaataaca	2220
ggcgcacgct tcattatcta atctcccagc gtggtttaat cagacgatcg aaaatttcat	2280
tgcagacagg ttcccaaata gaaagagcat ttctccaggc accagttgaa gagcgttgat	2340
caatggcctg ttcaaaaaca gttctcatcc ggatctgacc tttaaccaact tcatccgttt	2400
cacgtacaac atttttttaga accatgcttc cccaggcatc ccgaatttgc tcctccatcc	2460
acggggactg agagccatta ctattgctgt atttggtgaag caaaatacgt acatcaggct	2520
cgaacccttt aagatcaacg ttcttgagca gatcacgaag catatcgaaa aactgcagtg	2580
cggaggtgta gtcaaacaac tcagcaggcg tgggaacaat cagcacatca gcagcacata	2640
cgacattaat cgtgccgata cccaggttag gcgcgctgtc aataactatg acatcatagt	2700
catgagcaac agtttcaatg gccagtcgga gcatcagggtg tggatcggtg ggcagtttac	2760
cttcatcaaa ttgtcccatt aactcagttt caatacgggtg cagagccaga caggaaggaa	2820
taatgtcaag ccccggccag caagtgggct ttattgcata agtgacatcg tccttttccc	2880
caagatagaa aggcaggaga gtgtcttctg catgaatatg aagatctggt acccatccgt	2940
gatacattga ggctgtttcc tgggggtcgt taccttccac gagcaaaaaca cgtagcccct	3000
tcagagccag atcctgagca agatgaacag aaactgaggt tttgtaaacg ccacctttat	3060
gggcagcaac cccgatcacc ggtggaaata cgtcttcagc acgtcgcaat cgcgtaccaa	3120
acacatcacg catatgatta atttgttcaa ttgtataacc aacacgttgc tcaaccgctc	3180
ctcgaatttc catatccggg tgcggtagtc gccctgcttt ctcggcatct ctgatagcct	3240
gagaagaaac cccaactaaa tccgctgctt cacctattct ccagcgccgg gttattttcc	3300
tcgcttccgg gctgtcatca ttaaactgtg caatggcgat agccttcgtc atttcatgac	3360
cagcgtttat gcactgggta agtgtttcca tgagtttcat tctgaacatc ctttaatcat	3420
tgctttgcgt ttttttatta aatcttgcaa ttactgcaa agcaacaaca aaatcgcaaa	3480
gtcatcaaaa aaccgcaaag ttgtttaaaa taagagcaac actacaaaag gagataagaa	3540
gagcacatac ctcagtcact tattatcact agcgctcgcc gcagccgtgt aaccgagcat	3600
agcgagcgaa ctggcgagga agcaaagaag aactgttctg tcagatagct cttacgtca	3660
gcgcaagaag aaatatccac cgtgggaaaa actccaggta gaggtacaca cgcggatagc	3720

caattcagag taataaactg tgataatcaa ccctcatcaa tgatgacgaa ctaacccccg	3780
atatcaggtc acatgacgaa gggaaagaga aggaaatcaa ctgtgacaaa ctgccctcaa	3840
atttggtctc cttaaaaatt acagttcaaa aagtatgaga aaatccatgc aggctgaagg	3900
aaacagcaaa actgtgacaa attaccctca gtaggtcaga acaaattgtga cgaaccaccc	3960
tcaaatctgt gacagataac cctcagacta tcctgtcgtc atggaagtga tatcgcggaa	4020
ggaaaatacg atatgagtcg tctggcggcc tttctttttc tcaatgtatg agaggcgc	4080
tggagtcttg ctgttgatct cattaacaca gacctgcagg aagcggcggc ggaagtcagg	4140
catacgctgg taactttgag gcagctggta acgctctatg atccagtcga ttttcagaga	4200
gacgatgcct gagccatccg gcttacgata ctgacacagg gattcgtata aacgcatggc	4260
atacggattg gtgatttctt ttgtttcact aagccgaaac tgcgtaaacc ggttctgtaa	4320
cccgataaag aagggaatga gatatgggtt gatatgtaca ctgtaaagcc ctctggatgg	4380
actgtgcgca cgtttgataa accaaggaaa agattcatag cttttttcat cgccggcatc	4440
ctcttcaggg cgataaaaaa ccacttcctt ccccgcgaaa ctcttcaatg cctgccgtat	4500
atccttactg gcttccgcag aggtcaatcc gaatatttca gcatatttag caacatggat	4560
ctcgcagata ccgtcatgtt cctgtagggg gccatcagat tttctgatct ggtcaacgaa	4620
cagatacagc atacgttttt gatcccgga gagactatat gccgcctcag tgaggctcgtt	4680
tgactggacg attcgcgggc tattttttacg tttcttgtga ttgataaccg ctgtttccgc	4740
catgacagat ccatgtgaag tgtgacaagt ttttagattg tcacactaaa taaaaaagag	4800
tcaataagca gggataactt tgtgaaaaaa cagcttcttc tgagggaat ttgtcacagg	4860
gttaagggca atttgtcaca gacaggactg tcatttgagg gtgatttgtc acactgaaag	4920
ggcaatttgt cacaacacct tctctagaac cagcatggat aaaggcctac aaggcgctct	4980
aaaaaagaag atctaaaaac tataaaaaaa ataattataa aaatatcccc gtggataagt	5040
ggataacccc aagggaagtt ttttcaggca tcgtgtgtaa gcagaatata taagtgtgt	5100
tccctggtgc ttcctcgtc actcgaccgg gagggttcga gaaggggggg cccccctt	5160
cggcgtgcgc ggtcacgcgc acagggcgca gccctggta aaaacaaggt ttataaatat	5220
tggtttaaaa gcaggttaaa agacaggtta gcggtggccg aaaaacgggc ggaaaccctt	5280
gcaaatgctg gattttctgc ctgtggacag ccctcaaat gtcaataggt gcgccctca	5340

tctgtcagca ctctgcccct caagtgtcaa ggatcgcgcc cctcatctgt cagtagtcgc	5400
gcccctcaag tgtcaatacc gcagggcact tatccccagg cttgtccaca tcatctgtgg	5460
gaaactcgcg taaaatcagg cgttttcgcc gatttgcgag gctggccagc tccacgtcgc	5520
cggccgaaat cgagcctgcc cctcatctgt caacgccgcg ccgggtgagt cggcccctca	5580
agtgtcaacg tccgcccctc atctgtcagt gagggccaag ttttccgca ggtatccaca	5640
acgccggcgg ccggccgcgg tgtctcgcac acggcttcga cggcgtttct ggcgcgtttg	5700
cagggccata gacggccgcc agcccagcgg cgagggcaac cagccgaggg cttcgccctg	5760
tcgctcgact gcggcgagca ctactggctg taaaaggaca gaccacatca tggttctgtg	5820
ttcattaggt tgttctgtcc attgctgaca taatccgctc cacttcaacg taacaccgca	5880
cgaagatttc tattgttcct gaaggcatat tcaaatcgtt ttcgttaccg cttgcaggca	5940
tcatgacaga aactacttc ctataaacgc tacacaggct cctgagatta ataatgcgga	6000
tctctacgat aatgggagat tttcccgact gtttcgttcg cttctcagtg gataacagcc	6060
agcttctctg tttaacagac aaaaacagca tatccactca gttccacatt tccatataaa	6120
ggccaaggca tttattctca ggataattgt ttcagcatcg caaccgcatc agactccggc	6180
atcgcaaact gcacccggtg ccgggcagcc acatccagcg caaaaacctt cgtgtagact	6240
tccgttgaac tgatggactt atgtcccatc aggctttgca gaactttcag cggatataccg	6300
gcatacagca tgtgcatcgc ataggaatgg cggaacgtat gtggtgtgac cggaacagag	6360
aacgtcacac cgtcagcagc agcggcggca accgcctccc caatccaggt cctgaccgtt	6420
ctgtccgtca cttcccagat ccgcgctttc tctgtccttc ctgtgcgacg gttacgccgc	6480
tccatgagct tatcggaat aaatacctgt gacggaagat cacttcgcag aataaataaa	6540
tcctggtgtc cctgttgata ccgggaagcc ctgggccaac ttttggcgaa aatgagacgt	6600
tgatcggcac gtaagagggt ccaactttca ccataatgaa ataagatcac taccgggcgt	6660
atTTTTTgag ttatcgagat tttcaggagc taaggaagct aaaatggaga aaaaaatcac	6720
tgatataacc accgttgata tatcccaatg gcatcgtaaa gaacattttg aggcatTTTca	6780
gtcagttgct caatgtacct ataaccagac cgttcagctg gatattacgg cttttttaa	6840
gaccgtaaag aaaaataagc acaagtttta tccggccttt attcacattc ttgccgcct	6900

gatgaatgct catccggaat ttcgtatggc aatgaaagac ggtgagctgg tgatatggga	6960
tagtgttcac ccttgttaca ccgttttcca tgagcaaact gaaacgtttt catcgctctg	7020
gagtgaatac cacgacgatt tccggcagtt tctacacata tattcgcaag atgtggcgtg	7080
ttacggtgaa aacctggcct atttcctaa agggtttatt gagaatatgt ttttcgtctc	7140
agccaatccc tgggtgagtt tcaccagttt tgatttaaac gtggccaata tggacaactt	7200
cttcgcccc gttttcacca tgggcaaata ttatacgcaa ggcgacaagg tgctgatgcc	7260
gctggcgatt caggttcac atgccgtttg tgatggcttc catgtcggca gaatgcttaa	7320
tgaattacaa cagtactgcg atgagtggca gggcggggcg taatTTTTTT aaggcagtta	7380
ttggtgccct taaacgcctg gttgctacgc ctgaataagt gataataagc ggatgaatgg	7440
cagaaattcg atgataagct gtcaaacatg agaattggtc gacggcccgg gcggccgcaa	7500
ggggtttcgcg ttggccgatt cattaatgca gctggcacga caggtttccc gactggaaag	7560
cgggcagtga gcgcaacgca attaatgtga gttagctcac tcattaggca cccaggctt	7620
tacactttat gcttccggct cgtatgttgt gtggaattgt gagcggataa caatttcaca	7680
caggaaacag ctatgaccat gattacgcca agctattttag gtgagactat agaatactca	7740
agcttgcacg cctgcaggtc gactctagag gatccacagg ttcttctccc ggggtccaccc	7800
ccggttccgg acgccgtacc ggccgacgat cctgctcggc gtcgccatcg cgatcctcgc	7860
cggcttcacc ccgctgaacg agctcgcggc gctgggtgaac atcggcaccc tgttcgcctt	7920
cgtgatcgtc gcgatcagcg tgatcatcct ccgcaggacc cggcccgacc tgccccgcgc	7980
cttccgcacg ccctgggtgc ccgtgctgcc gatcgtctcg gtcgccgcgt ccctgtggct	8040
gatgctgaac ctgccggccg agacctgggt ccgcttcggc atctggatgg cggtcggcgt	8100
cgtcgtctac ttctgtaca gccgcaaaca cagccgtctg gccgaggagc gcggcgggga	8160
acggacgtcg tcctgaggcc ggcgctccc ggcgctaccc gtcccgccgc ggcgcccagg	8220
cgccgcccgc gggccgcacc gcgcgcgggc cggtgacgtc agcgcccagg gacgtcaccc	8280
ggcggcgcag ctcgcggtcg gccgtgacca ccaggaccgg gcggtcggcc gcctccgcca	8340
ccaggtcgac catgcggtcg tcgccgtgc ccggggccgg gtccacccgg acaccgggga	8400
ccgactccac cccgcgggcc gccccctcgg tcaccaggac gatctccacc gggcccgggt	8460
gacccggcac gccctccgcg gccagccgggt cgcgcagccg ttccgcggcg ccccgccgggt	8520

cccgccacca tccgtcgggc accgacccga ccacgttggc ggcgtcgacg atcacgagca	8580
ggccggtgtc atccatggcg tcagggtccc acgcggcgcc ccggccgtcc gggtcgcggg	8640
ggtgcgaccg cgcgcggggg aggacgcccg ggaacacatg aacgatcgac atatgatgtg	8700
tcgagtttca gcgttcccgc gtcggtgacg cggcgggccc caggaagggg cacctggcga	8760
tgcggagaag gtcggcgggc cccgtgggcg cctccgtcaa ggacggacgt gccgccgggg	8820
agcaccggga ggccgcgtcc ggggccgccc acggtgactg gctcaccgc ggcaaggacg	8880
gccggctgac gctgtacgtc cccaccgacg gcggtctgct gcgctggacg gagaccgccg	8940
tgggtggccc cggctggagc ggtccgcact tcgtcccggg ggccgggctg acgcacctgg	9000
cgggtggctca gggagccaac ggctacgtcc acttcctcgg ccgcagggag cgcgagggcg	9060
ccgactccac gccgggcgtg gacatcgtgc acgcgatcca gtaccagacc ggactcgcct	9120
tcagcgactg gcggtccctc ggcaaccgc accgggtccc ggaggagccc ggaccgctcg	9180
ccgtgccggt cggggcggtc gcccgggacg gcaccgtgca cgtgttcgtg cggggcgcg	9240
acggagggct gatgctgcgg cgcgaggccc cgaacggcaa gtggaaggcg tgggaggacc	9300
tgggcggcgg cggcgccggc gcccagcccg cggcgctcgc gctcaccgac gggcggatcg	9360
aggctctcgt ggccgccgag acgggcgtgc tggcgtggag ccagtccaag cccggcggcg	9420
acttcaccgg gccccggggt ttctcgctgc gccccgtgcc gggcacggtc gcggccctgg	9480
agaccggtcc gggccgggccc acgttcttct ggacggacgc cgagagcggc ggtgcggcgg	9540
cctggcgggc gggggcgtgg cccgtcgcg cgtggcggtac cccggccgag cggccgtgcg	9600
cggtgctccg cacgtcgctg gacggctacg actgtgtcgt cctcgcctac cgtgaccagg	9660
acggcacggc cgtgctcggc atgggcggca cggagaacga ggccgccggc ttctggtggt	9720
acgcgctgac cgagtcctgc cagggcgctc cggccctggc cctggacggg cggggccgcg	9780
tggatgatggc gctgatcggc gccgacggca ggcccagggt cggccgccag gaggacggcg	9840
acggcctctc gctcaccgg tgggacgtcc tcgggggctg agcgcgccgt cgccgtgccg	9900
gaggtcagcc gcccgtcgcg ggcgacttct tcgcggactc ggggatctcc gcgtcggagc	9960
ggatcgcctt ccacagctgg tcggcctgcg ggtgcgcggc caccaccggg ttcggtcgg	10020
tcttgtcgta cgccacgggg agcatgacgg tctccatgga cgccgggtcg acccgttca	10080

tgctgcgcgc gaagtccgcc aggctggtca gcgaggccag ctcgagtcg gtggtcaggg	10140
ccgaggtcag ggtgtcggcg atcttgtacg tcttggtggg gctgccgaac aggtcctgct	10200
tcttcacctc gctcagcagg gcgatcatga actgctgctg gaggcctatg cggcccaggt	10260
cgctgccgtc gccgacgccg tgccgggtcc ggacgaacgc caggagtcg gtgccgtcca	10320
gcttgtgcgt cccggcgctc aggtccaggc cgctcgtgct gtccttgatg ggctcgtcga	10380
cggtgaccgt gacgccgccg atcgcgtcga ccagcccctt gaagccggcg aagtcgatct	10440
ccatgtagtg gtccatgcgg acgccggaca tcttctccac ggtcttgacc acgcaggccg	10500
ggccggccgt cgagtacacg gagttgaaca tgacgcgctc ggcctgcgga agggcggagc	10560
cgtccgcctt ggtgcactcc gggcgggtga cgagggtgtc gcgcgggatg ctacggcga	10620
cggcctgccg gcggccctcc gggatgtgca tgaccagcgc ggtgtccgac cgggcgcccg	10680
ccaccttgcc ggtgcccagg ccggcgttgt ctccggcacg cgagtccgag ccgaggacca	10740
gcacgttctg tccggaggtc ggcagcttct cggggcggtc ctgccgagg gcctcgtcca	10800
ggtcgacccc gtcgatgttc ccgttcaggt cgctgtagag ccagtagccg gtgcccccg	10860
cggcgagcac gacggccagc agggacagca ggacgatgcg tccccagcgc cggcgcttcg	10920
gccggcgggc tccgccgccg gactcgggct gcgcacgccg ggactccgtg gtggcgctgt	10980
gcgtcatggt tcttggttcc ctctcctcac gggctccgcg cggcggagtg gggcccgggt	11040
ggggggagac cgggtccacag atgcgttcac gtcagtggga gaactatagg cagcgggtcca	11100
cggcacatcc tgcacgggtg gaattagcac ggctcagcga tgaatgccac atgaaccgac	11160
cttgaccact ctttaaggctg gaataagatg tgtcgggcct gtgaccgtcg tggagggggg	11220
ctttcacgag gtcggagggt ctggtggggc ggagcggccg gaagccgcga agcctcacat	11280
aaggggtggg cgtccggtgt ggacgatttg gtgtcccggc gggcgaaatg tcttggtagg	11340
agtacgagtt ggtgggttcag ccgcgtctga gtattgtcgt gcccttcag gacgtcgagg	11400
tgtacctcgc cgagtgtctg gaatcgatcg cgcggcagtc gttccgcgac ttcgaggtca	11460
tcctggtcga cgacggctcc accgacgggt ccgtgcggat cgcggcggac ttctgcgccg	11520
ccgaccgccg tttccggctg gtccgccagc acggccacgg accgggccac gcgcgcaaca	11580
ccggactgcg gaacacgcac cccgcggcgg agttcctcgc cttcgtggac ggtgacgacg	11640
tcatccccga gtacgcctac gaactcctgg tgcgcacgct cgaggagtcc gagtccgact	11700

tcgtctcggg caacgtgcag atgatgaact ccaccaagaa gtggcagtca ccgctgcaca	11760
agggccccat gcagaagaac cggcgcggga cgcacatcac gaagttcgac gcgctgatct	11820
acgaccgcac cgtctggaac aaactcttcc ggcgctcctt ctggaaccag aactccatca	11880
ggttccccga aggcgtgctg tacgaggact cgtgggtcaa catgtacgcc cacttccgcg	11940
ccgccaaggt cgacgtcatc acggacgtcg tctatttctg gcgccgccgg gacggcggag	12000
cggcgccctc catcaccag cgccactccg aactgtcgaa cctccgggac cgggtcgcgg	12060
ccgtgcagtc ggtgagccgc ttctcggcg accggcgctc gcgtgagtac gcggacagca	12120
agcggaagta cgatctcgcc tgcctgaagt ccgacctcct gctccatctg aagggtctcc	12180
cggacgcgga cgaggagtac cagcacgcct tcatgaagtg ggccaacgag ttctcgcacg	12240
agacggatct caccatcatc gacgagctgc ccgcggactc ccgctcaag tggctcctgg	12300
tgcgcgagga gcggtggcc gaactgctcg aggtcatcga gttcgagcgc cgcggcggtc	12360
ccatgcccgt gcagcggcgt ttccggcgct acctgaacta cccgtacctc ggggaccggg	12420
gggtgggcct cgacaagaag gcctaccggc tggacaagga gctctcgctg cacggctcgc	12480
tgtccggagc ccgctggagc accggctccg acctgtcac cctcaccgga acggcgtacg	12540
tccgcttcat caacgtgcac aagaagcaca tgtcggtgaa ggcgatcgcc ctgcggaaca	12600
agaagcaggg gcgcatgcag atcacgacgg cgaagaccgt ctacgcgccg caggcgaccg	12660
aggacagtaa gcagaatcgt tactgctatg actgggcccgg cttcgaggcg cgcacgcaca	12720
ccacccgcct caagcgcaag ggccagtggg tcgagggcac ctgggacgtg gccgccgggtg	12780
tcctcagccg gggactgttc cgctaccggg gcatcgaccg gggcggcgcg ggcagcgccg	12840
ccaacccgcc ctaccgctac gtcgacaaga acacccgcat cctcccggtc ttctccagg	12900
gcaaactcaa gctgcgcgtc gagatcgtgc gctgccggat caccaagcac cgtgtcgtcg	12960
gcgaccagct ggagctgcgc ggcgtctacc tcggcccaa ggtcccggag tggggcaagc	13020
tccgcgtcac cagcatgagc ggcgcgggac gccacgacgc acgcgtccac ttcacccgg	13080
gcggtgaggg ctggtgcacc ttctccgcca agctccccct gagccgtctg gtgccaagt	13140
cccgcgtcca ggcggaacc gacgcggacg tcccgcagtc ctggggcatg ggcagcaacg	13200
gctggaagac caccttccac gtcgagggcc gcaagtcggc catctatccc gtgatggcgg	13260

aggagacccc ggacgggcac tactccatgc cgtcctccct gcagaccccc gagcgcgacc 13320
 gggagatcgt cgtgcaccgc aacggctccg gctatctcgt gctcttcgaa cgagcgaccc 13380
 tgcccctggc gacccggtgc gactggcagg aggacggctc gctgtggatc cagggccggtt 13440
 acctggccgc ggaccagctg accccggagc agtaccgctc cgcccacctc gtggtgcgct 13500
 cgcgcgccca cggcgcgga cgtccgtac cgctcacctg ggacgggcac gagttccgct 13560
 gcgtcctggc ccccgccgcg atgcggaccc tggccgggga catcccgctg gcggccggac 13620
 ggtgggactt cttcctgcgc cgccaggacc tgtcggccgt ggcccgcgag gaccggctcg 13680
 aagacctcat ggtgaagatc gagcaggatc tcatcgaggc gttcccgag gagtacgaga 13740
 gaaacgaacg ccgctacgag acgcaggccg aggcctacga ccggctgtcg ctgctcgtcc 13800
 actcggcgat gcccgaccac gcccgcggcc cctaccggca gaagctcctg aggaccaagg 13860
 cctaccccgga cgcccggcgc cggccggtgc gtgacgccgt gctgttcgac gccttcaagg 13920
 gcacccagta ctcgacagc ccccgcgccc tgcacgagga actcgtgcgc cgccgcaccg 13980
 gcctggaaca cctctgggtg gtgcgcgacg accagggtgca ggtgccgccc acggcgacgc 14040
 ccgtccgcat gtggtcgccg gagggttacg aggccctcgc caccagccgc tacgtcgtcg 14100
 ccaacaacca cctcccggac tggttcaaga agcgggacgg acaggtcgtc gtgcagacct 14160
 ggcacggcac gccgctgaag aagatcggcc acgacatcga gtccatccac ttcgccgacc 14220
 agcgctatct ggaacgcgtc gagaaggagg tgcagaactg ggacatgctg gtgtcgccca 14280
 acagcttctc caccgcatc ctcaagcgcg ccttcggctt ccccggcgag atggtggaga 14340
 gcggctaccc gcgcaacgac atcctgcgcc ggccggacac cggggcccgg gagcaggaga 14400
 tccgccgag catcgggctg ccggagggca agcgggtggt gctgtacgcg ccgacctggc 14460
 gcgacgacca gttctacgcg cccggcaagt acaagctgga cttccggatc gacctggccg 14520
 ccgcgcgtgc gcagctcggc cccgaccacg tcctcatggt gcgccgccac cccaacgtcg 14580
 tggacccggt gccgggcgcc ggcgacggat tcgtcttcga cgtgtccgac taccgggaca 14640
 tggccgacct ctcgctgatc accgacgtga tgatcaccga ctactcctcc ctgatgttcg 14700
 actacgtgaa caccgggcgg cccatcctgt tcttcaccta cgacctggac cactaccggg 14760
 acaccctgcg cgggttctac ttcgacttcg agggcagcgc gccgggcccc ctctctaca 14820
 cgtccgagga actggtggcg gcgatccgtg acatcgacgc catccaggac ctctacgccg 14880

agcgggtaccg ctggttccag cgggagttct gcgacctgga cgacggttac gccgcggccc	14940
ggctcgccga ccgatgctg gtcgcgggcg gcgacctcgc ccccgggcag gcgcacgcgc	15000
cggccgtcgg cgcggtcgac acccggcaca ccggaaggcc gatgaccccc ctccagtggg	15060
ggaactcgga gtggttcgcc ggccccgcc cgccggcggg tctcgtcgac gccgtgccc	15120
cccagccgc cccggcgtag gacgccgtac cgcagcacca ggcgggtccg ttcggccata	15180
ccccgcccgc cggcgaccgc agctacgaag gcgtgatcgc gtgacccgc cgacccgc	15240
cggcgcgagg tccccggcgt gccgcgcggt gccggaaggc cccggcccc gatgaccgc	15300
gtgccgttcg ccacgcgggc gccggccgc gctgccggtc cgtcgtccgg cggccgccc	15360
ggcgcgtcct cagccggtgc cgggtgtcac ccgatcagc agcgcgcggt ggtcggagac	15420
cccgtgtcg ctcaccggc acccgagcac gggcagccc gtgaagagg agtccagctt	15480
gtggtgcgag acgtgcgtcg gccgggccgg ccggagggga ccgggcgtcc cgtcgcactc	15540
ccggtgcgtg ccgtaaggct ggtcgggcca gaccgggag agcgggttgc gctctcccgg	15600
cggatccacg ttgaggtcgc cgccgtagac ggtgcgccgc tcgggcaccg cgtccaccag	15660
ggccttcagc tgtccggcgc ggaactcgc gtccggatgc gccagatcgc cgccgcgcgg	15720
ggtcagatgc gcggtgcaca ccgtgaggtc gtgcgccgc acgaacgcgc agagtattcc	15780
gcgctgcacc ccgaccgcgg gctggggcgc gggcaccgc cgcacggacg acagcggata	15840
cgccgacagc agggcgtagc ccgcggagcc ccggccgggc gcccgcagc gcaccgcggt	15900
gcggcgccc tcacggccgc gccaggtgta ggccctgaac tccgcgtgcc acgacgccc	15960
gagggaggcg cgcaccgcct cgacgtccgc cgcgcaggtc tcctgcagga acagcaccg	16020
ggccccggat tccgcggcga ggcgttccgt ccggccgcgc ttggcgtcct cgccgcccgt	16080
gccctgcag ttccattccc tgaccccgca catgttccag gtcgccacgg tcagtgtccg	16140
gtccccggtg gcggagcggc ccgctccatt gccgtccgc tcgtgggtca ccacggcggc	16200
gagccccgc agggcggcgg ccgccgtgac ggcggcgagc aggcgccgtc cgcgcgggcg	16260
ggcgaccgg gttcggttcc tgagcaccg gccatcatga ccgatccgc gctgaccagc	16320
caaaaccacc cggcgccagg cgcttccgg caggtgtcct cccctgccc gcgcaccgc	16380
ccgtcgtcca agtacctctg agagtggagt ccgtacatgt ccaaggcacc ctcgaacggg	16440

cggcagctgc tcaacggcat cgaagcctcg ggaacgttcc cggaggagta ccggttcacc	16500
cacgccaaga gcggcaaccg gcacctcgtg gtggtcttcg ccaacttctc ggcacccgag	16560
gactacggct ggtcgaacgg cgtcttcgac aacgtccgtg ccaacatcct gtggatccgt	16620
gaccggttcg acgggatgaa cgcctactac ctgtgccgga acatggactt cgggtctggcg	16680
gactcggtagc agaccctgat cgcgaacgtc accggggcgc tcgggctgac gccggaccaa	16740
gtcacgctct ggggcggctc caagggcggc agcgccgcgc tgtacttcgg cctgcggtac	16800
ggctaccgga acatcgtcgc catcgtcccg cagttcctca tcggcgacgc cctggagaag	16860
cggcacccga aggtctccgc gtacatgctc ggcgaaaggg cgcaggcgca caacgcgcgg	16920
atcctggacg cgctcctgcc cgacctggtg cgcgccaagg ccaacccggg cgccaacatc	16980
tacgtgctct cctccccgca ggacgagcat tacgccgtgc aggtcgagcc gttcctcggc	17040
atgttccacg gctacgagaa cttcaatttc ctgtacagcg agtcgccgac catcacgggg	17100
cacgccacgg cgaccggcg gaacgtcccg gcgtggtagc gcctgctcaa cctgctcgcc	17160
gacggctacg ccccccggct gggcttcacc cgccacgccg ccgaggactt cgaccacgac	17220
cggtcggaca tcaacgccta cctcgccctg acctccaagg tccagggcgc cgacgcgttc	17280
gcgccgccgg tggtagaccac cccgggcttc aacagcgagg tcccgcgcac cggaccgtgg	17340
ttcaccggga cggccacagg agcggtagcg gtgagcatgt ggcgcaacgg caagttcgtg	17400
gcgtcgcccc aggtcgcggc cgacggcacc tggctcctggc agccgaccgg gccgtgggag	17460
gccgggaagc acatcgtcaa gatcttcgcg gtggacccgg cgggcttcca ctccgcccg	17520
gtcgagatcc ccttcaccgt ggtcgaccgg gatcccgtcc ctgccccgcc ggtcgtctcc	17580
gcaccggtgt ccgggcagca gaccggagcg gcggtcgggt tccacggcag cgcgccggga	17640
gcgtcacagg tcggcttcg ggagaacggc gtgctcctcg gcgcggtggc cgtcgcgccc	17700
gacggcacct ggggctggga ccccgcccg ccctggcccc aggggcagca cctggtcgag	17760
atcgtcgcg tcgacgcgta cggcatggag tccgcgcccc cgccgccgg cttcaccgtg	17820
ctcggccacg cggtgccgc cggacattc acgccgcggt actgaccgac ggcccaggac	17880
gacgacacac cacgagtggg aagcagacat gccgaaagaa gcgccgacaa cacgcgagct	17940
gatcaccggg atcgacacct ccggcgcgta tccgctcgag taccggttca cgcacgcaa	18000
ggggggcaac cggcacctcg tcgtcgtgtt cgccaacttc gcggtcaagg acgactacgg	18060

ctggtccaac ggcgtcctca acccggtgcg ggccaacatc ctgtggatcc gtgaccggtt	18120
ccgcgacatg aacagctact acctgtgcga ggggatggac ttctccctgg agcagtccgt	18180
gatcgggctc atctccaagg tgatgaacgc cctggaactc accccggagc aggtcacgat	18240
gtggggcggc tcgaagggcg gcagcgccgc gctctacttc ggcatgcgct acggcttcgg	18300
caacatcgtc tccatcgtgc cgcagttcct cgtcggcacc tatgtgaagc ggggtgcaccc	18360
caaggttgcc cggttcatgc tgggcgaggc ggtgccggag gagaacgtcc gcgcggtcga	18420
cgcgctcatc ccggacctgg cccgttcggg cgtcgcccgg cactccaaca tctatctgct	18480
ctcctcgccg caggacgagc agtaccagga gcaggtcgag ctttctctcg gactgttcca	18540
ggggtacgac aacttcaact tcgtgttcag cgagtcccc cacatcacc gtcactcgga	18600
cgtcaccgg cgcaacgtcc ctttctgat gggcctcgtg aacatgctcg ccgacgggat	18660
gtccccgcgg ctgggcctgg tgcgcaacgg gtacgaggag ccggaccgcg acaggtccgc	18720
catcgagggc ttcttggcgg ccacttcggc ggagcggccc agcgccatcc cgatgcccgt	18780
ggtgacgcat ccgcttcgcg acatggaact gcccacggac ggcgtgtact tcacaggac	18840
ggccccggc gcggtgcggg tgagcctgtg ggagcacggc aagttcctgg gttcgccgtc	18900
ggtggcgccg gacggcacct ggtcctggaa gcgggacaag ccgtggagca agggcgacca	18960
tctggtcaag gccgtcggct gggacgcgga gaagcgccgc accaagggca ccgtggtccc	19020
gttcaccacg gtcgccggcg cgaacgccgc cgcgcccggg gcaccggccg ccgcgcccct	19080
ggcgcccggg cagccgctgc cggcgccgac ggtccacacg ccgggggctg acgagcagat	19140
caccggcacg gccgtgcgct tcagcggctt cgcggggg gcgcccagg tgggattcag	19200
ggcggggggc accctccttg gcacgagccg ggtcgcggcc gacggaacgt gggcctggga	19260
ctccggctgg ccctggcagg cgggcatgca caccgtggag gtgttcgccg tggacgccgc	19320
gggatccgag tcgcccgtgg cgccggtgcc cttcgacgtc atgcacgcca cggcgggcgc	19380
ctcgccgttc gcctactgac cggtcgccac gcgcggaagg gtcctggag atcgccggcg	19440
atctccagga gcccttcgcg ggtctacggg gactaggcgg gcacgctcgc cgtgcccggc	19500
tccaggaacc gtttcccgtt caccgcctcg gagacgccct cgcggtccag gtacggcgtg	19560
atgccgcca ggtggaaggg ccagccggcg ccggtgatca ggcacaggtc gatgtcctgg	19620

gcctcggcga cgacgccctc gtcgagcatg agcccgatct cctgggccac cgcgtccagg	19680
acgcgggcg c ggacctgctc ctcggtgagg acggtgtcgc cctgcttcag gagcgcggcg	19740
acctccgggt ccagctcggg cttgccgctg tcgtagaggt agaagccacg cttgccggcc	19800
tcgacgaccg ccctgagggt cggggagacc gtgaagcgct ccgggaacgc cctgttgagg	19860
gtctcggaca cgtgcagacc gatcgcgggg ccgaccagct ccagcagcac cagcggggac	19920
atcggcaggc cgagcggctc gacggccttc tccgcgacct cgaccgggggt gccctcgtcg	19980
atgacgttct ggatctcgcc catgaagcgg gtcaggatgc ggttcacgac gaacgccggg	20040
gcgtccttga ccaggaccgc ggtcttcttc agcttcttg cgacaccgaa cgccgtggcc	20100
agcgccgcgt cgtcgggtcc ctcgcccggg acgatctcca ggagcggcag gatcgcgacc	20160
gggttgaaga agtggaagcc gacgaccgc tcgggggtgct tcagcttcga cgccatctcg	20220
gagacggaga gcgaggaggt gttggtggcg aggatcgcgt gcgccggggc gaccgcctcg	20280
acctccgcga acacctgctg cttgacgccc atctcctcga agacggcctc gatcacgaag	20340
tcggcgtccg cgaagccctc ggccttgtcc agcacaccgg tgaccagggc cttgaggcgg	20400
ttggccttgt cctggttgat ccggcccttg ccgagcagct tgcgatctc ggcggtggacg	20460
tagcccacac ccttgctgat gcgcgcctgg tcgatgtcgg tcagcacgac cggcacctcg	20520
aggcggcgca ggaacagcag cgcgagctgg gaggccatca gaccggcgcc gaccacgccc	20580
accttggtga ccgggcgggc cagcgacttg tccggggcg cggccggccg cttgccgcgc	20640
ttctgcacca ggttgaacgc gtagatgccg gagcgcagtt caccgcccac gatcaggtcg	20700
gcgagcgctt ggtcctcggc gtcgtagccc tgctggagggt cgccgttctt ggcgggcggcg	20760
atgatgtcca gggcgcggtta ggcgggccgg gcggcgccgt gcaccttgga gtcggcgatg	20820
aagcgccctt tggcgacggc ctggtcccag gcctcgccgc ggtcgatcac cgggcgctcg	20880
atccggatct cgtccttgag gacggccgcg gtccagatca gcgactgctc caggaagtcc	20940
gcgccctcga agatcgcgtc cgcgatgccg agttcgaaga cctgcgcgcc cttgagctgc	21000
ttgttctggt tgaggctgtt ctcgatgatg accgagacgg ctttctcggc gccgatcagg	21060
ttcggcagca gcgtgcagcc gccccagccg gggacgagac cgaggaagac ctcggggagc	21120
gagaacgccg ggagggcggc cgacaccgtg cggtaggtgc agtgcagacc gacctcgacg	21180
ccaccgcca tcgcccgcc gttgtagtac gcgaaggctc gcacggccag cgtcgacagc	21240

cgcttgaaga cgtcgtggcc gcccttgccg atggccagcg cgtcctcgtg ccgcttcagc	21300
agctcgacgc ccttgaggtc ggcgccgacg gcgaagatga acggcttgcc ggtgacgccg	21360
acgccgacga tctcgccgtc cgcggcctcc ttctcgaccc ggtcgatcgc ggcgtcgatg	21420
ttcgccagcg actgcgggcc gagcgtggtc ggcttggtgt ggtcgtggcc gttgtccagg	21480
gtgatgagcg cgaagcgccc ggcgcccagg gggaggtcga agtggcgcac gtgcgcgctg	21540
gtgacgacct cgccggggaa cagctcggcc gcacccttca gaagctctgc ggtggtgctc	21600
acttgtcccc ctcgaagtgc gggttctccc agatgaccgt cgcgcccattg ccgaagccga	21660
cgcacatggt ggtgaggccg taacggacct gcggctgctc ctcgaactgg cgggccagct	21720
gcgtcatcag acggacgccg gaggaggcca gcgggtggcc gaacgcgatg gcgccgccgt	21780
actggttgac gcgcgcgtcg tcgtcggcga tgccgtagtg ctccaggaag gccagcacct	21840
ggacggcgaa ggcctcgttg atctcgaaca gaccgatgtc ggagatggac agccccgcct	21900
gggcgagggc cttctccgtg gccgggatcg ggccgtagcc catgacctcg ggctccacgc	21960
cggcgaagga gtaggagacc aggcgcattc tgaccgggag gccgttctcg cgggcgaagt	22020
cctcgctcgc gatgaccgag gcgggtggcgc cgtcgttcag accggccgcg ttgccggcgg	22080
tgaccgggcc gtggacgcgg aagggggtct tcaggccggc caggttctcc agggtggtgc	22140
ccgggcgcac cggctcgtcg gcggtgacca ggccccagcc ggtctcaccg gcctcctcgt	22200
tggtgcggcg caccgagacc gggaccaggc cggcctggat cttgccgttg gcgtaggcct	22260
tggcggcctt ctcctgggag cgcacggcgt actcgtcggc gcgctgcttg gtgatcgagg	22320
ggtagcggtc gtgcaggttc tccgcggtca tgcccatgaa cagggcggac tcgtcgacca	22380
gcttctcgga gacgaagcgc gggttggggc cgacgccctc gcccatcggg tggcggccca	22440
tgtgctcgac gccgcccgcg atggcgacgt cgtacgcgcc gaaggcgacg gagccggcga	22500
ccgtggtgac ggcggtcagg gcgccggcgc acatgcggtc gatggagtag cccgggaccg	22560
aggtgggcag gcccgcgagg atgccggccg tgcggccgat ggtcaggccc tggtcgccga	22620
tctgcgtggt cgcggcgacg gcgacctcgt cgatcttctt cgggtcgaga ccgggggtgc	22680
ggcgcagcag ctcccggatc gccttcacga ccaggctcgt ggcgcgggtc tcgtggtaga	22740
tgcccttcgg gcccgccctg ccgaacgggg tacggacgcc gtcgacgaag acgacgtccc	22800

tgacggtacg aggcacgatg gctctcctcc caggggtgcgg gacgctgagc gcttgcttac	22860
gccatgctac ttatgagtaa cgtgactgcc cagtcccggc ccccgagcg gcgaacatca	22920
cacgtacggc ggcgcccgcc aaacgccgga ggggctggaa tcagccccct ccggcgtttg	22980
aggagcggga acccctcacg gtggagccgt ggagcctcgc ggggtcagga ccggggtcgg	23040
taccgggtgc gaccccggt gtcctcccc ggtgatgacg ccgaacagcg tgcgcgccac	23100
ctccgccccg aaccgtgca catcgtgact catcgcggac agcgtcggat gcgtcagccg	23160
gcacagctgc gagtcgtccc acgccagcag cgacacgtcg tccggcacc gcagccccat	23220
ctccgcccgc accgacagcc ccgccaccgc catgatgtcg ttgtcgtaca cgatcgccgt	23280
gggccgttcc cccggcgcg gcccagcag cgaacgcgtc gcccgcgccc ccggtcccc	23340
cgagaagtcc gtggccgtct gccacgccc cgcgggcggc tccagcgccc ggaccgcctc	23400
gtcgaacgcc gccgtgcgga tcgaggtgtg cccgagcgcc gcccgcgcc ccacccgggc	23460
gatccgcccg tgcccagcg ccgccagata ccgcacggcc tccgtcacgg ccgtggcgtc	23520
gtccgtccac acggaggtga gcccgcctgt cagcgccggg tggccgacgg ccaccgccgg	23580
cagcccgagc cgctccgcca ccgccggacg ggggtcgccg gcccggaagt ccaccaggat	23640
cgagccgccc atctgcccgc cgcgccacca cgactccatc agcccgaact cctcctccgg	23700
gctgcgcacc agccgcagca gcagcgagca ggaccgctcc accaggacgc tctccacccc	23760
ggagatgaac tccatgtaga acggctccag gccgagcagc cgggcggggc ggagaccgc	23820
gagaccacc acgtccacc gcgaccggc cagcgtgcgc gcggttcggc tcggcgccca	23880
ccccagctcc cgcgccgccc ggaagatgcg gtcccgggtc gcctccgaca gcccgggctt	23940
ccggttgaag gcgagggaca cggcgccctt ggacacgccg gcgcgcgcgg cgacgtccct	24000
gatggtgacg cgaggggtcg gcgttgccgt catcgagtgg gttccacgca gtacagggcg	24060
gaacgggcgg tgtccgggtc cggagtcttc ccccccgca ccccgatggt cacctgttcc	24120
ccggggagca gggtcaccag ccccggtcg gcccgcgccc cggggtccag ccggtcggcc	24180
tggagcagca ggtcccgtac gaggggtgcgg gccgtgaccg tgatcccgtc cggcgcgagg	24240
gcgacctga actccggcgg ggggtagggg atctcccgtt ccggcgccgg gaagtgccac	24300
gcccgcccc cgtccgcgtc ggcgaccagg aactccccgg gccgtccgg cagcagttcg	24360
accgggacct cgaccaggg caccgtccgc ccccggcgt ccagcgccgg ggccgcctcc	24420

gcgatcgggg cgccgtcgac ggacatccgg cgagccgca gcgttccccg ccagtcctcc	24480
gcggactggt tgaccgccgc caccaccaga ccgtcaccgt ccgcgcgcac ggtcagcagc	24540
cggtcgcgt acagccggcg cagctcgtgg tagagcggct tctcccgccc gtccccgtcg	24600
atcgcggccc acgacgtcac cggccagcag tcgttgagct gccagaccac cgtgcccgcg	24660
cacaccggcc agtgcgagcg ccagtgtcg acaccggccg ccaccgcacg cgcctggttg	24720
acctgcgtca gatagtcca gcggtcgaag tcgccctccg gcacggcgaa gtggcgggcg	24780
aggccgcgt ccagcttgcc gttgccgtcc tccgccttct ggtggtgcag catgccgggg	24840
gagtcggcg cggggtcctc cccgggcagc gcccgccgca gcgtggcgtg cgcgggaggc	24900
gcctgccagc cgaactcggc cacgaagcg gggacgtcg gccggtagtc ggcgtagtcg	24960
gcgcggttcc acacctcca ggagtgggtg gtgccgtcg ccggatcggt ggggtggtgc	25020
cgccaggaac cggaccaggg actgcccgc gtgtacggcc gcgtcgggtc cagctccgcg	25080
accacccgcg gcaggacgcc gaggtagtag ccctcgcccc aggagtcccc ggcgagcccc	25140
tgctcccagt ccagtcccg gaacccccac aggttctcgt tgttgccgtt ccacagcacc	25200
agggaggggt gcggcatcag ccgtacgacg ttctcccggg cctccgcctc cacctccccg	25260
cgcagcggct gtcctcggg gtaggcggcg cacgcgaacg ggaagtcctg ccagaccagc	25320
agccccaact cgtcgcaggc gtcgtagaag tcctcgtcct cgtagatccc gccgccccag	25380
accggacca ggtccacccc cgcgccggcc gcctgctcca gccggtgccg gtagcgctcc	25440
cgggtgatcc gggacgggaa cacgtcgtcc gggatccagt tgacgccccg cgcgaacagc	25500
cgtcaccgt tgacgaccag ggtgaacccg gtgccgtcg cgtcggccga ggtgtccagc	25560
tcaaccgtcc ggaacccggt cctgcgccgc caggcgtcca gcgcctcgtc accgtgggac	25620
aacgtcagct cgacgtcgta cagcggctgt tcgccgtatc cgcgcggcca ccacaggcgg	25680
acgtccggca cccggagccg cacggtcccg gccgtcccat cgaccgcgc ccgggcgcgc	25740
acgccccgg cgctgcctc cagggtgagc ggtgcctcga cccgggagcg ctccacgtcg	25800
accgccagct cgatctgccc caccctgtcc tcgacggtga ccagcgggcg caccgggcg	25860
atccgcgccg tcgaccagcg ctccagccgc accggccgcc agatcccggc cgtcaccagc	25920
gtcggccccc agtcccagcc gaacgagcag gccatcttcc gcagggtactg gtacggctcg	25980

gcgtacgctc cggggcgctc gcccagcctg ccgcgccaccg cctccgcctc ggcgtagcgc	26040
gaggcgaacc gcaccgtgag ccggccgctc agtcccgtca cgtcgaagcg gtacgagcgg	26100
tgcattgttc gcgtccggcc cagtggccgg ccgtcgagca ggatctcggc gacgggtgtcg	26160
agaccgtcga agacgaggtc cgtctgctcg tgcggggccc tcccggcggc cagctccgtc	26220
tcgtacgtcc actcccgcg gcccacccag gccacctcgg tctcgttgcg gccgaggaac	26280
ggatcgggga tcagcccggc cgccagcaga tcggtgtgca cacaccccg caccgaggcg	26340
gggagggcgt ccccggtgcc gtccgggtgt cgcaggatcc atccctcggc gagcgggtgtg	26400
acctgacgca tgcacactcc ctaaaccggc tgagccttct ctgaagagtg gtctggcatc	26460
gttggcgcga ttgcgacttt accggttcag ttcagggtcg ccagagtgcc gaatcagcca	26520
tcccactcgt gctcgtccgt cccgtgaacg gagccgtgat gcatctgaac cgccgtacga	26580
cactcaccgg atcgctcgcc ctgctcgccc tcctggcctc cgcctgcacg ggcacggggg	26640
gttcctcgaa gggcgccggac gccaaaggct ccgacgaccc gtcaaaggct aaggggtccc	26700
tcacggctct caccaccgg accgatctgg tgcaggacgg gacgatgaag aagtacgccg	26760
ccgagttcaa cgagacctat cccgggggtga aggtggagtt cgacggcctc accgactacg	26820
agggcgaggt caagatccgt atgaacacgg agaactacgg cgacgtcctc atgatcccg	26880
cggtcgtcga gaagaaggac taccgaagt tcttcgcctc cctgggcacc aaggccgaac	26940
gcgccgcaa gtaccggttc accgactact ccaccgtcga cggcaaggct tacgggcaga	27000
gccccgtcgg cgtcgtcccc gggttcattc acaacaagcg ggtgtggagc gaggccggcg	27060
tcaccgactg gcccaccacc cccgccgagt tcctggacga cctgaaggcg atccggtcga	27120
agaccgacgc ggtgccgtac tacaccaact tcaaggacat gtggccgctg acccagtgga	27180
ccaacgtcaa cggctccgtc ggctgcgacc cgcacgccac cacgaagctc gccgagggcg	27240
acccgtgggc cgagggggcc gacctgcgcg tgggcgacac cctgctccac gacatcgtgc	27300
gcggcggact cgccgagaag gacccgacca ccaccaactg ggagggctcc aagcccaagc	27360
tggccaaggg cgagatcgcc accatgtggc tgggctcctg ggccgtcgtg cagatgcggg	27420
acgcggcgaa gcaggccggc gccgaccccg ccgacatcgg cttcatgccc ttccccgcac	27480
agcgggacgg cacgttctgc gcggtgacct cccgggacta ccagcaggcg gtcaacgtca	27540
actccgacaa caaggaggcc gcccgcgctt ggatcgactg gttcaccgac aagtccggct	27600

acgccgaggc caacctcgcc ctatcccccc tgaaggacgc cccgctgccc gccgtcctcg	27660
agccctacga gaaggccggc gtgaagctcc tggacctcga ggacagcaag ggcgccgagg	27720
tgaagtccct cgacaaccgc tccgaggtcg gcatctacaa gcccgactac cgccaggaac	27780
tcgtcgacct cgcccgcggc gcccgcaagg gcggcctgga cgactacctc ggccggcctcg	27840
gcgagcgctg ggccgaggcg cgcagcgcgc tgggggcctg atgacggaca ccacccgcaa	27900
ggcggcgcgg ccggttcccc cggccgcgcc cgccgggccc ggccgcgcgg cgccggcccc	27960
gcgccgcacc cggctgtcgc gccgcctcac cccgtggctg ttcctggccg caccgctggc	28020
cctgctcctg accttcacct acgcgcccga tcgccaacat ggtcgcgtac agcttcaccg	28080
actgggacgg cgtgagcccg gagctgaact ggacgggcac cgggaactac accgaactcc	28140
tcacccgctc cgagctgttc gaggtcttct tcgtcagcgg ctactacctc gtcgcctccg	28200
cgggtgcagat cgtgctcgcc ctctacttcg ccacggtcct cagcttcgac gtccgcttcc	28260
ggaacttctt caagggcgtg ctgttcttcc cgtacctcat caacgggggtg gccatcggct	28320
tcgtcttctt ctacttcttc caggacggcg gcaccctcga ctccgtactg ggcttgctcg	28380
gcgtcgagac cgaccacgcc tggctgggca cgccgttctc cgcgaacacc tcgttgccg	28440
gcgtctccgt ctggcgctac ctccgactga acttcgtcct cttcctcggc gcgatccagt	28500
ccatcccggg cgagctgtac gaggcggccg agatcgacgg cgcgaaccgc tggcagcagt	28560
tccggcacat catcgcgccc ggcatcagac ccgtgctgag cctgagcgtg atcctctcgg	28620
tctccggctc gctgtcggtc ttcgagatcc cgtacatcat gaccggcggc gccaccggca	28680
cggagacctt cgtgatccag accgtgaagc tggcgttcca gttcaacaag acgggactcg	28740
cctcggccgc cgccgtcgtc ctgctgctga tcgtcctggc ggtcacctgg gtgcagcggc	28800
gcatcgtccc cgacgagaag gtggacctcg tatgaccgc cgtaccgcgg cacgcgccct	28860
ggtcctgacg tccctgatcc tggcgacgct ggtggtgctg ctgccgctcg ccgtggtctt	28920
cctgacctcg ctgaagtcct ccgaggagat ggcaacggc agcggagcgc tgacgccgcc	28980
cgacgacccg ctgaacttcg gcaactacgt gacggcgttc cgggacggcc agatgctgtc	29040
cgcgttcggg aacacggccg tcctcctggt cgtggccgtc ggcggaacga tcctgatcgg	29100
ctcgatgacg gcgtacgcga tcgaccgctt ccggttccgc ttcaagaagc tggctcgtggc	29160

gctgttcctg ctggccgcgc tgggtccccgg ggtgaccacc caggtggcga ccttccagat	29220
cgtaacagc ttcggcatgt tcgacagcct gtgggcgccg atcgccctct acatgggcac	29280
ggacatcgtc tcgatctacg tcttcctgca gttcatccgc tccatccccg tctccctgga	29340
cgaggcggcg cgcctggacg gcgccaacgc gttcaccgtc taccgcaagg tgatcttccc	29400
gctgctcaag ccggcgatcg cgacgggtggt gatcgtaaag gggatcaacg tctacaacga	29460
cttctacatc cccttcctct acatgccctc cgaggacctg ggggtcatct cgacgtccct	29520
gttccgcttc aagggtcccct tcggcgcgca ctgggagacg atctcggcgg gcgcggtcct	29580
ggtcatcctg cccaccttga tcgtcttctt gttcctccag cgcttcatct acaacggggtt	29640
catgcggggg gcgacgaagt agccagcgcg gccaccagca cgggagtgac ctggtcgacc	29700
tgccacgcgc gtgccccgtg cgccgtcagc gccgcggcga cggaccgctc gtcgggcccc	29760
ggcggctccc agcagaccct gcgcaccgtg tccgggggtga tcaggttctc cggcggcatg	29820
ttcagccgct cggcgagttc ggcgaccccc gcgcggggcg ccgacagccg ggccgcggca	29880
acggggctct tgtccgcca ggcgcgcggc ggcgaggggc cggtcaccgg ctggccgggc	29940
tgccgcagct gggcctcgct cagcgcttc gcgcggtcga cggccgcctg cactgctcc	30000
agctggcgcc gccccacccg ctgcccgaac ccgttgagcg cggccatggc gtgcaggttg	30060
gcgggcagcg cgagcgcggc ctccacgatc gccgcgtcgg aaagcacctt gccgggggag	30120
acgtcacggc gccgggcgat ccggtcgcgg gtctcccaca gctccgcac caccgccatc	30180
tggcggcgcc ggcgcacctt gtgcatgccg gaggtgcggc gccaggggtc cttgcggggc	30240
tccggcggcg gggccgaggc gatcgctcg aactcctgcc gggcccagtc cagcttgccc	30300
tggcgggtcca gctccttctc cagggcgtcc cgcagatcga ccagcagttc gacgtcgagg	30360
gcggcgtagc gcagccaggg ctccgggcagc ggacgggtgg accagtcgac ggcgaggtgg	30420
cccttctcca ggacgaagcc gagcacgttc tcgaccatcg cgccgagccc gacgcggggg	30480
aacccggcaa ggcggccggc cagctcgggt tcgaagaggc gggagggcac catgcctatc	30540
tcgcgcagac agggcaggtc ctgggtggcg gcgtgcagca cccactcgac gccggacagc	30600
gcctcgccga gggcgagacg gtcggggcag gccacgggggt cgatcagcgc ggtacccgca	30660
ccctcgcggc gcagctggac gaggtaggcg cgctggccgt agcggtagcc ggaggcgcgc	30720
tcggcgtcca cggcgacggg tccgctgccg gccgcgaagg cggcgaccgc ctcggcgagg	30780

gcggcctcgt ccgctatcac gggcggaatg ccctcgcggg gttccagcaa gggggtcggc	30840
gcctccgtaa cagaagatcc gccgtcgtcc ggaggagcgc ctccggtggg gcgcagtga	30900
ctgtctgctg cgggtgtcgtg ggcgtcggtc acctgtcaag ggtatccgtg ccgcgaaggc	30960
gcccgtcgac ggttgtgctc cgtgacgggc gccgggtggg cgtattccgg tcagaagagt	31020
gaaagaacgt gttcgttgg ccgtgggcgg gcggatcggg gacgggcgga tcgggggcgg	31080
gacggaaggg tcagtggatg atgccggtgc gcagggccac ggccaccatg ccggcgcggg	31140
cggcgtgcc gagcttgcgg gcgatccggg ccagggtggc cttgacggtc agtgcggaca	31200
ggcccatcga gacgccgatc gccttggttc actggccttc cgccaccagc cgcagcacct	31260
ccacctcgcg gccggacagc tcgcggtagc cgccgggtg gctcggggca cccggggggc	31320
ggcgggtgcag gcgcgcggcg gcggcgccga tgggagcggc acccgccgg gtggggagcc	31380
cgaggttggg gcgggtgccg gtgacgacgt agcccttgac tccgcccgcg agggcgttgc	31440
gcacggcgcc gatgtcgtcc gccgcggaga gggcgaggcc gttgggccag cccgcggcgc	31500
gggtctcgga gaggagggtg aggccggaac catccggcag atggacttcg gcgacgcaga	31560
tgctcggggg gttgccgatg cggggacgag cctccgcgac ggacgaggcc tcgatgacgt	31620
cgcgtacgcc gagcgccac aggtggcggg tgacggtcga acggacgcgg gggtcggcca	31680
cgaccaccat ggcggtcggc ttgttcgggc ggtaggcgac caggcttgcg ggctgctcga	31740
ggagaacgga caccaggcct cctgggggtgc gggacgggcc ggctcgtggg ggtgaaggcg	31800
ggacgaaccg tgctttcaag gtcacagtcg tcttcggcag caaacctggg gtcctttaac	31860
gaatgatcac gaagtgatga gtaacaatcc gggcaattcg gacgcacgat cgatcattcg	31920
aagatcgaac ggtttcggtc tgcgtcgaa cgcttccgaa agtggccgta tcgacaaaga	31980
gagatgcagg aggccggtcg tcgggacccc gcagcgggag gtcagcgcg actgcggccc	32040
cctccgctgc ggcagcgtca ccacggacgc gtccccgga gcggccggcg gcagccccgc	32100
gacctgcgcc agcagatcgg accacgcgac cagatgggcg gccgtgtccg gaacccccgc	32160
cagaccctca cgggcggtcc acgaggcacg gatctcgatc tgggaggcgg cgggccgcgc	32220
ggacagcccc ccgaagtagt gcgaactcgc ccgcgtcacc gtgccgctcg gctcgccgta	32280
cgacaggccg cgcgccgcca gcgcgccggg cagccaggac cagcacacgt ccggcagcag	32340

cggatccgcc gccatctccg gctccagctc ggcgcgacc agcgtcacca gacggaaggt	32400
ccccgccag gcgtcgtgtc cggccgggtc gcacagcagc accagccggc cgtcggccag	32460
atcctcctcg ccgtcgacga ccgccgcctc cagcgcgtgc gcgtacgggg cgagccgttt	32520
cggcgcgggc accgtctcca cctcgatctg cggccgcagc cgggcgctct gcagcgctc	32580
gacagcggcc cggaagggcg gcggaggcgc acctccgtgc ccgggatccc ccccgccctc	32640
cttcggttcg tccattccgc cagcgccgtc cgacagtcgt ccctgagccg cagccatgcc	32700
gggaagatta agcggaacgg gccccggcg cagggaggga caccgcgcc gcccgcgct	32760
gtccggatcc tgcaccgcgg ccccgcccg cggacgccc tggggtccgg ggcggcggac	32820
gggtcgtgcg agactggccg gtgtgagtgc caacacgagc ccgaagggcc agacgcctac	32880
cgcgaccccc gaccccgta agaacgacgc cgtccgggaa tcagccttcc tcaaggcgtg	32940
ccggcgcgag ccggtgccgc acacgccgt gtggttcattg cggcaggccg ggcgctcact	33000
gccggagtac cgcaagggtgc gcgaggcat cgggatgctc gactcctgca tgcggcccga	33060
gctggtcacc gagatcacc tccagccgtt gcgccgccac cacgtcgacg cggcgatcta	33120
cttcagcgac atcgtcgtcc cgctcaaggc catcggcatc gacctcgaca tcaagcccgg	33180
catcggcccc gtcgtcgagc agccggtgcg caccgcgcc gacctcgccc ggctgcgcga	33240
cctgaccccc gaggacgtct cctacgtcac cgaggccatc ggcatgctga cccgtgagct	33300
cgggtccacc ccgtgatcg gtttcgcggg cgccccgttc acccttgca gttacctgt	33360
cgaggcggc ccgtcccgt cgtacgagaa cgccaaggcg atgatgtacg gcgaccccga	33420
gctctgggcc gacctgctcg accgcctcg gcacatcacg gcggccttcc tcgacgtcca	33480
gatccgggcc ggcgcctcgg ccgtgcagct cttcgactcc tgggcccggc cgctcgcccc	33540
ctccgactac cggcggttcg tgctgcccgc ctcggcgaag gtgttcgcg cggtggccgg	33600
ccacggcgtc ccgcgcatcc acttcggcgt cggcaccggc gagctgctgg ggctcatggg	33660
cgaggccggc gcggacatcg tcggcgtcga ctggcgcgtc ccgatggacg aggccgccc	33720
ccgcgtcggc cccggcaagg cgctccagg caacctggac ccgaccgtgc tggtcgccgg	33780
ccgggaggcc gtcgagacga aggcgcgcga ggtcctggac accgccgcg gcctggagg	33840
ccacatcttc aacctcggtc acggagtgat gccctccacc gaccggacg ccctcaccg	33900
tctcgtggag tacgtccaca cgacagggc gcgctgacct accgctcacg cgccggacgc	33960

gagtcggaat cccggggcg ggtactgggc acgggtgccc accacgttca cccccgggta	34020
cgggcaggtg gagggcccat gaggctcgag atgttcgacc ccgccccgat cggcgctcgtg	34080
ttcaccaggg ggccggagca ccggctcgcg tacaccaacg ccgtctaccg ggagaccttc	34140
ggcgaccgcc cgctggggcg gacgatccgc gaggccttcc ccgacctcgc gcagtccggc	34200
tacttcgaca tcttcgaccg ggtcctcacc acgggcgcgg ccgaggtggt caccgcggtg	34260
cccctcgacc tgatctaccc cggtccacg ggcgagggca ggcgctactt cacgttcagc	34320
atctcccgcg ccacgatgag cgacggccgg ccgggagtgcc tcggcgtgat cgtggaggtg	34380
accgcgcagg tgaccgccgc ggaacggatc cgtgtgctgg ccgaggagcg ccgccgcgcg	34440
ctgcagcgct accgcagcct ggtgaacgcc ggaacgcaga tgggtgtgggt ggcgagcgcc	34500
aagggccgga tcaccgagcc gagccccggc tgggaacgcg tgaccgggca gacctgggag	34560
gagttccgcg gcgagggctg gatgaacgcc gtccaccccg acgaccgcgc cgctcggtc	34620
gaggcgtggc ggcgggcgac gaccgaacag gtgccgcgct ggatccacac ctaccggctg	34680
cggctggccg ccggcgggta ccggcacttc gtcgtcgacg ccgcgcccgt gcgcgacggg	34740
aacacggtga tcgaatgggt gggcacctgc acggacatcg agcgggaatg gcaggagggc	34800
cgccgtacgg aactgctggc gcgggcccgc accgccacgt ccggcatcgc gcggctggac	34860
gagatgctcg ccgccctggc cgatgtgatc gtgcccgcgc tcgccgacaa ctgcaccatc	34920
cacctctgc cgaggccct gcaccgtctg ccgggcaccc cgctgaccac cgaacgcgtc	34980
gccgcggtca cccgccggg gctcccggac ctgccccgc accacgagga gcacctgcgg	35040
cccggcagcc cgctggcccg cgccgccgac cgccgcagcc cgctccactt cgtcttcccg	35100
cccggcgagc cgccggccga cctcgctccg ctcgacggcg agccctggat ggccgaggac	35160
gtcaacagcg tcgtgctgct gcccgtcgtc gtcgacggca ccaccgccgc cctggtcgcc	35220
gtctccacca gcggcgcccg cccgcccctc ggccaggcgg agatcggcct gctgcagaca	35280
ctcctggaac gcgcccacac cccctcagc aacgccctgg agtaccagcg caccggcag	35340
gtggccctgg ccctgcagaa cagcctgctc accgaccgc cgacgcgcc cggcctggac	35400
atcgccgtcc gctaccggcc cagcaccgcc gccgccgagg tcggcgggga ctggtacgac	35460
gcgttcgtgc tgcgcgacgg cgccaccgtc ctcaccatcg gcgacgtctc cggccacgac	35520

ctgccggccg ccgtcacat gagccagctg cgcaacatgc tgcgcgggct cacgctggac	35580
cgccaggaac cgaccggcac catcctgcgc cggctggaca tcgccgtgca gaccctctat	35640
acggagtgca ccgccacctg cgtgctggcc cgggtggaac gcccggactc cggcggcgctc	35700
cggctgcact actccgtcgc cggtcacccg ccgccgctgc tcgtcgaggc ggacggctcc	35760
gcgcgcttcc tgaccggggc gcggtccccg atgctcgggc tcgtccccgc gccggagtac	35820
tcgagcgcca tggaaccgct gccgccggc tccaccctgc tgctgtacac cgacgggctg	35880
gtggagcgcc gcgacgagga tctcaccgtg ggcttgagc ggctgcggca ccacgcctcg	35940
gaggcggta gccgcccgt gcaggacttc tgcgacacac tgctcaccgg ccagctcacc	36000
gtcgacaacg acgacgacgt ggcgatgctg gtcctgcgcc ggtaggagcg tgccgaggag	36060
cgccactctg gccgatttta cccttgcttt tccatcggga ttggttctcc ggatttcccg	36120
atccggcgcc gacggcgaga ccgttgggat caccaatacc ccggaattcc cgcctccgcc	36180
accgttgggc agcgacggat cctgtgatat ttgactacg cgcggtgatg aattggctcg	36240
gtgccggtcg cggccggctg tagcagttct ggagcgctc tggacatcgt cacgagcgct	36300
tgtgattctt ggtcctgtac acgcaagccg gcgcaacgtc cacgttgccc atcagcggtt	36360
atcggcggtc caccggcgcg acggtgaccg cgggcgggta ctcatagggg gaactgcaat	36420
gaattactca aaagcagcga gaggaatgcc gacagccgga caaggtgccg ttcgggcggc	36480
gcgcgtcgtc cgtgaaagtc cggcggaatc agaaacggtc acagttcaga tagcgtcgtt	36540
attaccgggt gagtcgctgc gctcgaaagg gatcgagcag aaccacgtcg cggcactcgc	36600
ggaggtagac gcgccgcttc cgcccatact ggtggaccgg aagacgatgc gggtcgtcga	36660
cgggatgcac cggctcctcg cggctctgct caacggacgg cagacgatcg aggccgaact	36720
gttcgacgga accgcggatg agggattcct gcgcgccgtc cgggagaacg tggcgcacgg	36780
actcccgtg tcgcaggcgg accgccgggc cgccgctgcg cgcatcatcg tgtcccaccc	36840
gcatctgtcg gacagggcga tcgcccgggc gtccgggctc ggggcgaaga ccgtcgcggc	36900
cgtgcggcgc agttcaactg ccgtcgtgcc gcagttgaac acccgggtgg gccaggacgg	36960
cagggtccgg ccgtgaacg ggggcgaggg gcggcgagg gccatggcgg tactggccga	37020
acaccccgac gcgtccctgc gcgaggtcgc ccgtctgtcc ggggtgtcgc ccgcgacggt	37080
cagcgacgta cgccggcggc tggccgccgg cgagtcgccc ctgccgtcga gacgggaacc	37140

ggccgaaccg cggacgggcg ccgactccca ccgcaaccag agcttcgtgg atccccgtccc 37200
 ggtgctggag aagctgctgc gcgaccctc tctgcggcac aaggagggcg gccgccagct 37260
 gctccagctg ctccgccaga acgcggtcgg cgtgcaggac ctgatggagc tgtccgacgc 37320
 cgtgccgtcc cactgcaggt ccctgggtgat ccatctcgcg cagcagtacc gggacgcctg 37380
 gcagtccttc gcggagaagc tggacgagcc cgcctgcgcc tgtcccgggt gacgaacggg 37440
 cggcacggac ccgttcaccg gacatgaccg gcgccgcgcc gcgttcacgg cgcgccgccg 37500
 gcactcccac ggcacccgga ccaccgccgc gtatccggcg gacccggggc cgggcggggc 37560
 ggattcagcg ggcggggggc caggtgccac ccgattccag ccaccgggag agctccgccg 37620
 ccgagtcctt gcgcacgacc agttcgacga ggccgcgggt ctggtcctga ccgtgctcga 37680
 tgcggacgtc ctcatgttg acgccaagt cgcgatcga cgtgaacagt tcggccaggg 37740
 cgccgggctt gtcggagatg gtcaccgaga cggtcgcgag ctccgtccgg cgcgtaccgg 37800
 gtttgcgcac gatcctggcg ccccccggt tcccctcccg caacagctcc tcgagctcct 37860
 cctgcgcgcg gcggcgggacc agcgggtcgg cgtcggagac ggcgcgcagc gcgccgacgg 37920
 cccggcccag gccggcgggc agggagtcga gaacgtccgc cacggccgtg gcgttggaac 37980
 gcaggatgtc cccccagagc cgggcgtcac cggccgcgat ccgggtgacg tcggcgacgc 38040
 cctgccccgc cagccggacg ctgtcctccg ccgcgtgctc cagccgcgcg gcgagcaggg 38100
 aggagagccg atggggcgcg tgcgagacga gggccaccgc gtggtcgtgc acaccggcgt 38160
 ccatgaccac cggcatgccg tcgcacaacg acaccatctc cagggcgggtg ttcagcacgt 38220
 cctgcccgggt cagctccgac ggggtgagca cccagggggc cccctcgaag aggtccgccc 38280
 gggcggcgag cggcccggaa cgctcgggtc cggccagcgg atggcttcct atgtagctgg 38340
 ccgggtcggc ccgcatcgcg cgcacgtcgt cgtgcgggac cttcttgacg ctggcgacat 38400
 cgaggtaggc tcgggccagc ccgctctcct gtgcgcgcgc gagcacgcgt ccgacctgtg 38460
 ccgggggcac ggccagcacc gccaggtcga cctgacggtc cggctctctc agggatcccg 38520
 cgcccatcgc ctccgccgtc ctggcggcgt tccggtcgac gtcctccagg tgcacgccga 38580
 ccccgcgcg ggtcagcgcg agagcgacgg acgtgccgat ggccccgggtg ccgatgactg 38640
 tgggtggtcct caacgcgcgc ccccagggtg ggtgatccga aatcgggtcg gacaagtgcc 38700

gtgcccggca cgggaaaagg gaattcccat ggcgccgtgc gccgccaatt taacgcttcg	38760
gcgcgcatgt tcaactgcgg cgtcgcagcg gtcgaacaca gtagcggtag accggaccat	38820
tgaggcatcg tgctcagttg gcgacaccgg gtcggataaa cgccggaatc cgaggagttg	38880
acgttgacgt cagcgtgag acacgacgac ctgcatccga tagaagaagt ggaaataagt	38940
tcgctctcca ccgacggctc cccgcggatc gacggggaga gtcccagca cgtggaaatg	39000
ctggccgccc ccgacaccgc gcttccaccg atcatgggtgc accgccgcac cgggcgggtc	39060
atcgacggca tgcaccggct gcgcgccgcg atgtgacgg gccgtacgac gatcgcggtg	39120
aggttcttcg acggcaccga ggaggacgcc ttcttcctcg ccgtgaagtc gaacatcgcg	39180
cacggactgc cgctgtccgc cgccgaccgc cggcggggcg ccgggcgcat catggccacc	39240
catccccggt ggtcggaccg gatgatcgcc tcggtgggtcg gcacctccgc caggacggtc	39300
gccgagatcc gccgcgacgc cggcgccgcc ggggcggggg agcccacccg catcggccgg	39360
gacggcaggg tacggcccgt cgacgtgagc gagggccgca gactggccca cgacatgatc	39420
gtccgcgacc cgggcctgtc gctgcgccag gtcgcccgcg ccgccgggat ctgccggag	39480
accgtcaggg acgtcagaca ccggatgctc cgcggtgagg acccggtgcc cgcgccgcgg	39540
ccgcggaccc tgggtggagcg cggcgccggac cgccggggcg agccggccgg gaaggccgcc	39600
gcgccgtgcg ggacggagcc gccgcccgcc gtcgtgatga agcggctgag ggccgatccg	39660
gcgctgcgtc tcaacgagaa cggacgcgac ctgctgcggc ttctggatat ccacacggtc	39720
cggctggagg actggaaccg cattatcgaa agcgtgccgc cgcaccgtct ggagacggtg	39780
gcgcagctgg cacgctcctg cgccgacaaa tgggtccgaga tcgcgtcacg catcgaaagc	39840
aacgcatcac atctggccgg gtgaacgagg aaacacacga atccttcgag gagccgtcgg	39900
agaaagcggg acggcccgtc ggaacaccct tgtggagggg caatggagat acggtcgatc	39960
gatcacgtcg aattgttcgt cgaggacgcc caggacacgg ccggcaggct gtgcgactcc	40020
ttcggcttcg tccgcgtggg ccgcggcgcc gggaccaccg gactgcgcgg ctgcgagtcc	40080
gtcctgctgc gccagaacga catcgccctg ctgctgacca cggccaccga cgccgaccac	40140
cgtgccgccg agtacgtgaa gcagcacggg gacggggtcg cgggtgatcg gtgggatccc	40200
cgggtaccga gtcgaattc gccctatagt gagtctgatt acaattcact ggccgtcgtt	40260
ttacaacgtc gtgactggga aaaccctggc gttaccaac ttaatcgctt tgcagcacat	40320

ccccctttcg ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag	40380
ttgcgcagct gaatggcgaa tggcgcctga tgcggtatct tctccttacg catctgtgcg	40440
gtatttcaca ccgcatatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa	40500
gccagccccg acacccgcca acacccgctg acgcgaaccc cttgcggccg c	40551