



(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)



1-0050468

(51)^{2021.01} C12P 13/04; C12N 9/00; C12N 9/10

(13) B

(21) 1-2022-04534

(22) 14/12/2020

(86) PCT/EP2020/085882 14/12/2020

(87) WO2021/122400 24/06/2021

(30) 19218204.6 19/12/2019 EP

(45) 25/08/2025 449

(43) 26/09/2022 414A

(73) EVONIK OPERATIONS GMBH (DE)

Rellinghauser Strasse 1-11 45128 Essen, Germany

(72) SCHNEIDER, Frank (DE); JANKOWITSCH, Frank (DE).

(74) Công ty Luật TNHH Phạm và Liên danh (PHAM & ASSOCIATES)

(54) VI SINH VẬT CÓ HOẠT TÍNH ENZYM CÓ CHỨC NĂNG

CARBAMOYLPHOSPHAT SYNTHAZA VÀ PHƯƠNG PHÁP LÊN MEN TỔNG
HỢP AXIT GUANIDINO AXETIC

(21) 1-2022-04534

(57) Sáng chế đề cập đến vi sinh vật được biến nạp để có khả năng tổng hợp axit guanidinoaxetic (GAA) và phương pháp lên men tổng hợp GAA bằng cách sử dụng vi sinh vật này. Sáng chế còn đề cập đến phương pháp lên men tổng hợp creatin.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến vi sinh vật được biến nạp để có khả năng tổng hợp axit guanidinoaxetic (GAA) và phương pháp lên men tổng hợp GAA bằng cách sử dụng vi sinh vật này. Sáng chế còn đề cập đến phương pháp lên men tổng hợp creatin.

Tình trạng kỹ thuật của sáng chế

GAA là hợp chất hữu cơ được sử dụng làm phụ gia thức ăn chăn nuôi (WO2005120246 A1/ US2011257075 A1). GAA là tiền chất tự nhiên của creatin (ví dụ, Humm và đồng tác giả, Biochem. J. (1997) 322, 771-776). Do đó, việc bổ sung GAA cho phép cung cấp creatin một cách tối ưu trong cơ thể sinh vật.

Sáng chế đề cập đến phương pháp tổng hợp GAA bằng phương pháp lên men, sử dụng nguồn thức ăn công nghiệp (ví dụ, amoniac, muối amoni và cơ chất chứa glucoza hoặc đường) làm nguyên liệu đầu vào. Trong hệ thống sinh học, GAA và ornitin được tạo ra từ nguyên liệu đầu vào là arginin và glyxin nhờ hoạt tính xúc tác của L-arginin: glyxin- amidinotransferaza (AGAT; EC 2.1.4.1), là bước đầu tiên trong quá trình sinh tổng hợp creatin (US 20060200870 A1):



Guthmiller và đồng tác giả (J Biol Chem. 1994 Jul 1;269(26): 17556-60) đã xác định đặc điểm AGAT của thận chuột bằng cách tách dòng và biểu hiện khác loại enzym trong *E. coli*. Muenchhoff và đồng tác giả (Tạp chí FEBS 277 (2010) 3844–3860) báo cáo đặc điểm đầu tiên của AGAT từ sinh vật nhân sơ cũng bằng cách tách dòng và biểu hiện khác loại enzym trong *E. coli*. Sosio và đồng tác giả (Cell Chemical Biology 25, 540–549, 17 tháng 5 năm 2018) đã làm sáng tỏ

con đường sinh tổng hợp pseudouridimyxin ở *Streptomyces* sp.. Họ mô tả là phản ứng trung gian tạo ra GAA và L-ornitin bằng phản ứng giữa L-arginin với glyxin được xúc tác bởi PumN, L-arginin: glyxin- amidinotransferaza (AGAT). Humm và đồng tác giả biểu hiện gen tái tổ hợp mã hóa cho AGAT ở người trong *Escherichia coli* và nhận ra xystein-407 là gốc vị trí hoạt tính của AGAT (Biochem. J. (1997) 322, 771-776).

Mijts và đồng tác giả (WO 2018079687 A1) bộc lộ trong bản mô tả về sự tổng hợp các chất đích ở vi sinh vật, ví dụ, vanillin và axit vanilic, creatin mà có thể được tổng hợp từ L-arginin và glyxin. Các tác giả sáng chế đó còn đề xuất rằng điều này có thể đạt được bằng cách sử dụng các enzym sinh tổng hợp L-arginin, enzym sinh tổng hợp glyxin và enzym xúc tác chuyển L-arginin và glyxin thành creatin. L-arginin và glyxin có thể được hóa hợp với nhau để tạo ra guanidinoaxetat (GAA) và ornitin nhờ hoạt tính của AGAT (EC 2.1.4.1) và GAA có thể được metyl hóa để tạo ra creatin nhờ hoạt tính guanidinoaxetat N-metyltransferaza (GAMT, EC 2.1.1.2), sử dụng S-adenosyl metionin (SAM) làm chất cho metyl. Các tác giả sáng chế đó đề cập trong bối cảnh tổng hợp polyamin mà ví dụ về các enzym sinh tổng hợp L-arginin cũng có thể bao gồm các enzym sinh tổng hợp L-ornitin đã biết, cũng như các enzym đã biết từ chu trình L-ornitin, tức là carbamoyl phosphat synthaza (*carAB*), ornitin carbamoyl transferaza (*argF*, *argI*), argininosuccinat synthetaza (*argG*), argininosuccinat lyaza (*argH*) (xem Marc và đồng tác giả, Eur. J. Biochem. 267, 5217-5226, 200).

Tuy nhiên, một số cách tiếp cận cụ thể hơn nhằm tăng cường tổng hợp một trong các nguyên liệu đầu vào trong quá trình tổng hợp GAA, tức là L-arginin ở vi sinh vật, cụ thể là ở vi khuẩn, cũng đã được biết đến từ các tài liệu. Tổng quan về kỹ thuật trao đổi chất của *Corynebacterium glutamicum* (*C. glutamicum*) để tổng hợp L-arginin được cung cấp bởi Park và đồng tác giả (NATURE COMMUNICATIONS | DOI: 10.1038/ncomms5618). Họ đề xuất gây đột biến ngẫu nhiên và sàng lọc các vi khuẩn tổng hợp L-arginin thuộc các chủng *C. glutamicum* đã tổng hợp L-arginin, ví dụ, ATCC 21831 (Nakayama và Yoshida 1974, US3849250 A) và kỹ thuật chuyển hóa hợp lý từng bước dựa trên phân tích

toàn hệ thống về sự trao đổi chất dẫn đến việc dần gia tăng sự tổng hợp L-arginin trong suốt các bước kỹ thuật thiết kế. Yim và đồng tác giả (Yim và đồng tác giả (J Ind Microbiol Biotechnol (2011) 38: 1911–1920) đề xuất bất hoạt *argR*, gen mã hóa cho protein ức chế trung tâm ArgR, kiểm soát con đường sinh tổng hợp L-arginin bằng cách phá vỡ gen *argR* trên nhiễm sắc thể của *C. glutamicum*, dẫn đến cải thiện sản xuất arginin. Ginesy và đồng tác giả (Microbial Cell Factories (2015) 14: 29) báo cáo phương pháp thành công trên vi khuẩn *E. coli* để tăng cường tổng hợp arginin. Trong số các phương pháp đó, các tác giả đề xuất loại bỏ gen ức chế *argR*.

Kurahashi và đồng tác giả (EP1057893 A1) báo cáo các phương pháp làm gia tăng khả năng tổng hợp L-arginin của vi sinh vật nhờ tăng cường enzym sinh tổng hợp L-arginin bằng cách sử dụng các kỹ thuật ADN tái tổ hợp, ví dụ, bằng cách sử dụng vi sinh vật thuộc giống *Corynebacterium* hoặc *Brevibacterium*, được tổng hợp để chứa ADN tái tổ hợp bao gồm vectơ ADN và đoạn ADN chứa các gen mã hóa cho axetylornitin deaxetylaza, axit N-axetylglutamic- γ -semialdehyt dehydrogenaza, N-axetyl glutamokinaza có nguồn gốc từ argininosucxinaza của vi sinh vật thuộc giống *Escherichia*. Để cải thiện sự tổng hợp L-arginin, các tác giả sáng chế còn đề xuất vi sinh vật được tăng cường hoạt tính của glutamat dehydrogenaza nội bào (GDH) và có khả năng tổng hợp L-arginin.

Phương pháp sử dụng chủng tái tổ hợp di truyền, trong đó gen ức chế sự biểu hiện của operon sinh tổng hợp arginin, *argR*, bị bất hoạt đã được báo cáo bởi Suga và đồng tác giả (US7160705 B2). Cụ thể là, sự loại bỏ *argR*, kiểm soát operon arginin, được coi là yếu tố quan trọng trong tổng hợp arginin.

Ở vi sinh vật *Corynebacterium*, gen *argCJBDFR*, tham gia vào quá trình sinh tổng hợp arginin, được cấu tạo dưới dạng operon và bị arginin nội bào ức chế-phản hồi (Sakanyan và đồng tác giả, Microbiology, 142: 9-108, 1996), do đó hạn chế giới hạn tổng hợp L-arginin năng suất cao của nó.

Tuy nhiên, Bae và đồng tác giả (EP3153573 A1) đã nỗ lực tăng năng suất tổng hợp L-arginin ở *C. glutamicum*, đã phát hiện ra rằng L-arginin có thể được tổng hợp với năng suất cao hơn so với chủng tổng hợp L-arginin bố mẹ, bằng cách

tăng cường hoạt tính của arginin operon và ornitin carbamoyltransferaza (ArgF, ArgF2) mà không cần loại bỏ gen ức chế arginin (*argR*).

Operon arginin là operon bao gồm các gen mã hóa cho các enzym tham gia vào cơ chế sinh tổng hợp L-arginin, và cụ thể là, operon arginin bao gồm các gen mã hóa cho các enzym tạo nên các bước của chu trình sinh tổng hợp L-arginin. Cụ thể, operon arginin bao gồm N-axetylglutamyl phosphat reductaza (ArgC), glutamat N-axetyltransferaza (ArgJ), N-axetylglutamat kinaza (ArgB), axetylornitin aminotransferaza (ArgD), ornitin carbamoyltransferaza (ArgF), và gen ức chế arginin (ArgR). Các enzym này tham gia vào các phản ứng enzym liên tục của quá trình sinh tổng hợp L-arginin và được điều khiển bởi chất ức chế arginin, được mã hóa bởi *argR* (WO 2006/057450 A1).

Fan Wenchao bộc lộ phương pháp tổng hợp creatin bằng cách lên men vi sinh vật không gây bệnh, như *Corynebacterium glutamicum* (CN106065411 A). Vi sinh vật này có các chức năng chuyển hóa sinh học như sau: chuyển hóa glucoza thành axit L- glutamic; chuyển hóa axit L-glutamic thành axit N-axetyl-L-glutamic; chuyển hóa axit N- axetyl-L-glutamic thành semialdehyt axit N-axetyl-L-glutamic; chuyển hóa semialdehyt axit N-axetyl-L-glutamic thành N-axetyl-L-ornitin; chuyển hóa N-axetyl-L-ornitin thành L-ornitin; chuyển hóa L-ornitin thành L-xitrulin; chuyển hóa L-xitrulin thành axit arginino-succinic; chuyển hóa axit arginino-succinic thành L-arginin; chuyển hóa L-arginin thành axit guanidinoaxetic; và cuối cùng là chuyển hóa axit guanidinoaxetic thành creatin. Fan Wenchao đề xuất rằng vi sinh vật biểu hiện quá mức hoặc nhiều enzym được chọn từ nhóm bao gồm N-axetylglutamat-synthaza, N-axetylornitin- δ -aminotransferaza, N-axetylornitinaza, ornitin-carbamoyl transferaza, argininosuccinat synthetaza, glyferaza amidino-transferaza (EC: 2.1, 4.1), và guanidinoaxetat N- metyltransferaza (EC: 2.1.1.2). Tốt hơn nếu vi sinh vật này biểu hiện quá mức glyxin aminotransferaza (L-arginin:glyxin amidinotransferaza) và guanidinoaxetat N- metyltransferaza.

Cho đến nay, các vi sinh vật thích hợp để tăng cường tổng hợp GAA so với các kiểu đại của chúng và phương pháp tương ứng để tổng hợp GAA bằng

cách sử dụng các vi sinh vật đó vẫn chưa được báo cáo.

Bản chất kỹ thuật của sáng chế

Do đó, sáng chế nhằm cung cấp vi sinh vật được biến nạp để có khả năng tổng hợp axit guanidinoaxetic (GAA) và phương pháp lên men tổng hợp GAA bằng cách sử dụng vi sinh vật này.

Vấn đề này được giải quyết bởi vi sinh vật có hoạt tính gia tăng bằng cách biểu hiện quá mức enzym có chức năng carbamoylphosphat synthaza (EC 6.3.4.16) so với hoạt tính enzym tương ứng ở vi sinh vật kiểu dại và mang ít nhất một gen dị hợp mã hóa cho protein có chức năng L-arginin:glyxin amidinotransferaza (AGAT).

Gen dị hợp có nghĩa là gen đã được đưa vào sinh vật chủ không có gen này một cách tự nhiên. Việc đưa gen dị hợp vào vật chủ được thực hiện bằng kỹ thuật ADN tái tổ hợp. Các vi sinh vật đã trải qua kỹ thuật ADN tái tổ hợp được gọi là vi sinh vật chuyển gen, biến đổi gen hoặc tái tổ hợp. Do đó, vi sinh vật theo sáng chế là vi sinh vật tái tổ hợp.

Sự gia tăng hoạt tính của enzym có chức năng enzym của carbamoylphosphat synthaza có thể đạt được bằng cách gây đột biến và/hoặc biểu hiện quá mức của gen mã hóa cho enzym có chức năng carbamoylphosphat synthaza.

Hoạt tính của L-arginin:glyxin amidinotransferaza cũng có thể tăng lên bằng cách gây đột biến và/hoặc biểu hiện quá mức của gen mã hóa cho L-arginin:glyxin amidinotransferaza.

Protein có chức năng L-arginin:glyxin amidinotransferaza (AGAT) thuộc họ amidinotransferaza. Họ amidinotransferaza bao gồm glyxin (EC: 2.1.4.1) và inosamin (EC: 2.1.4.2) amidinotransferaza, các enzym lần lượt tham gia tham gia vào quá trình sinh tổng hợp creatin và streptomycin. Họ này còn bao gồm các deiminaza arginin, EC: 3.5.3.6. Các enzym này xúc tác cho phản ứng: arginin + H₂O $\rightleftharpoons$ xitrulin + NH₃. Glycoprotein chống khối u Streptococcus cũng được tìm thấy trong họ này. Các enzym hoặc protein có hoạt tính L-arginin: glyxin-

amidinotransferaza (AGAT) cũng được mô tả là có vùng bảo thủ thuộc họ PFAM: Amidinotransf (PF02274(Marchler-Bauer A và đồng tác giả (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", Nucleic Acids Res. 45(D1): D200-D203.) như được mô tả trong các đơn sáng chế sau: Pissowotzki K và đồng tác giả, Mol Gen Genet 1991;231: 113-123 (PUBMED: 1661369 EPMC: 1661369); D'Hooghe I và đồng tác giả, J Bacteriol 1997;179: 7403-7409 (PUBMED: 9393705 EPMC: 9393705); Kanaoka M và đồng tác giả, Jpn J Cancer Res 1987;78: 1409-1414 (PUBMED: 3123442 EPMC: 3123442). Các ví dụ cụ thể về AGAT là AGAT của các sinh vật *Moorea producens*, *Homo sapiens*, *Rattus norvegicus*, *Galeopterus variegatus*, và *Cylindrospermopsis raciborskii*.

Lý tưởng nhất là vi sinh vật theo sáng chế có khả năng tổng hợp L-arginin cải thiện hơn so với khả năng tổng hợp L-arginin của vi sinh vật kiểu đại. Đặc tính này có thể đạt được bằng cách chọn lọc các vi sinh vật là các vi sinh vật tổng hợp L-arginin tự nhiên hoặc có thể có khả năng tổng hợp L-arginin bằng cách gây đột biến.

Vi sinh vật theo sáng chế và có khả năng tổng hợp L-arginin cải thiện hơn so với khả năng tổng hợp L-arginin của vi sinh vật kiểu đại có thể có hoạt tính gia tăng của enzym có chức năng argininosuccinat lyaza (EC 4.3.2.1) so với hoạt tính của enzym tương ứng ở vi sinh vật kiểu đại.

Ngoài ra, ở vi sinh vật theo sáng chế, hoạt tính của enzym có chức năng ornitin carbamoyltransferaza (EC 2.1.3.3) có thể tăng lên so với hoạt tính enzym tương ứng ở vi sinh vật kiểu đại.

Ở vi sinh vật theo sáng chế, hoạt tính của enzym có chức năng của enzym argininosuccinat synthaza (E.C. 6.3.4.5) cũng có thể tăng lên so với hoạt tính enzym tương ứng ở vi sinh vật kiểu đại.

Có thể đạt được sự gia tăng hoạt tính của enzym ở vi sinh vật, ví dụ, bằng cách gây đột biến gen nội sinh tương ứng. Phương pháp khác để tăng hoạt tính của enzym có thể ổn định mã hóa mRNA cho các enzym này.

Hoạt tính gia tăng của các enzym nêu trên cũng có thể đạt được bằng cách biểu hiện quá mức các gen mã hóa cho các enzym tương ứng. Nói cách khác, tốt hơn là vấn đề này được giải quyết bởi vi sinh vật có khả năng tổng hợp L-arginin cải thiện hơn so với khả năng tổng hợp L-arginin của vi sinh vật kiểu đại và/hoặc vi sinh vật có ít nhất một hoặc nhiều gen biểu hiện quá mức (ví dụ, *carA*, *carB*) mã hóa cho protein có chức năng tổng hợp carbamoylphosphat synthaza (EC 6.3.4.16) và còn mang gen mã hóa cho protein có chức năng L-arginin:glyxin amidinotransferaza (AGAT, ví dụ, EC 2.1.4.1). Tốt hơn nếu vi sinh vật theo sáng chế còn mang ít nhất một hoặc nhiều gen biểu hiện quá mức được chọn từ nhóm bao gồm gen (ví dụ, *argF/argF2/argI*) mã hóa cho protein có chức năng ornitin carbamoyltransferaza (EC 2.1.3.3), gen (ví dụ, *argG*) mã hóa cho protein có chức năng argininosuccinat synthetaza (EC 6.3.4.5) và gen (ví dụ, *argH*) mã hóa cho protein có chức năng argininosuccinat lyaza (EC 4.3.2.1).

Sự biểu hiện quá mức của gen thường đạt được bằng cách tăng số lượng bản sao của gen và/hoặc bằng cách liên kết gen về mặt chức năng bằng vùng khởi động và/hoặc bằng cách tăng cường vị trí liên kết của riboxom và/hoặc bằng cách tối ưu hóa việc sử dụng mã bộ ba của mã bộ ba khởi đầu hoặc của hệ gen hoặc tổ hợp bao gồm việc lựa chọn tất cả các phương pháp được đề cập trên đây.

Trong bản mô tả sáng chế, vi sinh vật có khả năng tổng hợp L-arginin được cải thiện có nghĩa là vi sinh vật tổng hợp L-arginin vượt quá nhu cầu của chính nó. Ví dụ về các vi sinh vật tổng hợp L-arginin này là, ví dụ, *C. glutamicum* ATCC 21831 hoặc các vi sinh vật được công bố bởi Park và đồng tác giả (NATURE COMMUNICATIONS | DOI: 10.1038/ncomms5618) hoặc bởi Ginesy và đồng tác giả (Ginesy và đồng tác giả (Microbial Cell Factories (2015) 14:29).

Theo phương án vi sinh vật theo sáng chế, operon arginin (*argCJBDFR*) có thể được biểu hiện quá mức.

Ngoài ra, ở vi sinh vật theo sáng chế, gen *argR* mã hóa cho protein ức chế đáp ứng arginin ArgR có thể bị giảm độc lực hoặc bị loại bỏ.

Theo phương án khác của sáng chế và, tùy ý ngoài các biến đổi nêu trên, ít nhất một hoặc nhiều gen mã hóa cho các enzym tham gia vào con đường sinh tổng

hợp L-arginin, bao gồm *gdh*, *argJ*, *argB*, *argC* và/hoặc *argD* lần lượt mã hóa cho glutamat dehydrogenaza, ornitin axetyltransferaza, axetylglutamat kinaza, axetylglutamylphosphat reductaza và axetylornitin aminotransferaza, được biểu hiện quá mức ở vi sinh vật theo sáng chế.

Bảng 1 thể hiện tên của các enzym khác nhau tham gia hoặc góp phần vào quá trình sinh tổng hợp arginin ở các loài khác nhau, tức là *E. coli*, *C. glutamicum* và *Pseudomonas putida* (*P. putida*).

Ở vi sinh vật theo sáng chế, gen mã hóa cho protein có chức năng L-arginin:glyxin amidinotransferaza có thể được biểu hiện quá mức. Sự biểu hiện quá mức của gen thường đạt được bằng cách gia tăng số lượng bản sao của gen và/hoặc bằng cách liên kết chức năng giữa gen này với vùng khởi động và/hoặc bằng cách tăng cường vị trí liên kết của riboxom và/hoặc bằng cách tối ưu hóa việc sử dụng mã bộ ba của mã bộ ba khởi đầu của hệ gen hoặc kết hợp bao gồm bước chọn lọc hoặc tất cả các phương pháp nêu trên.

Bảng 1: Tên enzym

Tên enzym	Tên riêng	Số EC	<i>E. coli</i>	<i>C. glutamicum</i>	<i>P. putida</i>
Glutamat dehydrogenaza	GLUTAMAT DEHYDROGENAZA ĐẶC HIỆU NADP	1.4.1.4 (1.4.1.2 ở <i>P. putida</i>)	<i>gdhA</i>	Gdh	<i>gdhA/g dhB</i>
N-axetyl glutamokinaza	AXETYLGLUTAMAT KINAZA	2.7.2.8	<i>argB</i>	<i>argB</i>	<i>argB</i>
N-axetylglutamyl phosphat reductaza	N-axetyl-gamma-glutamylphosphat reductaza	1.2.1.38	<i>argC</i>	<i>argC</i>	<i>argC</i>
N-axetylglutamic axit-γ-semialdehyt dehydrogenaza	AXETYLORNITIN AMINOTRANSFERAZA (axetylornitin aminotransferaza hai chức trong e.c./succinyldiaminopimelat aminotransferaza)	2.6.1.11/ 2.6.1.17	<i>argD</i>	<i>argD</i>	<i>aruC</i>
Axetylornitin deaxetylaza	axetylornitin deaxetylaza hai chức /GLUTAMAT N-AXETYLTRANSFERAZA	2.3.1.35 / 2.3.1.1		<i>argJ</i>	<i>argJ</i>
Axetylornitin deaxetylaza		3.5.1.16	<i>argE</i>		<i>argE</i>
	GLUTAMAT N-AXETYLTRANSFERAZA	2.3.1.1	<i>argA</i>		<i>argA</i>

Ornitin carbamoyltransferaza	ornitin carbamoyltransferaza 1	2.1.3.3	argI	argF/argF2	arcB/arg F
Argininosuccinat synthetaza	ARGININOSUCXINAT SYNTAZA	6.3.4.5	argG	argG	argG
Argininosuccinat lyaza	ARGININOSUCXINAT LYAZA	4.3.2.1	argH	argH	argH
Carbamoyl-phosphat synthaza	carbamoyl-phosphat synthaza	6.3.5.5	carAB	carAB	carAB
Carbamat kinaza	carbamat kinaza	2.7.2.2	ybcF /yqeA		arcC

Protein có chức năng L-arginin:glyxin amidinotransferaza (AGAT) trong vi sinh vật theo sáng chế có thể bao gồm trình tự axit amin tương đồng ít nhất là 70%, tốt hơn nếu ít nhất là 80% hoặc ít nhất là 90% so với trình tự axit amin theo SEQ ID NO:2 hoặc SEQ ID NO:4 (“AGAT_Mp” của *Moorea producens*). Theo phương án khác của sáng chế, trình tự axit amin của L-arginin:glyxin amidinotransferaza tương đồng với trình tự axit amin theo SEQ ID NO:2 hoặc SEQ ID NO:4 (xem Cơ sở dữ liệu UniProt, ngày 15 tháng 2 năm 2017, “Glycine amidinotransferase”, XP055706853, số đăng ký EBI UNIPROT: A0A1D8TKD3). Trình tự ADN kiểu đại mã hóa cho *Moorea producens* AGAT là SEQ ID NO:1, trình tự ADN tương ứng đã được tối ưu hóa mã bộ ba cho *C. glutamicum* là SEQ ID NO:3. Trình tự ADN tương ứng của gen AGAT ở *Moorea producens* đã được tối ưu hóa mã bộ ba cho *P. putida* là SEQ ID NO:33.

Protein có chức năng L-arginin:glyxin amidinotransferaza ở vi sinh vật theo sáng chế có thể bao gồm trình tự axit amin tương đồng ít nhất là 70%, tốt hơn nếu ít nhất là 80% hoặc ít nhất là 90% so với trình tự axit amin theo SEQ ID NO:16 hoặc SEQ ID NO:26 (“AGAT_cyrA” của *Cylindrospermopsis raciborskii* ATW205). Theo phương án khác của sáng chế, trình tự axit amin của L-arginin:glyxin amidinotransferaza tương đồng với trình tự axit amin theo SEQ ID NO:16 hoặc SEQ ID NO:26. Trình tự ADN kiểu đại mã hóa cho *Cylindrospermopsis raciborskii* AGAT là SEQ ID NO:15, trình tự ADN tương ứng đã được tối ưu hóa mã bộ ba cho *C. glutamicum* là SEQ ID NO:25.

Protein có chức năng L-arginin:glyxin amidinotransferaza ở vi sinh vật theo

sáng chế có thể bao gồm trình tự axit amin tương đồng ít nhất là 70%, tốt hơn nếu ít nhất là 80% hoặc ít nhất là 90% so với trình tự axit amin theo SEQ ID NO:23 (“AGAT_Gv” của *Galeopterus variegatus*), tốt hơn là SEQ ID NO:24 hoặc SEQ ID NO:32. Theo phương án khác của sáng chế, trình tự axit amin của L-arginin:glyxin amidinotransferaza tương đồng với trình tự axit amin theo SEQ ID NO:24 hoặc SEQ ID NO:32 Trình tự của ADN AGAT của *Galeopterus variegatus* tương ứng đã được tối ưu hóa mã bộ ba cho *C. glutamicum* là SEQ ID NO:31.

Protein có chức năng L-arginin:glyxin amidinotransferaza ở vi sinh vật theo sáng chế có thể bao gồm trình tự axit amin tương đồng ít nhất là 70%, tốt hơn nếu ít nhất là 80% hoặc ít nhất là 90% so với trình tự axit amin theo SEQ ID NO:18 (“AGAT_Hs” của *homo sapiens*), tốt hơn là SEQ ID NO:20 hoặc SEQ ID NO:28, ví dụ, trình tự axit amin theo SEQ ID NO:21 hoặc SEQ ID NO:22 hoặc SEQ ID NO:30 (mỗi “AGAT_Rn” của *Rattus norvegicus*; ADN, mã bộ ba tương ứng được tối ưu hóa cho *C. glutamicum* là SEQ ID NO:29). Theo phương án khác của sáng chế, trình tự axit amin của L-arginin:glyxin amidinotransferaza tương đồng với trình tự axit amin theo SEQ ID NO:18. Trình tự của ADN kiểu đại mã hóa cho *Homo sapiens* AGAT là SEQ ID NO:17, trình tự ADN tương ứng đã được tối ưu hóa mã bộ ba cho *C. glutamicum* là SEQ ID NO:27.

Vi sinh vật theo sáng chế có thể thuộc giống *Corynebacterium*, tốt hơn là *Corynebacterium glutamicum* (*C. glutamicum*), hoặc giống *Enterobacteriaceae*, tốt hơn là *Escherichia coli* (*E. coli*), hoặc giống *Pseudomonas*, tốt hơn là *Pseudomonas putida* (*P. putida*).

Nói chung, có thể đạt được sự gia tăng hoạt tính của enzym ở vi sinh vật, ví dụ, bằng cách gây đột biến gen nội sinh tương ứng. Hoạt tính của enzym cũng có thể được tăng cường bằng cách biểu hiện quá mức của gen tương ứng.

Theo sáng chế, nói chung, sự biểu hiện quá mức của gen, đạt được bằng cách tăng số lượng bản sao của gen và/hoặc bằng cách tăng cường các yếu tố điều hòa, ví dụ, bằng cách liên kết chức năng gen với vùng khởi động và/hoặc bằng cách tăng cường vị trí liên kết của riboxom và/hoặc bằng cách tối ưu hóa việc sử dụng mã bộ ba của mã bộ ba khởi đầu hoặc của hệ gen. Ví dụ, việc tăng cường các

yếu tố điều hòa ảnh hưởng tích cực đến sự biểu hiện gen có thể đạt được bằng cách biến đổi trình tự vùng khởi động ngược dòng của gen cấu trúc để tăng hiệu quả của trình tự khởi động hoặc bằng cách thay thế hoàn toàn trình tự khởi động này bằng trình tự có hiệu quả hơn - được gọi là vùng khởi động mạnh. Vùng khởi động nằm ở phía ngược dòng của gen. Vùng khởi động là trình tự ADN bao gồm khoảng 40 đến 50 cặp bazơ và tạo thành vị trí liên kết cho holoenzym ARN polymeraza và điểm khởi đầu phiên mã, do đó có thể ảnh hưởng đến cường độ biểu hiện của polynucleotit hoặc gen được kiểm soát. Nói chung, có thể đạt được sự biểu hiện quá mức hoặc gia tăng sự biểu hiện của các gen ở vi khuẩn bằng cách chọn các vùng khởi động mạnh, ví dụ, bằng cách thay thế trình khởi động ban đầu bằng trình khởi động tự nhiên, mạnh (ban đầu được gán cho các gen khác) hoặc bằng cách biến đổi các vùng nhất định của vùng khởi động gốc tự nhiên đã nêu (ví dụ, cái gọi là vùng -10 và -35 của nó) hướng tới trình tự bảo thủ, ví dụ, như được đề xuất bởi M. Patek và đồng tác giả (Microbial Biotechnology 6 (2013), 103-117) cho *C. glutamicum*. "Sự liên kết chức năng" được hiểu là sự sắp xếp trình tự của vùng khởi động với gen, dẫn đến phiên mã gen.

Mã di truyền bị thoái hóa có nghĩa là axit amin nhất định có thể được mã hóa bởi một số bộ ba khác nhau. Thuật ngữ sử dụng mã bộ ba được dùng để quan sát mà một số sinh vật thường sẽ không sử dụng tất cả mã bộ ba có thể có cho một axit amin nhất định với cùng tần số. Thay vào đó, sinh vật thường sẽ thể hiện một số ưu tiên nhất định đối với các mã bộ ba cụ thể, có nghĩa là các mã bộ ba này được tìm thấy thường xuyên hơn trong trình tự mã hóa cho các gen đã phiên mã của sinh vật. Nếu gen nào đó ngoại lai với vật chủ tương lai của nó, tức là từ loài khác thì sẽ phải được biểu hiện trong sinh vật chủ tương lai, sau đó, trình tự mã hóa cho của gen nêu trên sẽ được điều chỉnh để sử dụng mã bộ ba của sinh vật chủ tương lai đó (tức là tối ưu hóa việc sử dụng mã bộ ba).

Mô tả chi tiết các phương án thực hiện sáng chế

Vấn đề nêu trên tiếp tục được giải quyết bằng phương pháp lên men tổng hợp axit guanidino axetic (GAA), bao gồm các bước a) nuôi cấy vi sinh vật theo sáng chế như đã xác định trên đây trong môi trường thích hợp ở các điều kiện thích hợp,

và b) tích lũy GAA trong môi trường để tạo ra môi trường lên men chứa GAA.

Phương pháp theo sáng chế còn có thể bao gồm bước bổ sung glyxin và/hoặc bổ sung L-arginin và/hoặc bổ sung L-ornitin vào môi trường. Tốt hơn là môi trường này được bổ sung glyxin với hàm lượng nằm trong khoảng từ 0,1 đến 300g glyxin/l môi trường, tốt hơn là môi trường này chứa glyxin với lượng 0,82g/l, và/hoặc chứa L-arginin sao cho hàm lượng nằm trong khoảng từ 0,1 đến 200g/l môi trường, tốt hơn nếu chứa L-arginin với lượng 1,9g/l môi trường.

Phương pháp theo sáng chế có thể bao gồm thêm bước tách GAA ra khỏi môi trường lên men.

Phương pháp theo sáng chế có thể bao gồm thêm bước làm khô và/hoặc tạo hạt môi trường lên men chứa GAA.

Sáng chế còn đề cập đến vi sinh vật như được định nghĩa trên đây, còn mang gen mã hóa cho enzym có hoạt tính guanidinoaxetat N-metyltransferaza (EC: 2.1.1.2). Tốt hơn nếu gen mã hóa cho enzym có hoạt tính guanidinoaxetat N-metyltransferaza được biểu hiện quá mức.

Sáng chế cũng đề cập đến phương pháp lên men tổng hợp creatin, bao gồm các bước a) nuôi cấy vi sinh vật theo sáng chế mang gen mã hóa cho enzym có hoạt tính guanidinoaxetat N-metyltransferaza trong môi trường thích hợp ở điều kiện thích hợp và tích lũy creatin trong môi trường để tạo ra dịch lên men chứa creatin.

Tốt hơn nếu phương pháp này còn bao gồm bước tách creatin từ môi trường lên men creatin. Creatin có thể được chiết xuất từ môi trường lên men bằng phương pháp điểm đẳng điện và/hoặc phương pháp trao đổi ion. Ngoài ra, creatin có thể được tinh chế thêm bằng phương pháp kết tinh lại trong nước.

Ví dụ thực hiện sáng chế

Phần thử nghiệm

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Hóa chất

Dung dịch kanamycin từ *Streptomyces kanamyceticus* được mua từ Sigma Aldrich (St. Louis, Mỹ, mã số K0254). IPTG (Isopropyl β -D-1-thiogalactopyranoside) được mua từ Carl-Roth (Karlsruhe, Đức, mã số 2316.4.). Nếu không có quy định khác, tất cả các hóa chất khác được mua từ Merck (Darmstadt, Đức), Sigma Aldrich (St. Louis, Mỹ) hoặc Carl-Roth (Karlsruhe, Đức) là tinh khiết về mặt phân tích.

Nuôi cấy tăng sinh tế bào

Nếu không có quy định khác, phương pháp nuôi cấy/ủ được thực hiện như sau:

a. môi trường LB (MILLER) từ Merck (Darmstadt, Đức; mã số 110285) được sử dụng để nuôi cấy các chủng *E. coli* trong môi trường lỏng. Các mẫu cấy lỏng (10ml môi trường lỏng trên mỗi bình Erlenmeyer 100ml có 3 vách ngăn) được ủ trong tủ ẩm lắc tiêu chuẩn Infors HT Multitron của Infors GmbH (Bottmingen, Thụy Sĩ) ở 30°C và 200 vòng/phút.

b. Thạch LB (MILLER) từ Merck (Darmstadt, Đức, mã số 110283) được sử dụng để nuôi cấy các chủng *E. coli* trên đĩa thạch. Các đĩa thạch được ủ ở 30°C trong tủ ẩm mini INCU-Line® của VWR (Radnor, Mỹ).

c. Môi trường nuôi cấy tim não (BHI) từ Merck (Darmstadt, Đức, mã số 110493) được sử dụng để nuôi cấy các chủng *C. glutamicum* trong môi trường lỏng. Các mẫu cấy lỏng (10ml môi trường lỏng trên bình Erlenmeyer 100ml có 3 vách ngăn) được ủ trong tủ ẩm lắc tiêu chuẩn Infors HT Multitron của Infors GmbH (Bottmingen, Thụy Sĩ) ở 30°C và 200 vòng/phút.

d. Thạch tim não (BHI-agar) từ Merck (Darmstadt, Đức, mã số 113825) được sử dụng để nuôi cấy các chủng *C. glutamicum* trên đĩa thạch. Các đĩa thạch này được ủ ở 30°C trong tủ ẩm của Heraeus Instruments bằng bộ điều khiển nhiệt độ Kelvitron® (Hanau, Đức).

e. Để nuôi cấy *C. glutamicum* sau khi điện di, thạch BHI (Merck, Darmstadt, Germany, mã số 113825) được bổ sung sorbitol (Carl Roth GmbH + Co. KG, Karlsruhe, Germany) với lượng 134g/l, chiết xuất nấm men (Oxoid/ThermoFisher

Scientific, Waltham, Mỹ, mã số LP0021) với lượng 2,5g/l và kanamycin với lượng 25mg/l. Các đĩa thạch được ủ ở 30°C trong tủ ẩm của Heraeus Instruments có bộ điều khiển nhiệt độ Kelvitron® (Hanau, Đức).

Xác định mật độ quang học của huyền phù chứa vi khuẩn

a. Mật độ quang học của huyền phù chứa vi khuẩn trong môi trường nuôi cấy bình lắc được xác định ở bước sóng 600 nm (OD₆₀₀) bằng máy quang phổ của Eppendorf AG (Hamburg, Đức).

b. Mật độ quang học của huyền phù chứa vi khuẩn được tạo ra trong hệ thống lên men vi mô Wouter Duetz (WDS) (tấm 24 giếng) được xác định ở bước sóng 660 nm (OD₆₆₀) bằng đầu đọc đĩa GENios™ của Tecan Group AG (Männedorf, Thụy Sĩ).

Ly tâm

a. Huyền phù chứa vi khuẩn có thể tích tối đa là 2ml được ly tâm trong các ống phản ứng 1,5ml hoặc 2ml (ví dụ, ống Eppendorf® 3810X) bằng máy ly tâm để bàn Eppendorf 5417 R (5 phút ở tốc độ 13.000 vòng/phút).

b. Huyền phù chứa vi khuẩn có thể tích tối đa là 50ml được ly tâm trong ống ly tâm 15ml hoặc 50ml (ví dụ, ống ly tâm hình nón Falcon™ 50ml) bằng máy ly tâm để bàn Eppendorf 5810 R trong 10 phút với tốc độ 4.000 vòng/phút.

Phân lập ADN

ADN plasmit được phân lập từ tế bào *E. coli* bằng Kit QIAprep Spin Miniprep của Qiagen (Hilden, Đức, mã số 27106) theo hướng dẫn của nhà sản xuất.

Phản ứng chuỗi polymeraza (PCR)

PCR bằng polymeraza đọc sửa (độ chính xác cao) được sử dụng để khuếch đại đoạn ADN mong muốn nhằm xác định trình tự Sanger hoặc nối ADN. Kit polymeraza không đọc sửa được sử dụng để trực tiếp xác định sự có mặt hoặc không có mặt của đoạn ADN mong muốn từ các khuẩn lạc *E. coli* hoặc *C. glutamicum*.

a. Kit Phusion® ADN Polymeraza có độ trung thực cao (Kit PhusionKit) từ New England BioLabs Inc. (Ipswich, Mỹ, mã số M0530) được sử dụng để khuếch đại đúng khuôn mẫu các vùng ADN đã chọn theo hướng dẫn của nhà sản xuất (xem bảng 2).

Bảng 2: Các điều kiện chu trình nhiệt cho PCR với Kit Phusion® ADN Polymeraza có độ trung thực cao từ New England BioLabs Inc.

Chương trình PCR			
Bước	Thời gian [phút : giây]	Nhiệt độ [°C]	Mô tả
1	05: 00	94	Bước biến tính ban đầu
2	00: 30	94	Bước biến tính
3	00: 30	52	Bước gắn môi
4	01: 20	72	Bước kéo dài 1 phút/kb ADN
			Lặp lại bước 2 đến 4: 35 x
5	04: 00	72	Bước kéo dài cuối cùng
6	Giữ	4	Bước làm mát

b. Kit Taq PCR Core (Taq Kit) của Qiagen (Hilden, Đức, mã số 201203) được sử dụng để khuếch đại đoạn ADN mong muốn nhằm xác nhận sự có mặt của nó. Kit này được sử dụng theo hướng dẫn của nhà sản xuất (xem bảng 3).

Bảng 3: Các điều kiện chu trình nhiệt cho PCR bằng Kit Taq PCR Core (Taq Kit) của Qiagen.

Chương trình PCR			
Bước	Thời gian [phút : giây]	Nhiệt độ [°C]	Mô tả
1	05: 00	94	Bước biến tính ban đầu
2	00: 30	94	Bước biến tính
3	00: 30	52	Bước gắn mồi
4	01: 20	72	Bước kéo dài 1 phút/kb ADN
			Lặp lại bước 2 đến 4: 35 x
5	04: 00	72	Bước kéo dài cuối cùng
6	Giữ	4	Bước làm mát

c. SapphireAmp® Fast PCR Master Mix (Sapphire Mix) từ Takara Bio Inc (Takara Bio Europe SAS, Saint-Germain-en-Laye, Pháp, mã số RR350A/B) được sử dụng thay thế nhằm xác nhận sự có mặt của đoạn ADN mong muốn trong tế bào được lấy từ khuẩn lạc *E. coli* hoặc *C. glutamicum* theo hướng dẫn của nhà sản xuất (xem bảng 4).

Bảng 4: Các điều kiện chu trình nhiệt cho PCR với SapphireAmp® Fast PCR Master Mix (Sapphire Mix) từ Takara Bio Inc.

Chương trình PCR			
Bước	Thời gian [phút : giây]	Nhiệt độ [°C]	Mô tả
1	01: 00	94	Bước biến tính ban đầu
2	00: 05	98	Bước biến tính
3	00: 05	55	Bước gắn mồi
4	00: 05	72	Bước kéo dài
			Lặp lại bước 2 đến 4: 30 x
5	04: 00	72	Bước kéo dài cuối cùng
6	Giữ	4	Bước làm mát

d. Tất cả các mồi oligonucleotit được Eurofins Genomics GmbH (Ebersberg, Đức) tổng hợp bằng phương pháp phosphoramidit, phương pháp này đã được mô tả bởi McBride và Caruthers (1983).

e. Dung dịch ADN plasmit hoặc ADN tổng số được phân lập từ môi trường nuôi cấy lỏng hoặc ADN tổng số có trong khuẩn lạc (PCR khuẩn lạc) với độ pha loãng thích hợp đã được sử dụng làm khuôn PCR. Đối với PCR khuẩn lạc này, khuôn được chuẩn bị bằng cách dùng tăm để lấy chất liệu tế bào từ khuẩn lạc trên đĩa thạch và trực tiếp đưa chất liệu tế bào vào ống phản ứng PCR. Chất liệu tế bào được làm ấm trong 10 giây với 800W trong lò vi sóng loại Mikrowave & Grill của SEVERIN Elektrogeräte GmbH (Sundern, Đức) và sau đó bổ sung hóa chất PCR vào khuôn trong ống phản ứng PCR.

f. Tất cả các phản ứng PCR đều được thực hiện trong máy luân nhiệt PCR loại Mastercycler hoặc Mastercycler nexus gradient từ Eppendorf AG (Hamburg, Đức).

Enzym cắt giới hạn ADN

“Endonucleaza giới hạn FastDigest (FD)” (ThermoFisher Scientific, Waltham, Mỹ) hoặc các endonucleaza giới hạn từ New England BioLabs Inc. (Ipswich, Mỹ) được sử dụng để làm enzym cắt giới hạn. Các phản ứng này được thực hiện theo hướng dẫn của nhà sản xuất.

Xác định kích cỡ của đoạn ADN

a. Kích cỡ của các đoạn ADN nhỏ (<1000 cặp bazơ nitơ) thường được xác định bằng cách điện di mao quản tự động, sử dụng QIAxcel của Qiagen (Hilden, Đức).

b. Nếu các đoạn ADN cần được phân lập hoặc nếu các đoạn ADN có kích cỡ >1000 cặp bazơ nitơ thì ADN được tách bằng điện di trên gel TAE agarosa và được nhuộm bằng chất nhuộm gel axit nucleic GelRed® (Biotium, Inc., Fremont, Canada). ADN đã nhuộm được quan sát ở bước sóng 302 nm.

Tinh chế các đoạn khuếch đại PCR và các đoạn đã cắt

Các đoạn khuếch đại PCR và các đoạn đã cắt được tinh chế bằng cách sử dụng Kit tinh chế QIAquick PCR của Qiagen (Hilden, Đức; mã số 28106), theo hướng dẫn của nhà sản xuất. ADN được rửa giải bằng 30 µl Tris*HCl 10 mM (độ pH=8,5).

Xác định hàm lượng ADN

Hàm lượng ADN được đo bằng máy quang phổ NanoDrop ND-1000 của PEQLAB Biotechnologie GmbH, thương hiệu VWR năm 2015 (Erlangen, Đức).

Tách dòng

Vector plasmit được nối bằng cách sử dụng “Kit tách dòng NEBuilder HiFi ADN” được mua từ New England BioLabs Inc. (Ipswich, Mỹ, mã số E5520). Hỗn hợp phản ứng, chứa vectơ mạch thẳng và ít nhất một đoạn chèn ADN, được ủ ở 50°C trong 60 phút. Hỗn hợp tách dòng được sử dụng cho mỗi thí nghiệm biến nạp với lượng 0,5 µl.

Biến nạp *E. coli*

Để tách dòng plasmit, tế bào khả biến “Tế bào khả biến *E. coli* NEB® ổn định

(Hiệu quả cao)“ (New England BioLabs Inc., Ipswich, Mỹ, mã số C3040) đã được biến nạp theo quy trình của nhà sản xuất. Tế bào biến nạp thành công được lựa chọn trên thạch LB có bổ sung kanamycin với lượng 25mg/l.

Biến nạp vào *C. glutamicum*

Thực hiện biến nạp ADN-plasmit vào *C. glutamicum* bằng phương pháp điện di, sử dụng “Gene Pulser Xcell” (Bio-Rad Laboratories GmbH, Feldkirchen, Đức) như được mô tả bởi Ruan và đồng tác giả (2015). Quá trình điện di được thực hiện trong cuvet điện di 1 mm (Bio-Rad Laboratories GmbH, Feldkirchen, Đức) ở 1,8 kV và hằng số thời gian cố định được đặt là 5 mili giây. Tế bào biến nạp được chọn lọc trên thạch BHI chứa sorbitol với lượng 134g/l, chiết xuất nấm men với lượng 2,5g/l và kanamycin với lượng 25mg/l.

Xác định trình tự nucleotit

Trình tự nucleotit của phân tử ADN được Eurofins Genomics GmbH (Ebersberg, Đức) xác định bằng cách xác định trình tự chu kỳ, sử dụng phương pháp kết thúc chuỗi dideoxy của Sanger và đồng tác giả (Proceedings of the National Academy of Sciences USA 74, 5463 – 5467, 1977), trên máy phân tích ADN 3730xl ADN Applied Biosystems® (Carlsbad, CA, Mỹ). Phần mềm Clonemanager Professional 9 của Scientific & Educational Software (Denver, Mỹ) đã được sử dụng để trực quan hóa và đánh giá các trình tự.

Giữ chủng các chủng *E. coli* và *C. glutamicum* trong glyxerol

Đã thực hiện việc giữ chủng trong glyxerol để bảo quản các chủng *E. coli*- và *C. glutamicum* trong thời gian dài. Các dòng *E. coli* chọn lọc được nuôi cấy trong 10ml môi trường LB đã được bổ sung glucoza với lượng 2g/l. Các dòng *C. glutamicum* đã chọn được nuôi cấy trong 10ml môi trường BHI đậm đặc gấp đôi đã được bổ sung glucoza với lượng 2g/l. Nuôi cấy plasmit chứa các chủng *E. coli*- và *C. glutamicum* được bổ sung kanamycin với lượng 25mg/l. Môi trường được chứa trong các bình Erlenmeyer 100ml có 3 vách ngăn. Bước cấy được thực hiện bằng que cấy vi sinh mang theo tế bào lấy từ khuẩn lạc. Sau đó, mẫu nuôi cấy được ủ trong 18 giờ ở 30°C và 200 vòng/phút. Sau thời gian ủ này, glyxerol 85%

vô trùng (v/v) được bổ sung vào mẫu nuôi cấy với lượng 1,2ml. Sau đó, chia huyền phù tế bào chứa glyxerol thu được thành các phần có thể tích 2ml và bảo quản ở -80°C .

Tổng hợp GAA trong nuôi cấy quy mô mililit

Hệ thống nuôi cấy ở quy mô mililit theo Duetz (2007) được sử dụng để đánh giá việc tổng hợp GAA của các chủng này. Nhằm mục đích này, đã sử dụng các vi đĩa 24 giếng sâu (24 đĩa WDS giếng) từ EnzyScreen BV (Heemstede, Hà Lan, mã số CR1424) chứa đầy 2,5ml môi trường cho mỗi giếng.

Bước tiền nuôi cấy của các chủng được thực hiện trong 10ml môi trường nuôi cấy (SM). Môi trường này được đựng trong bình Erlenmeyer 100ml có 3 vách ngăn. Thực hiện nuôi cấy với 100 μl môi trường giữ chủng trong glyxerol và ủ mẻ cấy trong 24 giờ ở 30°C và 200 vòng/phút. Thành phần của môi trường nuôi cấy (SM) được thể hiện trong bảng 5.

Bảng 5: Môi trường nuôi cấy (SM)

Thành phần	Hàm lượng (g/l)
Chiết xuất nấm men FM902 (Angel Yeast Co.,LTD, Hubei, Trung Quốc)	10
Urê	1,5
KH_2PO_4	0,5
K_2HPO_4	0,5
$\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$	1
Biotin	0,0001
Thiamin hydroclorua	0,0001
$\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$	0,01
$\text{MnSO}_4 \cdot \text{H}_2\text{O}$	0,01
Glucoza	20
Độ pH=7,0	

Sau thời gian ủ trên đây, đã xác định mật độ quang học OD_{600} của bước tiền nuôi cấy. Thể tích, cần thiết để cấy 2,5ml môi trường tổng hợp (PM) đến

OD₆₀₀=0,1, được lấy mẫu từ tiền nuôi cấy, ly tâm (1 phút ở 8000 g) và loại bỏ dịch nổi. Sau đó, cấy tế bào trở lại vào 100 µl môi trường tổng hợp.

Quá trình nuôi cấy chính được bắt đầu bằng cách cấy mỗi 100 µl tế bào được tái nuôi cấy từ các mẫu tiền nuôi cấy vào đĩa WDS bao gồm 24 giếng chứa 2,4ml môi trường tổng hợp (PM). Thành phần của môi trường tổng hợp (PM) được thể hiện trong bảng 6.

Bảng 6: Môi trường tổng hợp (PM)

Thành phần	Hàm lượng (g/l)
Axit 3-(N-morpholino)propanesulfonic(MOPS)	40
Chiết xuất nấm men FM902 (Angel Yeast Co., LTD, Hubei, Trung Quốc)	1,5
(NH ₄) ₂ SO ₄	10
NH ₄ Cl	15
Trinatri xitrat * 2 H ₂ O	10
Urê	1
KH ₂ PO ₄	0,5
K ₂ HPO ₄	0,5
MgSO ₄ * 7 H ₂ O	1
Biotin	0,0001
Thiamin hydroclorua	0,0001
FeSO ₄ * 7 H ₂ O	0,01
MnSO ₄ * H ₂ O	0,01
ZnSO ₄ * 7 H ₂ O	0,000015
CuSO ₄ * 5 H ₂ O	0,0004
Chất chống nổi bọt XFO-1501 (Ivanhoe Industries Inc., Zion, Mỹ)	0,5
Glucosa	40
IPTG (Isopropyl β-D-1-thiogalactopyranozit)	0,3 mM
Độ pH=7,2	

Các mẫu cấy chính được ủ trong 72 giờ ở 30°C và 300 vòng/phút trong tủ

ấm lắc tiêu chuẩn Infors HT Multitron của Infors GmbH (Bottmingen, Thụy Sĩ) cho đến khi tiêu thụ hết glucoza. Hàm lượng glucoza trong hỗn dịch này được phân tích bằng máy đo đường huyết OneTouch Vita® của LifeScan (Johnson & Johnson Medical GmbH, Neuss, Đức).

Sau khi nuôi cấy, huyền phù nuôi cấy được chuyển sang tấm vi đĩa giếng sâu. Phần huyền phù nuôi cấy được pha loãng thích hợp để đo OD₆₀₀. Phần nuôi cấy khác được ly tâm và phân tích hàm lượng của GAA trong dịch nổi phía trên như được mô tả dưới đây.

Xác định hàm lượng L-arginin và glyxin trong pepton FM902 của nấm men

Do chiết xuất nấm men FM902 (Angel Yeast Co., LTD, Hubei, Trung Quốc) chứa nhiều peptit và axit amin khác nhau nên hàm lượng L-arginin và glyxin của chiết xuất nấm men này được đo như sau.

Để đo hàm lượng axit amin tự do, các mẫu được chuẩn bị bằng cách hòa tan 1g dịch chiết nấm men trong 20ml nước. Bổ sung nước vào dung dịch này đến tổng thể tích là 25ml, trộn đều và lọc bằng cách sử dụng bộ lọc ống tiêm nylon 0,2µm.

Để đo hàm lượng axit amin tổng số (axit amin tự do cộng với axit amin liên kết trong peptit), các mẫu được chuẩn bị bằng cách hòa tan 1g dịch chiết nấm men trong 10ml HCl 6M và ủ trong 24 giờ ở 110°C. Sau đó, bổ sung nước đến tổng thể tích là 25ml. Trộn đều dung dịch và lọc bằng bộ lọc ống tiêm nylon 0,2µm.

Hàm lượng L-arginin và glyxin trong các mẫu được xác định bằng phương pháp sắc ký trao đổi ion bằng cách sử dụng máy phân tích axit amin SYKAM S433 của SYKAM Vertriebs GmbH (Fürstentfeldbruck, Đức). Cột có thiết bị trao đổi cation gốc polystyren có hình cầu (Peek LCA N04/Na, kích cỡ 150 x 4,6 mm) của SYKAM đã được sử dụng làm pha rắn. Tùy thuộc vào axit L-amino, bước phân tách diễn ra ở chế độ hoạt động không đổi bằng cách sử dụng hỗn hợp bao gồm chất đệm A và B để rửa giải hoặc bằng cách rửa giải gradient, sử dụng bộ đệm này. Dung dịch bao gồm trinati xitrat với lượng 263g, axit xitric với lượng

120g, metanol với lượng 1100ml, HCl 37% với lượng 100ml và axit octanoic với lượng 2ml (độ pH cuối cùng là 3,5) trong 20l nước được dùng làm chất đệm A. Dung dịch bao gồm 392g trinitrat, 100g axit boric và 2ml axit octanoic (độ pH cuối cùng là 10,2) trong 20l nước được sử dụng làm chất đệm B. Axit amin tự do được bắt màu với ninhydrin thông qua quá trình tạo dẫn xuất sau cột và được phát hiện bằng phương pháp đo quang ở bước sóng 570 nm.

Bảng 7 thể hiện hàm lượng L-arginin và glyxin tổng số và tự do, như được xác định trong chiết xuất nấm men FM902 (Angel Yeast Co., LTD, Hubei, Trung Quốc), cũng như lượng hàm lượng L-arginin và glyxin tổng số và tự do thu được trong môi trường tổng hợp (PM).

Bảng 7: Hàm lượng L-arginin và glyxin trong chiết xuất nấm men (YE) FM902 và hàm lượng L-arginin và glyxin thu được trong môi trường tổng hợp (PM) chứa 1,5g/l YE.

Amino axit	Amino axit tự do trong YE	Amino axit tổng số trong YE	Amino axit tự do thu được trong PM	Amino axit tổng số thu được trong PM
L-arginin	15,1 g/kg	32,1 g/kg	22,7 mg/l	48,2 mg/l
glyxin	6,9 g/kg	30,5 g/kg	10,4 mg/l	45,7 mg/l

Định lượng GAA

Các mẫu được phân tích bằng hệ thống phân tích Agilent, bao gồm HPLC “Infinity 1260” cùng với máy phân tích khối lượng “Triple Quad 6420” (Agilent Technologies Inc., Santa Clara, Mỹ). Bước tách sắc ký được thực hiện trên cột Atlantis HILIC Silica, 4,6X250mm, 5 μ m (Waters Corporation, Milford, Mỹ) ở 35°C. Pha động A là nước cùng với amoni fomat 10mM và axit fomic 0,2%. Pha động B là hỗn hợp bao gồm 90% axetonitril và 10% nước, bổ sung amoni fomat 10 mM vào hỗn hợp này. Hệ thống HPLC được khởi động với 100% B, tiếp theo là gradient tuyến tính trong 22 phút và tốc độ dòng không đổi là 0,6ml/phút đến 66% B. Máy phân tích khối lượng được vận hành ở chế độ ion hóa ESI dương. Để phát hiện GAA, trị số m/z được theo dõi bằng cách sử dụng phân đoạn MRM [M + H] + 118 - 76. Giới hạn định lượng (LOQ) đối với GAA

được cố định đến 7ppm.

B) Kết quả thử nghiệm

Ví dụ 1: Tổng hợp gen mã hóa cho L-arginin:glyxin amidinotransferaza (AGAT, EC 2.1.4.1) từ các sinh vật khác nhau

Moorea producens là vi khuẩn lam dạng sợi. Hệ gen của chủng *Moorea producens* PAL-8-15-08-1 đã được công bố bởi Leao và đồng tác giả (Leao T, Castelão G, Korobeynikov A, Monroe EA, Podell S, Glukhov E, Allen EE, Gerwick WH, Gerwick L, Proc Natl Acad Sci USA. 2017 Mar 21;114(12): 3198-3203. doi: 10,1073/pnas,1618556114; số đăng ký CP017599.1). Hệ gen này chứa khung đọc mở được mã hóa, được coi là của L-arginin:glyxin amidinotransferaza (AGAT, EC 2.1.4.1, số đăng ký BJP34_00300, SEQ ID NO:1). SEQ ID NO:2 và SEQ ID NO:4 thể hiện trình tự mã hóa cho axit amin (số đăng ký WP_070390602) được xác định là AGAT_Mp.

Gen *cyrA* từ *Cylindrospermopsis raciborskii* AWT205 (số đăng ký EU140798.1) mã hóa cho L-arginin:glyxin amidinotransferaza (Mihali TK, Kellmann R, Muenchhoff J, Barrow KD, Neilan BA (2008) “Characterization of gen cluster responsible for cylindrospermopsin biosynthesis.”, Appl Environ Microbiol., 74(3): 716-22, doi: 10,1128/AEM,01988-07; SEQ ID NO:15). SEQ ID NO:16 và SEQ ID NO:26 thể hiện trình tự axit amin suy ra (số đăng ký ABX60160), được xác định là AGAT_cyrA.

Trình tự cADN của L-arginin ở người: glyxin amidinotransferaza đã được Humm và đồng tác giả, 1994 mô tả (Humm A, Huber R, Mann K (1994), The amino acid sequences of human and pig l-arginine: glycine amidinotransferase.”, FEBS Letters, Tập 339(1-2), 101-107, DOI: 10.1016/0014-5793(94)80394-3; số đăng ký NM_001482.3, SEQ ID NO:17). Trình tự axit amin (số đăng ký NP_001473.1, SEQ ID NO:18) có nguồn gốc bắt đầu bằng peptit vận chuyển của ty thể (axit amin 1 - 37) không có mặt trong enzym trưởng thành. Enzym bị cắt ngắn, bắt đầu từ axit amin 56, được phát hiện là có hoạt tính khi được biểu hiện ở *E. coli* (Humm A, Fritsche E, Mann K, Göhl M, Huber R (1997) “Recombinant expression and isolation of human L-arginine: glycine

amidinotransferase and identification of its active-site cysteine residue.” Biochem. J. 322, 771-776, DOI: 10.1042/bj3220771). Sự hợp nhất ở đầu tận cùng N của đuôi 7 axit amin (SEQ ID NO:19) đã được chứng minh là cải thiện sự biểu hiện của protein ở *E. coli* (Hansted JG, Pietikäinen L, Hög F, Sperling-Petersen HU, Mortensen KK (2011) “Expressivity tag: A novel tool for increased expression in *Escherichia coli*.” Journal of Biotechnology 155 (2011) 275– 283, DOI: 10.1016/j.jbiotec.2011.07.013). Do đó, protein dung hợp bao gồm đầu và AGAT bị cắt ngắn đã được thiết kế và được xác định là AGAT_Hs (SEQ ID NO:20 và SEQ ID NO:28).

Trình tự axit amin của L-arginin:glyxin amidinotransferaza từ *Rattus norvegicus* (số đăng ký NP_112293.1, SEQ ID NO:21) rất giống với enzym của người. Như được mô tả đối với enzym của người, trình tự được sử dụng để thiết kế protein dung hợp bao gồm đầu biểu hiện ở đầu tận cùng N và trình tự enzym bị cắt ngắn. Đã xác định protein dung hợp thu được là AGAT_Rn (SEQ ID NO:22 và SEQ ID NO:30).

Chồn bay sunda *Galeopterus variegatus* có L-arginin:glyxin amidinotransferaza dự đoán (số đăng ký NP_112293.1, SEQ ID NO:23). Như được mô tả đối với enzym của người, trình tự axit amin của nó được sử dụng để thiết kế protein dung hợp bao gồm đầu biểu hiện ở đầu tận cùng N và trình tự enzym bị cắt ngắn. Đã xác định protein dung hợp thu được là AGAT_Gv (SEQ ID NO:24 và SEQ ID NO:32).

Bằng cách sử dụng công cụ phần mềm “GeneOptimizer” (Geneart/ThermoFisher Scientific, Waltham, Mỹ), trình tự axit amin của AGAT_Mp, AGAT_cyrA, AGAT_Hs, AGAT_Rn và AGAT_Gv đã được dịch mã ngược lại thành trình tự ADN và được tối ưu hóa để sử dụng mã bộ ba của *C. glutamicum*. Phần cuối của chúng được mở rộng với các trình tự để tách dòng không sử dụng enzym giới hạn và trình tự Shine-Dalgarno được bổ sung vào phía ngược dòng của khung đọc mở. Các trình tự ADN thu được là SEQ ID NO:3 (mã hóa cho AGAT_Mp), SEQ ID NO:25 (mã hóa cho AGAT_cyrA), SEQ ID NO:27 (mã hóa cho AGAT_Hs), SEQ ID NO:29 (mã hóa cho AGAT_Rn) và

SEQ ID NO:31 (mã hóa cho AGAT_Gv). Chúng được sắp xếp để tổng hợp gen từ Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ). Các gen tổng hợp được phân phối là các phần của plasmit tách dòng được ký hiệu là pMA-T_AGAT_Mp, pMA-T_AGAT_cyrA, pMA-T_AGAT_Hs, pMA-T_AGAT_Rn và pMA-T_AGAT_Gv.

Ví dụ 2: Tách dòng AGAT_Mp vào plasmit biểu hiện pEC-XK99E

Plasmit con thoi pEC-XK99E của *E. coli*-*C. glutamicum* được cắt bằng cách sử dụng endonucleaza giới hạn SmaI. Đuôi phosphat được loại bỏ bằng cách sử dụng “Phosphatase kiềm nhạy với nhiệt FastAP” (Thermo Fisher Scientific, Waltham, Mỹ). Sau đó, ADN này được tinh chế bằng “Kit tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức).

Plasmit tách dòng pMA-T_AGAT_Mp được cắt bằng mlul + AatII và cắt bằng các đoạn thu được bằng cách sử dụng “Kit sửa chữa nhanh đầu tận cùng của ADN” (Thermo Fisher Scientific, Waltham, Mỹ). Chúng được phân tách bằng cách điện di trên gel agarosa (agarosa 0,8% trong đệm TAE) và cắt đoạn tương ứng với “AGAT_Mp” (1174 bp). ADN của nó đã được tinh chế bằng cách sử dụng “Kit Tinh chế trên Gel QIAquick” (Qiagen GmbH, Hilden, Đức).

Phân tích đoạn AGAT_Mp và pEC-XK99E đã được chuyển thành mạch thẳng bằng cách sử dụng “ADN ligaza T4 sẵn sàng hoạt động” (GE Healthcare Europe GmbH, Freiburg, Đức). Sản phẩm nối được biến nạp vào “Tế bào khả biến thích hợp NEB của *E. coli* (Hiệu suất cao)” (New England Biolabs, Ipswich, Mỹ) và các tế bào này được nuôi cấy trên thạch LB chứa kanamycin 25mg/l. Các dòng thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pEC- XK99E_AGAT_Mp.

Ví dụ 2.1: Tách dòng AGAT_Mp và AGAT_cyrA vào plasmit biểu hiện pEKEx2

Plasmit con thoi pEKEx2 (Eikmanns, 1991) của *E. coli*-*C. glutamicum* được cắt bằng cách sử dụng endonucleaza giới hạn PstI. Cắt bằng các đoạn thu được bằng cách sử dụng “Kit sửa chữa nhanh đầu tận cùng ADN” (Thermo Fisher Scientific) và đuôi phosphat được loại bỏ bằng “Phosphatase kiềm nhạy với nhiệt

FastAP” (Thermo Fisher Scientific, Waltham, Mỹ). Sau đó, ADN được tinh chế bằng “Kit tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức).

Các plasmit tách dòng pMA-T_AGAT_Mp và pMA-T_AGAT_cyrA đã được cắt bằng *mluI* + *AatII* và cắt bằng các đoạn thu được bằng cách sử dụng “Kit sửa chữa nhanh đầu tận cùng ADN” (Thermo Fisher Scientific, Waltham, Mỹ). Chúng được phân tách bằng điện di trên gel agarosa (agarosa 0,8% trong đệm TAE) và cắt các đoạn tương ứng với AGAT_Mp (1174 bp) và AGAT_cyrA (1204 bp). Tinh chế ADN này bằng cách sử dụng “Kit tách chiết gel QIAquick” (Qiagen GmbH, Hilden, Đức).

Mỗi đoạn AGAT được nối với pEKEx2 đã được chuyển thành mạch thẳng bằng cách sử dụng “ADN ligaza T4 sẵn sàng hoạt động” (GE Healthcare Europe GmbH, Freiburg, Đức). Các sản phẩm nối được biến nạp vào tế bào khả biến ổn định NEB của *E. coli* (Hiệu suất cao) “(New England Biolabs, Ipswich, Mỹ) và các tế bào này được nuôi cấy trên thạch LB chứa kanamycin 25mg/l. Các dòng thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Các plasmit thu được được gọi là pEKEx2_AGAT_Mp và pEKEx2_AGAT_cyrA.

Ví dụ 2.2: Tách dòng AGAT_Hs, AGAT_Rn và AGAT_Gv vào plasmit biểu hiện pEKEx2

Các plasmit tách dòng pMA-T_AGAT_Hs, pMA-T_AGAT_Rn và pMA-T_AGAT_Gv được cắt bằng *Eco31I* và các sản phẩm này được tinh chế bằng cách sử dụng “Kit tinh chế trên gel QIAquick” (Qiagen GmbH, Hilden, Đức).

Plasmit con thoi *E. coli*-*C. glutamicum* pEKEx2 (Eikmanns BJ, Kleinertz E, Liebl W, Sahm H (1991), “A family of *Corynebacterium glutamicum*/Escherichia coli shuttle vectors for cloning, controlled gene expression, and promoter probing.”, *Gene*. 1991 Jun 15;102(1): 93-8) được cắt bằng cách sử dụng endonucleaza giới hạn *SbfI* và *BamHI*. ADN này đã được tinh chế bằng “Kit tinh chế QIAquick PCR” (Qiagen GmbH, Hilden, Đức).

Mỗi đoạn AGAT được nối bằng pEKEx2 đã cắt bằng cách sử dụng “Kit tách

dòng ADN không sử dụng enzym giới hạn NEBuilder HiFi” (New England BioLabs Inc., Ipswich, Mỹ, mã số E5520). Các sản phẩm nối được biến nạp vào "Tế bào khả biến thích hợp NEB của *E. coli* (Hiệu suất cao)" (New England Biolabs, Ipswich, Mỹ) và các tế bào được nuôi cấy trên thạch LB chứa kanamycin 25mg/l. Các dòng thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Các plasmit thu được lần lượt được xác định là pEKEx2_AGAT_Hs, pEKEx2_AGAT_Rn và pEKEx2_AGAT_Gv.

Ví dụ 3: Tách dòng gen *argF* vào plasmit pCR-Đầu tù II-TOPO

Gen *argF* được khuếch đại bằng phương pháp PCR nhờ Kit ADN Polymeraza Phusion có độ trung thực cao (New England BioLabs Inc., Ipswich, Mỹ) sử dụng ADN hệ gen của *C. glutamicum* ATCC13032 và mỗi oligonucleotit *argF_1.p* (SEQ ID NO:5) và *argF_2.p* (SEQ ID NO:6). Tách dòng sản phẩm PCR thu được trong plasmit pCR-Đầu tù II-TOPO (Thermo Fisher Scientific/Invitrogen, Waltham, Mỹ) và dòng plasmit thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Plasmit này được gọi là pCRII-*argF*.

Ví dụ 4: Tách dòng gen *argG* và *argH* vào plasmit pCR-Đầu tù II-TOPO

Các gen *argG* và *argH* đã được khuếch đại bằng phương pháp PCR nhờ Kit ADN Polymeraza Phusion có độ trung thực cao (New England BioLabs Inc., Ipswich, Mỹ) sử dụng ADN hệ gen của *C. glutamicum* ATCC13032 và mỗi oligonucleotit *argG_1.p* (SEQ ID NO:7) và *argH_2.p* (SEQ ID NO:8). Tách dòng sản phẩm PCR thu được trong plasmit pCR-Đầu tù II-TOPO (Thermo Fisher Scientific/Invitrogen, Waltham, Mỹ) và dòng plasmit thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Plasmit này được gọi là pCRII-*argGH*.

Ví dụ 5: Tách dòng gen *argG* và *argH* vào plasmit pCRII-*argF*

pCRII-*argGH* được cắt bằng cách sử dụng HpaI + AvrII và đoạn giới hạn có kích cỡ 2773 cặp bazơ nitơ được phân lập từ gel agarosa. pCRII-*argF* được cắt bằng cách sử dụng SspI + AvrII và đoạn giới hạn có kích cỡ 4526 cặp bazơ

nitor được phân lập từ gel agarosa. Cả hai đoạn này đều được nối và sau đó được biến nạp vào *E. coli*. Các dòng plasmit thích hợp được xác định bằng cách cắt bằng enzym giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pCRII-argFGH.

Ví dụ 6: Tách dòng gen *argF*, *argG* và *argH* vào plasmit biểu hiện pEC-XK99E

pCRII-argFGH được cắt bằng cách sử dụng HpaI + AvrII và đoạn giới hạn có kích cỡ 2773 cặp bazơ nitor được phân lập từ gel agarosa. Plasmit pEC-XK99E được cắt bằng Ecl136II + XbaI. Đoạn giới hạn có kích cỡ 6999 cặp bazơ nitor được phân lập từ gel agarosa. Cả hai đoạn này đều được nối và sau đó được biến nạp vào *E. coli*. Các dòng plasmit thích hợp được xác định bằng cách cắt giới hạn và xác định trình tự ADN. Plasmit pEC-XK99E_argFGH thu được chứa các gen *argF*, *argG* và *argH* từ *C. glutamicum*.

Ví dụ 7: Tách dòng gen *argF*, *argG* và *argH* vào plasmit biểu hiện pEC-XK99E_AGAT_Mp

pCRII-argFGH được cắt bằng cách sử dụng XbaI + SpeI và đoạn giới hạn có kích cỡ 3868 cặp bazơ nitor được phân lập từ gel agarosa. Plasmit pEC-XK99E_AGAT_Mp được cắt bằng XbaI. Đoạn giới hạn có kích cỡ 8188 cặp bazơ nitor được phân lập từ gel agarosa. Cả hai đoạn này đều được nối và sau đó được biến nạp vào *E. coli*. Các dòng plasmit thích hợp được xác định bằng cách cắt giới hạn và xác định trình tự ADN. Plasmit pEC-XK99E_AGAT_Mp_argFGH thu được chứa các gen *argF*, *argG* và *argH* từ *C. glutamicum* kết hợp với AGAT_Mp.

Ví dụ 8: Tách dòng gen *argF* vào plasmit biểu hiện pEC-XK99E_AGAT_Mp

pCRII-*argF* đã được cắt bằng KpnI + XbaI + AseI và ADN này được tinh chế bằng “Kit tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức). Plasmit pEC-XK99E_AGAT_Mp được cắt bằng KpnI + XbaI, ADN được tinh chế bằng “Kit tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức). Cả hai chất rửa giải được trộn lẫn, các đoạn ADN được nối với nhau và sản phẩm được sử dụng để biến nạp *E. coli*. Các dòng plasmit thích hợp được xác định bằng cách cắt giới hạn

và xác định trình tự ADN. Plasmid pEC-XK99E_AGAT_Mp_argF thu được chứa gen *argF* từ *C. glutamicum* kết hợp với AGAT_Mp.

Ví dụ 9: Tách dòng gen *argG* vào plasmid biểu hiện pEC-XK99E_AGAT_Mp

pCRII-*argGH* được cắt bằng cách sử dụng XbaI + SalI và đoạn giới hạn có kích cỡ 1798 cặp bazơ nitơ được phân lập từ gel agarosa. Plasmid pEC-XK99E_AGAT_Mp được cắt bằng cách sử dụng XbaI + SalI, ADN được tinh chế bằng “Kit Tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức). Các đoạn ADN được nối lại và sản phẩm này được sử dụng để biến nạp vào *E. coli*. Các dòng plasmid thích hợp được xác định bằng cách cắt giới hạn và xác định trình tự ADN. Plasmid pEC-XK99E_AGAT_Mp_argG thu được chứa gen *argG* từ *C. glutamicum* kết hợp với AGAT_Mp.

Ví dụ 10: Chèn nhiễm sắc thể của vùng khởi động *sod* ngược dòng với operon *carAB* vào ATCC13032

Hoạt tính enzym của carbamoyl phosphat synthetaza được tăng lên bằng cách đưa operon *carAB* trong ATCC13032 vào hệ gen vùng khởi động *sod* ngược dòng. Do đó, plasmid pK18mobsacB_Psod-*carAB* được thiết kế như sau. Plasmid pK18mobsacB được cắt bằng EcoRI + HindIII và vector ADN đã được chuyển thành mạch thẳng (5670 cặp bazơ nitơ) được cắt ra khỏi gel agarosa. ADN này được tinh chế bằng cách sử dụng “Kit tinh chế trên gel QIAquick” (QIAGEN GmbH, Hilden, Đức).

Để thiết kế đoạn chèn, ba đoạn ADN đã được tổng hợp bằng phương pháp PCR với các cặp mồi sau (ADN hệ gen của ATCC13032 được dùng làm khuôn mẫu):

PsodcarAB-LA-F (SEQ ID NO: 9) + PsodcarAB-LA-R (SEQ ID NO:10) = nhánh tương ứng bên trái (1025 cặp bazơ nitơ)

PsodcarAB-F (SEQ ID NO: 11) + PsodcarAB-R (SEQ ID NO:12) = *sod*-vùng khởi động (250 cặp bazơ nitơ)

PsodcarAB-RA-F SEQ ID NO: 13) + PsodcarAB-RA-R (SEQ ID NO:14) = nhánh tương ứng bên phải (944 cặp bazơ nitơ)

Các ADN sản phẩm được tinh chế bằng cách sử dụng “Kit tinh chế PCR QIAquick” (Qiagen GmbH, Hilden, Đức). Sau đó, plasmit đã được chuyển thành mạch thẳng và các sản phẩm PCR được nối bằng cách sử dụng “Kit tách dòng không sử dụng enzym giới hạn NEBuilder HiFi ADN” (New England BioLabs Inc., Ipswich, Mỹ, mã số E5520). Các dòng plasmit thích hợp được xác định bằng cách cắt giới hạn và xác định trình tự ADN.

pK18mobsacB_Psod-carAB được sử dụng để hợp nhất *sod* - vùng khởi động mạnh ngược dòng của gen *carAB* vào hệ gen của *C. glutamicum* ATCC13032. Plasmit này được biến nạp vào ATCC13032 bằng phương pháp điện di. Bước hợp nhất nhiễm sắc thể (kết quả từ lần tái tổ hợp thứ nhất) được chọn bằng cách cấy trải trên thạch BHI có bổ sung 134g/l sorbitol, 2,5g/l chiết xuất nấm men và kanamycin 25mg/l. Các đĩa thạch này được ủ trong 48 giờ ở 33°C.

Các khuẩn lạc riêng lẻ được chuyển lên đĩa thạch mới (chứa kanamycin 25mg/l) và ủ trong 24 giờ ở 33°C. Các mẫu nuôi cấy lỏng của các dòng này được nuôi cấy trong 24 giờ ở 33°C trong 10ml môi trường BHI, được chứa trong các bình Erlenmeyer 100ml có 3 vách ngăn. Để phân lập các dòng đã gặp trong lần tái tổ hợp thứ hai, phần mẫu đã lấy từ mỗi dịch nuôi cấy lỏng được pha loãng và cấy trải thích hợp (thường từ 100 đến 200 μ l) trên thạch BHI được bổ sung 10% sacaroza. Các đĩa thạch này được ủ trong 48 giờ ở 33°C. Sau đó, các khuẩn lạc phát triển trên đĩa thạch chứa sacaroza được kiểm tra độ nhạy với kanamycin. Để làm như vậy, đã sử dụng tăm để lấy chất liệu tế bào khỏi khuẩn lạc và chuyển chất liệu tế bào này lên thạch BHI chứa kanamycin 25mg/l và lên thạch BHI chứa 10% sacaroza. Các đĩa thạch này được ủ trong 60 giờ ở 33°C. Các dòng được chứng minh là nhạy cảm với kanamycin và kháng sacaroza đã được kiểm tra bằng phương pháp PCR và xác định trình tự ADN để tìm sự hợp nhất thích hợp của vùng khởi động *sod*. Chúng thu được được gọi là ATCC13032_Psod-carAB.

Bảng 8: Danh mục các chủng

Chủng	Chú thích
<i>Escherichia coli</i>	
NEB® ổn định	Chủng tách dòng thương mại (New England BioLabs Inc., Ipswich, Mỹ)
MG1655	Chủng <i>E. coli</i> viện dẫn (Blattner và đồng tác giả, 1997*)
<i>Corynebacterium glutamicum</i>	
ATCC13032	Chủng kiểu đại <i>Corynebacterium glutamicum</i> (Kinoshita và đồng tác giả, 1957**)
ATCC21831	Biến thể <i>C. glutamicum</i> tổng hợp L-arginin (Park và đồng tác giả, 2014***))
ATCC13032_Psod-carAB	Tích hợp vùng khởi động mạnh <i>sod</i> vào nhiễm sắc thể, ngược dòng với <i>carAB</i> ở <i>C. glutamicum</i> ATCC13032

*) Blattner và đồng tác giả, 1997: Blattner FR, Plunkett G thứ 3, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y, Science. 5 tháng 9 năm 1997;277(5331): 1453-62. (doi: 10.1126/science.277.5331.1453)

**) Kinoshita S, Udaka S, Shimono M., J. Gen. Appl. Microbiol. 1957; 3(3): 193-205.

***)) Park và đồng tác giả, 2014: Park SH, Kim HU, Kim TY, Park JS, Kim SS, Lee, Nat Commun. 5 tháng 8 năm 2014; 5: 4618. (doi: 10.1038/ncomms5618)

Bảng 9: Danh mục plasmid

Plasmid	Chú thích
pMA-T_AGAT_Mp	Tổng hợp gen và vector tách dòng cho AGAT_Mp (<i>Moorea producens</i>); plasmid độc quyền của Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ)
pMA-T_AGAT_cyrA	Tổng hợp gen và vector tách dòng cho AGAT_cyrA (<i>Cylindrospermopsis raciborskii</i> AWT205); plasmid độc quyền của Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ)
pMA-T_AGAT_Hs	Tổng hợp gen và vector tách dòng cho AGAT_Hs (<i>Homo sapiens</i>); plasmid độc quyền của Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ)
pMA-T_AGAT_Rn	Tổng hợp gen và vector tách dòng cho AGAT_Rn (<i>Rattus norvegicus</i>); plasmid độc quyền của Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ)
pMA-T_AGAT_Gv	Tổng hợp gen và vector tách dòng cho AGAT_Gv (<i>Galeopterus variegatus</i>); plasmid độc quyền của Invitrogen/Geneart (Thermo Fisher Scientific, Waltham, Mỹ)
pEKEx2	Plasmid con thoi rỗng của <i>E. coli</i> – <i>C. Glutamicum</i>
pEKEx2_AGAT_Mp	Biểu hiện AGAT_Mp (<i>Moorea producens</i>)
pEKEx2_AGAT_cyrA	Biểu hiện AGAT_cyrA (<i>Cylindrospermopsis raciborskii</i> AWT205)
pEKEx2_AGAT_Hs	Biểu hiện AGAT_Hs (<i>Homo sapiens</i>)
pEKEx2_AGAT_Rn	Biểu hiện AGAT_Rn (<i>Rattus norvegicus</i>)
pEKEx2_AGAT_Gv	Biểu hiện AGAT_Gv (<i>Galeopterus variegatus</i>)
pEC-XK99E	Plasmid con thoi rỗng của <i>E. coli</i> – <i>C. glutamicum</i>
pEC-XK99E_AGAT_Mp	Biểu hiện AGAT_Mp (<i>Moorea producens</i>)
pCR-Đầu tù II-TOPO	Plasmid rỗng để tách dòng sản phẩm PCR (Thermo Fisher Scientific/Invitrogen, Waltham, Mỹ)
pCRII-argF	pCR-Đầu tù II-TOPO chứa gen <i>argF</i> đã tách dòng (<i>C. glutamicum</i>)
pCRII-argGH	pCR-Đầu tù II-TOPO chứa các gen <i>argG</i> , <i>argH</i> đã tách dòng (<i>C. glutamicum</i>)
pCRII-argFGH	pCR-Đầu tù II-TOPO chứa các gen đã tách dòng <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. glutamicum</i>)
pEC-XK99E_argFGH	Biểu hiện <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. glutamicum</i>)

pEC-XK99E_AGAT_Mp_argFGH	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>C. glutamicum</i>)
pEC-XK99E_AGAT_Mp_argGH	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argG</i> , <i>argH</i> (<i>C. glutamicum</i>)
pEC-XK99E_AGAT_Mp_argF	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argF</i> (<i>C. glutamicum</i>)
pEC-XK99E_AGAT_Mp_argG	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argG</i> (<i>C. glutamicum</i>)
pK18mobsacB	Plasmit rỗng để biến đổi hệ gen của <i>C. glutamicum</i>
pK18mobsacB_Psod-carAB	Plasmit để hợp nhất vùng khởi động mạnh <i>sod</i> ngược dòng của <i>carAB</i> ở <i>C. glutamicum</i>

Ví dụ 11: Biến nạp các chủng *C. glutamicum* bằng các plasmit biểu hiện khác nhau:

Các chủng *C. glutamicum* sau đây đã được biến nạp bằng các plasmit khác nhau:

- *C. glutamicum* ATCC13032 (Kinoshita và đồng tác giả, J. Gen. Appl. Microbiol. 1957; 3(3): 193-205) là chủng kiểu đại thường được sử dụng.
- *C. glutamicum* chủng ATCC21831 (Park và đồng tác giả, Nat Commun. 5 tháng 8 năm 2014; 5: 4618) tổng hợp L-arginin từ cơ chất ban đầu như amoniac và glucoza.
- ATCC13032_Psod-carAB là biến thể của ATCC13032 chứa vùng khởi động mạnh *sod* ngược dòng của gen *carAB*.

Các chủng này được biến nạp bằng cách điện di với các plasmit khác nhau (như được thể hiện trong bảng 10). Tế bào chứa plasmit được chọn lọc bằng kanamycin 25mg/l.

Bảng 10: Danh mục các chủng *C. glutamicum* chứa plasmit

Plasmit	Gen đã tách dòng	Chủng nhận	Chủng thu được
pEKEEx2	không có	ATCC21831	ATCC21831/pEKEEx2
pEKEEx2_AGAT_Mp	AGAT_Mp	ATCC21831	ATCC21831/pEKEEx2_AGAT_Mp
pEKEEx2_AGAT_Hs	AGAT_Hs	ATCC21831	ATCC21831/pEKEEx2_AGAT_Hs
pEKEEx2_AGAT_Rn	AGAT_Rn	ATCC21831	ATCC21831/pEKEEx2_AGAT_Rn
pEKEEx2_AGAT_Gv	AGAT_Gv	ATCC21831	ATCC21831/pEKEEx2_AGAT_Gv
pEKEEx2_AGAT_cyrA	AGAT_cyrA	ATCC21831	ATCC21831/pEKEEx2_AGAT_cyrA
pEC-XK99E	không có	ATCC13032	ATCC13032/pEC-XK99E
pEC-XK99E_argFGH	<i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_argFGH
pEC-XK99E_AGAT_Mp	AGAT_Mp	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp
pEC-XK99E_AGAT_Mp_argG	AGAT_Mp, <i>argG</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argG
pEC-XK99E_AGAT_Mp_argF	AGAT_Mp, <i>argF</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argF
pEC-XK99E_AGAT_Mp_argGH	AGAT_Mp, <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argGH
pEC-XK99E_AGAT_Mp_argFGH	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032	ATCC13032/pEC-XK99E_AGAT_Mp_argFGH
pEC-XK99E	không có	ATCC13032_Psod-carAB	ATCC13032_Psod-carAB/pEC-XK99E
pEC-XK99E_AGAT_Mp_argFGH	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i>	ATCC13032_Psod-carAB	ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH

Ví dụ 12: Tác động của AGAT và độ khả dụng của cơ chất đối với sự tổng hợp GAA

Các chủng ATCC13032/pEC-XK99E_AGAT_Mp (mang gen mã hóa cho enzym AGAT từ *Moorea producens*) và ATCC13032/pEC-XK99E (vector rỗng để đối chứng) đã được phân tích khả năng tổng hợp GAA trong quá trình nuôi cấy theo mẻ bằng cách sử dụng hệ thống Wouter Duetz (như được mô tả trên đây). Môi trường tổng hợp (PM) chứa nguồn cacbon chính là D-glucosa với lượng 40g/l. Một số mẻ được bổ sung L-arginin và/hoặc glyxin theo hướng dẫn.

Bảng 11: Tổng hợp GAA bằng các chủng ATCC13032/pEC-XK99E và

ATCC13032/pEC-XK99E_AGAT_Mp

Chủng	Bổ sung	GAA
ATCC13032/pEC-XK99E	không bổ sung	không phát hiện được
ATCC13032/pEC-XK99E	1,90 g/l Arg và 0,818 g/l Gly	không phát hiện được
ATCC13032/pEC-XK99E_AGAT_Mp	không bổ sung	25 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	0,818 g/l Gly	31 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	1,90 g/l Arg và 0,818 g/l Gly	124 mg/l

Như được thể hiện trong bảng 11, chủng đối chứng ATCC13032/pEC-XK99E không thể tổng hợp GAA, ngay cả khi đã bổ sung các tiền chất L-arginin và glyxin. Các tác giả sáng chế kết luận rằng bản thân chủng này không có hoạt tính AGAT. Chủng ATCC13032/pEC-XK99E_AGAT_Mp chứa polynucleotit mã hóa cho AGAT giả định từ *Moorea producens*. Chủng này tổng hợp GAA với lượng 25mg/l trong PM không được bổ sung. Việc bổ sung glyxin làm lượng GAA tăng nhẹ lên 31mg/l. Việc bổ sung glyxin và L-arginin làm gia tăng mạnh lượng GAA lên 124mg/l.

Ví dụ 13: Tổng hợp GAA từ cơ chất ban đầu

Trong quy trình tổng hợp GAA trong công nghiệp, việc bổ sung L-arginin sẽ khá tốn kém so với các cơ chất ban đầu như amoniac, urê và glucoza. Do đó, vẫn có nhu cầu trực tiếp tạo ra GAA từ các cơ chất ban đầu này.

Nhằm mục đích này, chủng tổng hợp L-arginin *C. glutamicum* ATCC21831 đã được biến đổi bằng

- pEKEx2 (vector rỗng để đối chứng),

- pEKEEx2_AGAT_Mp (chứa gen AGAT_Mp từ *Morea producens*),
- pEKEEx2_AGAT_Hs (chứa gen AGAT_Hs từ *Homo sapiens*),
- pEKEEx2_AGAT_Rn (chứa gen AGAT_Rn từ *Rattus norvegicus*),
- pEKEEx2_AGAT_Gv (chứa gen AGAT_Gv từ *Galeopterus variegatus*) và
- pEKEEx2_AGAT_cyrA (chứa gen AGAT_cyrA từ *Cylindrospermopsis raciborskii*).

ATCC21831 được phân lập thành chủng đột biến kháng canavanin và đã phát hiện ra rằng chủng này tổng hợp L-arginin. Hệ gen của chủng này đã được xác định trình tự bởi Park và đồng tác giả (Nat Commun. 5 tháng 8 năm 2014;5: 4618. doi: 10.1038/ncomms5618; số đăng ký CP007722) và chủng này được công bố rộng rãi từ LGC Standards (LGC Standards GmbH, Wesel, Đức).

Tất cả các chủng ATCC21831 đã biến đổi đều được phân tích khả năng tổng hợp GAA trong quá trình nuôi cấy theo mẻ bằng cách sử dụng hệ thống Wouter Duetz (như được mô tả trên đây). Môi trường tổng hợp (PM) chứa D-glucosa với lượng 40g/l là nguồn cacbon chính. Một số mẻ được bổ sung glyxin và/hoặc L-arginin.

Bảng 12: Tổng hợp GAA từ các chủng ATCC21831 đã được biến nạp vectơ pEKEEx2 chứa AGAT từ các loài khác nhau trong môi trường CGAF và MOPS chứa dịch chiết nấm men với lượng 1,5g/l

Chủng	Bổ sung	GAA
ATCC21831/pEKEEx2	không bổ sung	không phát hiện được
ATCC21831/pEKEEx2	1,90g/l arginin, 0,82g/l glyxin	không phát hiện được
ATCC21831/pEKEEx2_AGAT_Mp	không bổ sung	26,5mg/l
ATCC21831/pEKEEx2_AGAT_Mp	0,82g/l glyxin	166,3mg/l
ATCC21831/pEKEEx2_AGAT_Hs	không bổ sung	3,6mg/l
ATCC21831/pEKEEx2_AGAT_Hs	0,82g/l glyxin	56,2mg/l
ATCC21831/pEKEEx2_AGAT_Rn	không bổ sung	0,9mg/l
ATCC21831/pEKEEx2_AGAT_Rn	0,82g/l glyxin	7,6mg/l

ATCC21831/pEKEEx2_AGAT_Gv	không bổ sung	0,7mg/l
ATCC21831/pEKEEx2_AGAT_Gv	0,82g/l glyxin	8,5mg/l
ATCC21831/pEKEEx2_AGAT_cyrA	không bổ sung	20,8mg/l
ATCC21831/pEKEEx2_AGAT_cyrA	0,82g/l glyxin	52,7mg/l

Như được thể hiện trong bảng 12, ATCC21831/pEKEEx2 không tổng hợp GAA, ngay cả khi có mặt các tiền chất L-arginin và glyxin. Các tác giả sáng chế kết luận rằng bản thân ATCC21831/pEKEEx2 không có hoạt tính AGAT. Các chủng đã biến nạp ATCC21831/pEKEEx2_AGAT_Mp, ATCC21831/pEKEEx2_AGAT_Hs, ATCC21831/pEKEEx2_AGAT_Rn, ATCC21831/pEKEEx2_AGAT_Gv và ATCC21831/pEKEEx2_AGAT_cyrA tổng hợp GAA với lượng từ 1 đến tối đa là 26mg/l trong PM không được bổ sung. Các tác giả sáng chế kết luận rằng tiền chất L-arginin và glyxin được tổng hợp từ cơ chất D-glucosa, amoni và urê ban đầu.

Khi bổ sung glyxin, sự tổng hợp GAA của ATCC21831/pEKEEx2_AGAT_Mp, ATCC21831/pEKEEx2_AGAT_Hs, ATCC21831/pEKEEx2_AGAT_Rn, ATCC21831/pEKEEx2_AGAT_Gv, và ATCC21831/pEKEEx2_AGAT_cyrA đã tăng lên đáng kể so với các thử nghiệm không được bổ sung. So với chủng ATCC13032/pEC-XK99E_AGAT_Mp (xem bảng 10), chủng tổng hợp L-arginin ATCC21831 mang gen AGAT tích lũy nhiều GAA hơn khi glyxin không bị giới hạn. Các tác giả sáng chế kết luận rằng khả năng cung cấp L-arginin bên trong cải thiện sự tổng hợp GAA.

Ví dụ 14: Tác động của quá trình tái tạo L-arginin đối với sự tổng hợp GAA

Ở các chủng tổng hợp L-arginin (ví dụ, ATCC21831), L-ornitin trung gian được tổng hợp *de novo* và tiếp tục được biến đổi thành L-arginin. Khi chủng này được cung cấp AGAT, enzym này sẽ tổng hợp GAA và L-ornitin với lượng phân tử bằng nhau. Mặc dù sự tạo thành L-ornitin tiêu thụ nguồn C và N ban đầu với lượng lớn và do đó làm giảm sản lượng GAA.

Các tác giả sáng chế phát hiện ra rằng việc tăng cường con đường sinh tổng

hợp L- ornitin thành L-arginin sẽ cải thiện sự tổng hợp GAA, có lẽ bằng cách cải thiện quá trình tái chế L-ornitin thành L-arginin.

Các chủng khác nhau có nguồn gốc từ ATCC13032, được nuôi cấy bằng cách sử dụng hệ thống Wouter Duetz và sau đó, được phân tích khả năng tổng hợp GAA (bảng 13, 14 và 15).

Bảng 13: Tác động của việc tái tạo L-arginin cải thiện đối với việc tổng hợp GAA trong PM chứa YE mà không bổ sung glyxin hoặc L-arginin

Chủng	GAA
ATCC13032/pEC-XK99E	0mg/l
ATCC13032/pEC-XK99E_argFGH	0mg/l
ATCC13032_Psod-carAB/pEC-XK99E	0mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	25mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	26mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argH	24mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argF	27mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	26mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	25mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	43mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argG	45mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH	45mg/l

Bảng 14: Tác động của việc tái tạo L-arginin cải thiện đối với việc tổng hợp GAA trong PM không có YE. Tất cả các mẫu cấy đều được bổ sung glyxin với lượng 0,82g/l

Chủng	GAA
ATCC13032/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_argFGH	0 mg/l
ATCC13032_Psod-carAB/pEC-XK99E	0 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp	22 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	24 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argH	22 mg/l

ATCC13032/pEC-XK99E_AGAT_Mp_argF	23 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	23 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	23 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	53 mg/l

Bảng 15: Tác động của việc tái tạo L-arginin cải thiện đối với việc tổng hợp GAA trong PM chứa YE. Tất cả các mẫu cấy đều được bổ sung glyxin với lượng 0,82g/l và L-arginin với lượng 1,9g/l.

Chủng	GAA
ATCC13032/pEC-XK99E	không phát hiện
ATCC13032/pEC-XK99E_argFGH	không phát hiện
ATCC13032_Psod-carAB/pEC-XK99E	không phát hiện
ATCC13032/pEC-XK99E_AGAT_Mp	124 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argG	136 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argF	136 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argGH	133 mg/l
ATCC13032/pEC-XK99E_AGAT_Mp_argFGH	154 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp	156 mg/l
ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH	171 mg/l

Như được thể hiện trong các bảng từ 13 đến 15, các chủng không mang gen AGAT không tổng hợp GAA với lượng có thể phát hiện được.

Biểu hiện của AGAT_Mp trong ATCC13032/pEC-XK99E_AGAT_Mp tạo ra GAA với lượng 124mg/l. Khuếch đại bổ sung *argG* (chủng ATCC13032/pEC-XK99E_AGAT_Mp_argG), *argF* (chủng ATCC13032/pEC-XK99E_AGAT_Mp_argF) hoặc *argG* + *argH* (chủng ATCC13032/pEC-XK99E_AGAT_Mp_argGH) đã cải thiện tổng hợp GAA (xem bảng 14).

Ở chủng ATCC13032/pEC-XK99E_AGAT_Mp_argFGH, sự biểu hiện của các gen *argF* (mã hóa cho ornitin carbamoyltransferaza), *argG* (mã hóa cho argininosuccinat synthetaza) và *argH* (mã hóa cho argininosuccinat lyaza) được tăng cường. Việc biểu hiện này tiếp tục cải thiện việc tổng hợp GAA với lượng tối đa là 154mg/l (xem bảng 14).

Sự biến đổi L-ornitin thành L-xitrulin được xúc tác bởi ornitin carbamoyltransferaza, phụ thuộc vào sự sẵn có của đồng cơ chất carbamoyl

phosphat. Carbamoyl phosphat được tổng hợp bởi carbamoyl phosphat synthaza, được mã hóa bởi các gen *carA* và *carB*. Ở chủng ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp, việc chèn vùng khởi động mạnh sod vào hệ gen sẽ tăng cường sự biểu hiện của *carA* và *carB*. So với ATCC13032/pEC-XK99E_AGAT_Mp, việc này làm cải thiện tổng hợp GAA (156mg/l so với 124mg/l).

Ở chủng ATCC13032_Psod-carAB/pEC-XK99E_AGAT_Mp_argFGH, sự biến đổi L-ornitin được cải thiện (biểu hiện quá mức *argF*, *argG* và *argH*) đã được kết hợp với sinh tổng hợp carbamoyl phosphat synthaza được cải thiện (biểu hiện quá mức của *carA* và *carB*). Sự kết hợp này tiếp tục cải thiện tổng hợp GAA với lượng 171mg/l.

Ví dụ 15: Thiết kế vector biểu hiện *P. putida* cho gen AGAT_Mp của *Moorea producens*

Để biểu hiện dị hình AGAT từ *Moorea producens* (EC 2.1.4.1, SEQ ID NO:2 và SEQ ID NO:4) ở *P. putida* KT2440, đã thiết kế plasmit pACYCATH-5{PRha}[agat_Mp (coPp)]. Gen AGAT_Mp đã tối ưu hóa mã bộ ba được tách dòng dưới sự kiểm soát của vùng khởi động cảm ứng ramnoza P_{rha} vào vector pACYATH-5. Trình tự kết thúc nằm ở vị trí xuôi dòng của gen AGAT_Mp. Gen AGAT_Mp được sắp xếp để tổng hợp gen từ Eurofins Genomics Germany GmbH (Ebersberg, Đức) và trình tự ADN của đoạn gen đã tối ưu hóa mã bộ ba để biểu hiện ở *P. putida* KT2440 (SEQ ID NO:33). Đã bổ sung trình tự Shine-Dalgarno ở vị trí ngược dòng của khung đọc mở. Catxet vùng khởi động P_{Rha} (SEQ ID NO:34) và trình tự kết thúc (SEQ ID NO:35) được khuếch đại từ ADN hệ gen của *E. coli* K12. Vector này được dựa trên pACYC184 (New England BioLabs Inc., Ipswich, Mỹ) và mang điểm khởi đầu sao chép p15A đối với *E. coli* và điểm khởi đầu sao chép pVS1 để sao chép ở *P. putida* KT2440. Điểm khởi đầu sao chép pVS1 bắt nguồn từ pVS1 của plasmit *Pseudomonas* (Itoh Y, Watson JM, Haas D, Leisinger T, Plasmid 1984, 11(3), 206-20). Trong bước tiếp theo, đoạn gen AGAT_Mp được khuếch đại bằng phương pháp PCR, sử dụng đoạn môi MW_20_01_xuôi (SEQ ID NO:36) và MW_20_02_ngược (SEQ ID NO:37) và

được tách dòng vào vector pACYCATH-5 bằng cách sử dụng các vị trí giới hạn *Apal*/*XhoI* và kit tách dòng liên hợp ADN NEBuilder® HiFi từ (New England BioLabs Inc., Ipswich, Mỹ, mã số E5520). Sản phẩm nối được biến nạp vào tế bào *E. coli* khả biến 10-beta (New England BioLabs Inc., Ipswich, Mỹ, mã số C3020K). Quy trình tinh chế sản phẩm PCR, tách dòng và biến nạp được thực hiện theo hướng dẫn của nhà sản xuất. Việc chèn đúng gen đích đã được kiểm tra bằng phân tích giới hạn và độ chính xác của các đoạn ADN đã đưa vào được xác nhận bằng phương pháp xác định trình tự ADN. Vector biểu hiện thu được được gọi là pACYCATH-5{PRha}[*agat_Mp* (coPp)] (SEQ ID NO:38, xem bảng 17).

Chủng *P. putida* KT2440 đã được biến nạp plasmit pACYCATH-5{PRha}[*agat_Mp* (coPp)] bằng các phương pháp điện di và được cấy trải lên đĩa thạch LB đã bổ sung tetracyclin (10mg/l). Các chủng biến nạp được kiểm tra sự có mặt của plasmit chính xác bằng cách chuẩn bị plasmit và phân tích giới hạn phân tích. Chủng thu được được gọi là *P. putida* KT2440/pACYCATH-5{PRha}[*agat_Mp* (coPp)] (xem bảng 18).

Ví dụ 16: Thiết kế vector biểu hiện *P. putida* cho gen *AGAT_Mp* của *Moorea producens* và các gen *argF*, *argG* và *argH* của *P. putida*

Để biểu hiện dị hợp của *AGAT_Mp* từ *Moorea producens* và *argF* (SEQ ID NO:39), *argG* (SEQ ID NO:41), *argH* (SEQ ID NO:43) từ *P. putida* KT2440, đã thiết kế plasmit pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*]. Operon tổng hợp này bao gồm *AGAT_Mp* mã hóa cho L-arginin:glyxin amidinotransferaza (*AGAT*, EC 2.1.4.1, SEQ ID NO:2 và SEQ ID NO:4), *argF* mã hóa cho ornitin carbamoyltransferaza (*ArgF*, EC 2.1.3.3, SEQ ID NO:40), *argG* mã hóa cho argininosuccinat synthaza (*ArgG*, EC 6.3.4.5, SEQ ID NO:42) và *argH* mã hóa cho argininosuccinat lyaza (*ArgH*, EC 4.3.2.1, SEQ ID NO:44), được tách dòng theo sự kiểm soát của vùng khởi động cảm ứng ramnoza *P_{rha}* vào vector pACYCATH-5. Ở vị trí xuôi dòng của operon tổng hợp có trình tự kết thúc. Gen *AGAT_Mp* được sắp xếp để tổng hợp gen từ Eurofins Genomics Germany GmbH (Ebersberg, Đức) và trình tự ADN của đoạn gen đã tối ưu hóa mã bộ ba để

biểu hiện ở *P. putida* KT2440. Các gen *argFGH* cũng được tổng hợp dưới dạng đoạn gen *argFGH* (SEQ ID NO:45). Catxet vùng khởi động P_{Rha} (SEQ ID NO:34) và trình tự kết thúc (SEQ ID NO:35) được khuếch đại từ ADN hệ gen của *E. coli* K12. Vector này dựa trên pACYC184 (New England BioLabs Inc., Ipswich, Mỹ) và mang điểm khởi đầu sao chép p15A đối với *E. coli* và điểm khởi đầu sao chép pVS1 để sao chép ở *P. putida* KT2440. Điểm khởi đầu sao chép pVS1 bắt nguồn từ plasmit pVS1 của *Pseudomonas* (Itoh Y, Watson JM, Haas D, Leisinger T, Plasmid 1984, 11(3), 206-20). Để tách dòng, AGAT_Mp và *argFGH* được khuếch đại bằng phương pháp PCR. Môi được sử dụng để tách dòng được liệt kê trong bảng 16. Các sản phẩm PCR được tách dòng vào vector pACYCATH-5 bằng cách sử dụng các vị trí giới hạn *ApaI/XhoI* và kit tách dòng liên hợp ADN NEBuilder® HiFi từ (New England BioLabs Inc., Ipswich, Mỹ, mã số E5520) để tổng hợp operon đã tối ưu hóa. Để khuếch đại, hỗn hợp phản ứng có độ tin cậy cao Phusion™ từ New England Biolabs (Ipswich, Mỹ) được sử dụng theo hướng dẫn của nhà sản xuất. Sản phẩm nối được biến nạp vào tế bào *E. coli* khả biến 10-beta (New England BioLabs Inc., Ipswich, Mỹ, mã số C3020K). Quy trình tinh chế sản phẩm PCR, tách dòng và biến nạp được thực hiện theo hướng dẫn của nhà sản xuất. Việc chèn đúng các gen đích đã được kiểm tra bằng phân tích giới hạn và độ chính xác của các đoạn ADN đã đưa vào được xác nhận bằng phương pháp xác định trình tự ADN. Vector biểu hiện thu được được gọi là pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*] (SEQ ID NO:49, xem bảng 17).

Chủng *P. putida* KT2440 được biến nạp plasmit pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*] bằng các phương pháp điện di và được cấy trải lên đĩa thạch LB đã bổ sung tetracyclin (10mg/l). Các chủng biến nạp được kiểm tra sự có mặt của plasmit chính xác bằng cách chuẩn bị plasmit và phân tích giới hạn phân tích. Chủng thu được được gọi là *P. putida* KT2440/pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*] (xem bảng 18).

Bảng 16: Các đoạn mồi được sử dụng để tách dòng AGAT_Mp và *argFGH* vào pACYCATH-5.

Mồi	SEQ ID	ADN khuôn	Sản phẩm PCR
MW_20_01_xuôi	SEQ ID NO:36	AGAT_đoạn gen Mp	AGAT_Mp
MW_20_03_ngược	SEQ ID NO:46	AGAT_đoạn gen Mp	AGAT_Mp
MW_20_04_xuôi	SEQ ID NO:47	đoạn gen <i>argFGH</i>	<i>argFGH</i>
MW_20_09_ngược	SEQ ID NO:48	đoạn gen <i>argFGH</i>	<i>argFGH</i>

Ví dụ 17: Tách dòng gen *carAB* từ *P. putida* KT2440 vào vector biểu hiện pACYCATH-5PRha}[*agat_Mp* (coPp)*argFGH_Pp*]

Để biểu hiện dị hợp của AGAT_Mp từ *Moorea producens*, đã thiết kế *argF* (SEQ ID NO:39), *argG* (SEQ ID NO:41), *argH* (SEQ ID NO:43), *carA* (SEQ ID NO:50) và *carB* (SEQ ID NO:52) từ plasmit pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*]{*carAB_Pp*}[*carAB_Pp*][*ter*] của *P. putida* KT2440. Gen *carA* (SEQ ID NO:50) và *carB* (SEQ ID NO:52) mã hóa cho carbamoyl-phosphat synthaza (CarAB, EC 6.3.5.5, SEQ ID NO:51 và SEQ ID NO:53) được khuếch đại từ ADN hệ gen của *P. putida* KT2440 bao gồm cả vùng khởi động gốc của operon *carAB* bằng phương pháp PCR, sử dụng mồi MW_20_35 xuôi (SEQ ID NO:54) và MW_20_36 ngược (SEQ ID NO:55). Để khuếch đại, hỗn hợp phản ứng có độ tin cậy cao Phusion™ từ New England Biolabs (Ipswich, Mỹ) được sử dụng theo hướng dẫn của nhà sản xuất. Sản phẩm PCR (SEQ ID NO:56) đã được tách dòng vào pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*], được cắt bằng *Bsu36I* bằng cách sử dụng kit tách dòng liên hợp ADN NEBuilder® HiFi từ (New England BioLabs Inc., Ipswich, Mỹ, mã số E5520). Sản phẩm nối được biến nạp vào tế bào *E. coli* khả biến 10-beta (New England BioLabs Inc., Ipswich, Mỹ, mã số C3020K). Quy trình tinh chế sản phẩm PCR, tách dòng và biến nạp được thực hiện theo hướng dẫn của nhà sản xuất. Việc chèn đúng các gen đích đã được kiểm tra bằng phân tích giới hạn và độ chính xác của các đoạn ADN đã đưa vào được xác nhận bằng phương pháp xác định trình tự ADN. Vector biểu hiện thu được được gọi là pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*]{*carAB_Pp*}[*carAB_Pp*][*ter*] (SEQ ID NO:57, xem bảng 17).

Chủng *P. putida* KT2440 được biến nạp plasmit pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] bằng các phương pháp điện di và cấy trải lên đĩa thạch LB đã bổ sung tetracyclin (10mg/l). Các chủng biến nạp được kiểm tra sự có mặt của plasmit chính xác bằng cách chuẩn bị plasmit và phân tích giới hạn phân tích. Chủng thu được được gọi là *P. Putida* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] (xem bảng 18).

Bảng 17: Danh mục các plasmit biểu hiện *P. putida*

Plasmit	Chú thích
pACYCATH-5	Plasmit con thoi trống <i>E. coli</i> – <i>P. putida</i>
pACYCATH-5{PRha}[agat_Mp(coPp)]	Biểu hiện AGAT_Mp (<i>Moorea producens</i>)
pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> (<i>P. putida</i>)
pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]	Biểu hiện AGAT_Mp (<i>Moorea producens</i>), <i>argF</i> , <i>argG</i> , <i>argH</i> , <i>carA</i> , <i>carB</i> (<i>P. putida</i>)

Bảng 18: Danh mục các chủng *P. putida* chứa plasmit

Plasmit	Gen đã tách dòng	Chủng nhận	Chủng thu được
pACYCATH-5	không chứa	<i>P. putida</i> KT2440	<i>P. putida</i> KT2440/pACYCATH-5
pACYCATH-5{PRha}[agat_Mp(coPp)]	AGAT_Mp	<i>P. putida</i> KT2440	<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)]
pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i>	<i>P. putida</i> KT2440	<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]
pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]	AGAT_Mp, <i>argF</i> , <i>argG</i> , <i>argH</i> , <i>carA</i> , <i>carB</i>	<i>P. putida</i> KT2440	<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]

Ví dụ 18. Ảnh hưởng của AGAT đến tổng hợp GAA ở *P. putida* KT2440

Chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)] mang gen mã hóa cho enzym AGAT từ KT2440/pACYCATH-5 của *M. producens* và *P.*

putida (vector rỗng đối chứng) đã được phân tích khả năng tổng hợp GAA khi nuôi cấy theo mẻ, sử dụng bình lắc. Trên đĩa thạch LB chứa tetracyclin với lượng 10mg/l, tạo vệt bằng que cấy vòng mang các chủng tương ứng đã được giữ lạnh trong glyxerol. Các đĩa thạch được ủ trong 24 giờ ở 30°C. Cấy một khuẩn lạc từ đĩa thạch vào bình có vách ngăn dung tích 100ml chứa 15ml môi trường nuôi cấy (đã hấp khử trùng: 4,4g/l $\text{Na}_2\text{HPO}_4 \times 2 \text{H}_2\text{O}$, 1,5g/l KH_2PO_4 , 1g/l NH_4Cl , 10g/l chiết xuất nấm men, khử trùng riêng: 20g/l glucoza, 0,2g/l $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$, 0,006g/l FeCl_3 , 0,015g/l CaCl_2 , dung dịch nguyên tố vi lượng SL6 có hàm lượng 1ml/l (đã lọc vô trùng: 0,3g/l H_3BO_3 , 0,2g/l $\text{CoCl}_2 \times 6 \text{H}_2\text{O}$, 0,1g/l $\text{ZnSO}_4 \times 7 \text{H}_2\text{O}$, 0,03g/l $\text{MnCl}_2 \times 4\text{H}_2\text{O}$, 0,01g/l $\text{CuCl}_2 \times 2 \text{H}_2\text{O}$, 0,03g/l $\text{Na}_2\text{MoO}_4 \times 2 \text{H}_2\text{O}$, 0,02g/l $\text{NiCl}_2 \times 6 \text{H}_2\text{O}$), độ pH=7) chứa tetracyclin với lượng 10mg/l và ủ trong tủ ấm lắc trong 18 giờ ở 30°C và 200 vòng/phút để tổng hợp tiền nuôi cấy. Tiền nuôi cấy này được sử dụng để cấy vào môi trường M12 với lượng 40ml (thành phần: (2,2g/l $(\text{NH}_4)_2\text{SO}_4$, 0,02g/l NaCl , 0,4g/l $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 0,04g/l $\text{CaCl}_2 \times 2\text{H}_2\text{O}$, được khử trùng riêng biệt: 2g/l KH_2PO_4 , 8,51g/l Na_2HPO_4 , 20g/l glucoza, 10ml/l dung dịch nguyên tố vi lượng M12 (đã lọc vô trùng: 0,2g/l $\text{ZnSO}_4 \times 7 \text{H}_2\text{O}$, 0,1g/l $\text{MnCl}_2 \times 4\text{H}_2\text{O}$, 1,5g/l $\text{Na}_3\text{-xitat} \times 2 \text{H}_2\text{O}$, 0,1g/l $\text{CuSO}_4 \times 5 \text{H}_2\text{O}$, 0,002g/l $\text{NiCl}_2 \times 6 \text{H}_2\text{O}$, 0,003g/l $\text{Na}_2\text{MoO}_4 \times 2\text{H}_2\text{O}$, 0,03g/l H_3BO_3 , 1g/l $\text{FeSO}_4 \times 7 \text{H}_2\text{O}$), độ pH=7,4) chứa tetracyclin với lượng 10mg/l, 3,48g/l arginin và 1,5g/l glyxin cho đến khi OD_{600} ban đầu bằng 0,1. Các chủng này được nuôi cấy trong 48 giờ. Tại OD_{600} nằm trong khoảng 0,5-0,8, sự biểu hiện gen được cảm ứng bằng cách bổ sung ramnoza với lượng 0,2% (khối lượng/thể tích). Sau khi cảm ứng 9 giờ và 24 giờ, lượng arginin và glyxin đã tăng đột biến với lượng lần lượt là 1,74g/l và 0,75g/l. Tại thời điểm cuối của quá trình nuôi cấy, các mẫu được lấy để xác định lượng GAA được tổng hợp.

Các kết quả được thể hiện trong bảng 19.

Bảng 19: Sự tổng hợp GAA bởi các chủng *P. putida* KT2440/pACYCATH-5 và *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)]

Chủng	GAA
<i>P. putida</i> KT2440/pACYCATH-5	không phát hiện được
<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)]	81,5 mg/l

Như có thể thấy trong bảng 19, chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)] mang gen AGAT_Mp từ *M. producers* có thể tổng hợp GAA với lượng khoảng 81,5mg/l. Chủng đối chứng *P. putida* KT2440/pACYCATH-5 hoàn toàn không tổng hợp GAA bất kỳ.

Ví dụ 19: Tác động của AGAT và sự cung cấp L-arginin gia tăng đối với sự tổng hợp GAA ở *P. putida* KT2440

Chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)] mang gen mã hóa cho enzym AGAT từ *M. producers* và chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)argFGH_Pp] còn mang các gen sinh tổng hợp arginin *argFGH* đã được phân tích khả năng tổng hợp GAA của chúng trong quá trình nuôi cấy theo mẻ bằng cách sử dụng bình lắc. Trên đĩa thạch LB chứa tetracyclin với lượng 10mg/l, đã vạch que cấy vòng mang các chủng tương ứng đã được giữ trong glyxerol. Đĩa thạch được ủ trong 24 giờ ở 30°C. Cấy một khuẩn lạc từ đĩa thạch vào bình có vách ngăn dung tích 100ml chứa 15ml môi trường nuôi cấy (đã hấp khử trùng: 4,4g/l Na₂HPO₄ x 2 H₂O, 1,5g/l KH₂PO₄, 1g/l NH₄Cl, 10g/l chiết xuất nấm men, khử trùng riêng: 20g/l glucoza, 0,2g/l MgSO₄ x 7 H₂O, 0,006g/l FeCl₃, 0,015g/l CaCl₂, dung dịch nguyên tố vi lượng SL6 có hàm lượng 1ml/l (đã lọc vô trùng: 0,3g/l H₃BO₃, 0,2g/l CoCl₂ x 6 H₂O, 0,1g/l ZnSO₄ x 7 H₂O, 0,03g/l MnCl₂ x 4H₂O, 0,01g/l CuCl₂ x 2 H₂O, 0,03g/l Na₂MoO₄ x 2 H₂O, 0,02g/l NiCl₂ x 6 H₂O), độ pH=7) chứa tetracyclin với lượng 10mg/l và ủ trong tủ ẩm lắc trong 18 giờ ở 30°C và 200 vòng/phút để tổng hợp tiền nuôi cấy. Tiền nuôi cấy được sử dụng để cấy vào môi trường M12 với lượng 40ml (thành phần: (2,2g/l (NH₄)₂SO₄, 0,02g/l NaCl, 0,4g/l MgSO₄ x 7H₂O, 0,04g/l CaCl₂ x 2H₂O, được khử trùng riêng biệt: 2g/l

KH₂PO₄, 8,51g/l Na₂HPO₄, 20g/l glucoza, dung dịch nguyên tố vi lượng M12 với lượng 10ml/l (đã lọc vô trùng: 0,2g/l ZnSO₄ x 7 H₂O, 0,1g/l MnCl₂ x 4H₂O, 1,5g/l Na₃-xítat x 2 H₂O, 0,1g/l CuSO₄ x 5 H₂O, 0,002g/l NiCl₂ x 6 H₂O, 0,003g/l Na₂MoO₄ x 2 H₂O, 0,03g/l H₃BO₃, 1g/l FeSO₄ x 7 H₂O), độ pH=7,4) chứa tetracyclin với lượng 10mg/l, 3,48g/l arginin và 1,5g/l glyxin cho đến khi OD₆₀₀ ban đầu bằng 0,1. Các chủng này được nuôi cấy trong 48 giờ. Ở OD₆₀₀ nằm trong khoảng 0,5-0,8, sự biểu hiện gen được cảm ứng bằng cách bổ sung ramnoza với lượng 0,2% (khối lượng/thể tích). Sau khi cảm ứng 9 giờ và 24 giờ, lượng arginin và glyxin đã tăng đột biến với lượng lần lượt là 1,74g/l và 0,75g/l. Tại thời điểm cuối của quá trình nuôi cấy, các mẫu được lấy để xác định lượng GAA được tổng hợp.

Các kết quả được thể hiện trong bảng 20.

Bảng 20: Sự tổng hợp GAA bởi chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)] và *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)argFGH_Pp]

Chủng	GAA
<i>P. putida</i> KT2440/pACYCATH-5	không phát hiện
<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp)]	81,5 mg/l
<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]	169,5 mg/l

Như có thể thấy trong bảng 20, chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)] mang gen AGAT_Mp từ *M. producents* có thể tổng hợp GAA với lượng khoảng 81,5mg/l. Ngoài ra, việc bổ sung *argF*, *argG* và *argH* ở chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)argFGH_Pp] đã cải thiện tổng hợp GAA đến 169,5mg/l.

Ví dụ 20: Tác động của AGAT và sự tái tạo L-arginin đối với việc tổng hợp GAA ở *P. putida* KT2440

Chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)] mang gen mã hóa cho enzym AGAT từ *M. producents*, chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp argFGH_Pp] còn mang gen sinh tổng

hợp arginin *argFGH* và chủng *P. putida* KT2440/pACYCATH-5{PRha}[*agat_Mp* (coPp)*argFGH_Pp*]{*carAB_Pp*}[*carAB_Pp*][*ter*] còn mang gen sinh tổng hợp carbamoyl-phosphat *carAB* được phân tích khả năng tổng hợp GAA khi nuôi cấy theo mẻ bằng cách sử dụng bình lắc. Sự biến đổi L-ornitin thành L-xitruilin được xúc tác bởi ornitin carbamoyltransferaza phụ thuộc vào sự sẵn có của đồng cơ chất carbamoyl phosphat. Carbamoyl phosphat được tổng hợp bởi carbamoyl phosphat synthaza, được mã hóa bởi các gen *carA* và *carB*. Trên đĩa thạch LB chứa tetracyclin với lượng 10mg/l, đã vạch que cấy vòng mang các chủng tương ứng đã được giữ trong glyxerol. Đĩa thạch được ủ trong 24 giờ ở 30°C. Cấy một khuẩn lạc từ đĩa thạch vào bình có vách ngăn dung tích 100ml chứa 15ml môi trường nuôi cấy (đã hấp khử trùng: 4,4g/l Na₂HPO₄ x 2 H₂O, 1,5g/l KH₂PO₄, 1g/l NH₄Cl, 10g/l chiết xuất nấm men, khử trùng riêng: 20g/l glucoza, 0,2g/l MgSO₄ x 7 H₂O, 0,006g/l FeCl₃, 0,015g/l CaCl₂, dung dịch nguyên tố vi lượng SL6 có hàm lượng 1ml/l (đã lọc vô trùng: 0,3g/l H₃BO₃, 0,2g/l CoCl₂ x 6 H₂O, 0,1g/l ZnSO₄ x 7 H₂O, 0,03g/l MnCl₂ x 4H₂O, 0,01g/l CuCl₂ x 2 H₂O, 0,03g/l Na₂MoO₄ x 2 H₂O, 0,02g/l NiCl₂ x 6 H₂O), độ pH=7) chứa tetracyclin với lượng 10mg/l và ủ trong tủ ẩm lắc trong 18 giờ ở 30°C và 200 vòng/phút để tổng hợp tiền nuôi cấy. Tiền nuôi cấy được sử dụng để cấy vào môi trường M12 với lượng 40ml (thành phần: (2,2g/l (NH₄)₂SO₄, 0,02g/l NaCl, 0,4g/l MgSO₄ x 7H₂O, 0,04g/l CaCl₂ x 2H₂O, được khử trùng riêng biệt: 2g/l KH₂PO₄, 8,51g/l Na₂HPO₄, 20g/l glucoza, 10ml/l dung dịch nguyên tố vi lượng M12 (đã lọc vô trùng: 0,2g/l ZnSO₄ x 7 H₂O, 0,1g/l MnCl₂ x 4H₂O, 1,5g/l Na₃-xitrat x 2 H₂O, 0,1g/l CuSO₄ x 5 H₂O, 0,002g/l NiCl₂ x 6 H₂O, 0,003g/l Na₂MoO₄ x 2 H₂O, 0,03g/l H₃BO₃, 1g/l FeSO₄ x 7 H₂O), độ pH=7,4) chứa tetracyclin với lượng 10mg/l, 3,48g/l arginin và 1,5g/l glyxin đến OD₆₀₀ ban đầu bằng 0,1. Các chủng này được nuôi cấy trong 48 giờ. Ở OD₆₀₀ nằm trong khoảng 0,5-0,8, sự biểu hiện gen được cảm ứng bằng cách bổ sung ramnoza với lượng 0,2% (khối lượng/thể tích). Sau khi cảm ứng 4 giờ/18 giờ/23 giờ, 6,97g/l Arg/1,5g/l Gly, 2,34g/l Arg/0,75g/l Gly và 6,97g/l Arg/1,5g/l Gly đã được tăng đột biến. Tại thời điểm cuối của quá trình nuôi cấy, các mẫu được lấy để xác định lượng GAA được tổng hợp.

Các kết quả được thể hiện trong bảng 21.

Bảng 21: Sự tổng hợp GAA bởi chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)], *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp) argFGH_Pp] và *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]

Chủng	GAA
<i>P. putida</i> KT2440/pACYCATH-5	không phát hiện
<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp argFGH_Pp]	589 mg/l
<i>P. putida</i> KT2440/pACYCATH-5{PRha}[agat_Mp(coPp) argFGH_Pp]{carAB_Pp}[carAB_Pp][ter]	693 mg/l

Như có thể thấy trong bảng 21, chủng mang gen AGAT_Mp từ *M. producens* và gen *argFGH* từ *P. putida* có thể tổng hợp GAA với lượng 589mg/l. Ngoài ra, việc bổ sung *carAB* ở chủng *P. putida* KT2440/pACYCATH-5{PRha}[agat_Mp (coPp)argFGH_Pp]{carAB_Pp}[carAB_Pp][ter] đã cải thiện tổng hợp GAA lên 693mg/l.

DANH MỤC TRÌNH TỰ

<110> Evonik Operations GmbH

<120> Vi sinh vật có hoạt tính enzym có chức năng carbamoylphosphat synthaza và phương pháp lên men tổng hợp axit guanidino axetic

<130> 000

<160> 57

<170> PatentIn bản 3.5

<210> 1

<211> 1146

<212> ADN

<213> Moorea producens chủng PAL-8-15-08-1

<220>

<221> CDS

<222> (1)..(1143)

<223> Mã số đăng ký Ngân hàng gen WP_070390602

<400> 1

atg tcg gaa aaa att gtt aat tcc tgg aat gaa tgg gat gaa ttg gaa	48
Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu	
1 5 10 15	

gaa atg gtg gta gga att gca gac tat gct agc ttt gaa cca aaa gaa	96
Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu	
20 25 30	

cca ggg aat cat ccg aaa tta aga aat caa aat tta gcg gaa atc att	144
Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile	
35 40 45	

cct ttc ccc agt gga cct aaa gac cct aaa gtc ctt gaa aaa gct aat	192
Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn	
50 55 60	

gaa gaa tta aat gga ctg gct tat tta tta aaa gac cac gat gtg ata	240
Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile	
65 70 75 80	

gta aga aga ccc gaa aaa att gat ttt act aaa tct cta aaa aca cct	288
Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro	
85 90 95	

tac ttt gaa gtt gca aat caa tac tgt gga gtc tgt cct cgg gat gtc	336
Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val	
100 105 110	

atg att acc ttt ggg aat gaa atc atg gaa gcg act atg tcg aag aga	384
Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg	
115 120 125	

gct aga ttt ttt gaa tac tta cct tac cgg aaa ttg gtc tat gaa tat	432
Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr	
130 135 140	

tggt aat aaa gac gag cat atg att tgg aat gct gcg cct aaa ccg act	480
Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr	
145 150 155 160	

atg cag gat agt atg tat cta gag aat ttc tgg gag ctg tct tta gaa	528
Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu	
165 170 175	

gaa cga ttt aag cgt atg cat gat ttt gaa ttt tgt att aca caa gat	576
Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp	
180 185 190	

gaa gta att ttt gat gcg gct gac tgt agc aga tta gga aag gat ata	624
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile	
195 200 205	

```

180              185              190
gaa gta att ttt gat gcg gct gac tgt agc aga tta gga aag gat ata      624
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile
      195              200              205

tta gtt cag gaa tcg atg aca aca aat aga aca gga att cgg tgg tta      672
Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu
      210              215              220

aaa aag cac cta gaa cca aga gga ttt cgg gtt cac cct gtt cat ttt      720
Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe
      225              230              235              240

ccc ctt gat ttt ttc ccc tca cac att gac tgt acg ttt gtt cct ttg      768
Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
      245              250              255

cga cca ggt ctt att ttg aca aac cct gaa aga cct ata cgg gaa gag      816
Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu
      260              265              270

gag gag aag att ttt aaa gag aat ggc tgg gag ttg atc aca gtt cct      864
Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro
      275              280              285

caa ccg act tgc tcg aat gat gaa atg cca atg ttt tgc cag tcc agt      912
Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser
      290              295              300

aag tgg ttg tca atg aat gtt ctg agt ata tca ccg aca aag gtt atc      960
Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile
      305              310              315              320

tgt gag gaa aga gaa aaa cct ctc caa gaa ttg ttg gat aag cat gga      1008
Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly
      325              330              335

ttt gag gtt ttt cct tta ccc ttt aga cat gtc ttt gaa ttt ggg ggg      1056
Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly
      340              345              350

tct ttt cat tgt gca act tgg gat att cgc cga aaa ggt gag tgt gaa      1104
Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu
      355              360              365

gat tat tta cca aat tta aac tat caa ccg att tgt ggt taa      1146
Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly
      370              375              380

```

```

<210>  2
<211> 381
<212>  PRT
<213>  Moorea producens chủng PAL-8-15-08-1

```

```

<400>  2

```

```

Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu
1              5              10              15

```

```

Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu
20              25              30

```

```

Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile
35              40              45

```

```

Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn
50              55              60

```

Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile
 65 70 75 80

Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro
 85 90 95

Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val
 100 105 110

Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg
 115 120 125

Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160

Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu
 165 170 175

Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp
 180 185 190

Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile
 195 200 205

Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu
 210 215 220

Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe
 225 230 235 240

Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu
 260 265 270

Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro
 275 280 285

Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser
 290 295 300

Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile
 305 310 315 320

Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly
 325 330 335

Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly
 340 345 350

Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu
 355 360 365

Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly
 370 375 380

<210> 3
 <211> 1248
 <212> ADN
 <213> ADN tổng hợp

<220>
 <221> RBS
 <222> (49)..(52)
 <223> Trình tự Shine-Dalgarno AGGA

<220>
 <221> CDS
 <222> (58)..(1200)
 <223> khung đọc mở của AGAT_Mp đã tối ưu hóa; protein tương đồng với
 mã số đăng ký Ngân hàng gen WP_070390602

<400> 3
 cgtctctgtg gataactgag cggataagtt cctagtagc gtgcgagcag gaagaac 57
 atg agc gag aaa att gtg aac agc tgg aat gaa tgg gat gaa ctg gaa 105
 Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu
 1 5 10 15
 gaa atg gtt gtt ggt att gca gat tat gca agc ttt gaa ccg aaa gaa 153
 Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu
 20 25 30
 ccg ggt aat cat ccg aaa ctg cgt aat cag aat ctg gca gaa att att 201
 Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile
 35 40 45
 ccg ttt ccg agc ggt ccg aaa gat ccg aaa gtt ctg gaa aaa gca aat 249
 Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn
 50 55 60
 gaa gaa ctg aat ggt ctg gcc tat ctg ctg aaa gat cat gat gtt att 297
 Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile
 65 70 75 80
 gtt cgc cgt ccg gaa aaa atc gac ttt acc aaa agc ctg aaa acc ccg 345
 Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro
 85 90 95
 tat ttc gaa gtt gcc aat cag tat tgt ggt gtt tgt ccg cgt gat gtt 393
 Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val
 100 105 110
 atg att acc ttt ggc aac gaa att atg gaa gcc acc atg agc aaa cgt 441
 Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg
 115 120 125
 gcc cgt ttt ttt gaa tat ctg ccg tat cgt aaa ctg gtg tat gag tat 489
 Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140
 tgg aac aaa gat gag cat atg atc tgg aat gca gca ccg aaa ccg acc 537
 Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160
 atg cag gat agc atg tat ctg gaa aac ttt tgg gaa ctg agc ctg gaa 585
 Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu

165	170	175	
gaa cgt ttt aaa cgt atg cac gat ttt gag ttt tgc atc acc cag gat			633
Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp			
180	185	190	
gaa gtg att ttt gat gca gca gat tgt agc cgt ctg ggt aaa gat att			681
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile			
195	200	205	
ctg gtt caa gaa agc atg acc acc aat cgt acc ggt att cgt tgg ctg			729
Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu			
210	215	220	
aaa aaa cat ctg gaa ccg cgt ggt ttt cgt gtt cat ccg gtt cat ttt			777
Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe			
225	230	235	240
ccg ctg gat ttt ttt ccg agc cat att gat tgt acc ttt gtt ccg ctg			825
Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu			
245	250	255	
cgt ccg ggt ctg att ctg acc aat ccg gaa cgt ccg att cgt gaa gaa			873
Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu			
260	265	270	
gaa gag aaa atc ttc aaa gag aat ggc tgg gag ctg att acc gtt ccg			921
Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro			
275	280	285	
cag ccg acc tgt agc aat gat gaa atg ccg atg ttt tgt cag agc agc			969
Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser			
290	295	300	
aaa tgg ctg agc atg aat gtt ctg agc att agc ccg acc aaa gtt att			1017
Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile			
305	310	315	320
tgt gaa gaa cgt gaa aaa ccg ctg caa gaa ctg ctg gat aaa cat ggt			1065
Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly			
325	330	335	
ttt gaa gtg ttt ccg ctg ccg ttt cgt cat gtt ttt gaa ttt ggt ggt			1113
Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly			
340	345	350	
agc ttt cat tgt gcc acc tgg gat att cgt cgt aaa ggt gaa tgt gaa			1161
Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu			
355	360	365	
gat tat ctg ccg aat ctg aat tat cag ccg att tgt ggt taataagacg			1210
Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly			
370	375	380	
tcgcgcgaggg ccgtgttgcc gggtttttcca gagagacg			1248
<210> 4			
<211> 381			
<212> PRT			
<213> ADN tổng hợp			
<400> 4			
Met Ser Glu Lys Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Glu			
1	5	10	15
Glu Met Val Val Gly Ile Ala Asp Tyr Ala Ser Phe Glu Pro Lys Glu			
20	25	30	

Pro Gly Asn His Pro Lys Leu Arg Asn Gln Asn Leu Ala Glu Ile Ile
 35 40 45

Pro Phe Pro Ser Gly Pro Lys Asp Pro Lys Val Leu Glu Lys Ala Asn
 50 55 60

Glu Glu Leu Asn Gly Leu Ala Tyr Leu Leu Lys Asp His Asp Val Ile
 65 70 75 80

Val Arg Arg Pro Glu Lys Ile Asp Phe Thr Lys Ser Leu Lys Thr Pro
 85 90 95

Tyr Phe Glu Val Ala Asn Gln Tyr Cys Gly Val Cys Pro Arg Asp Val
 100 105 110

Met Ile Thr Phe Gly Asn Glu Ile Met Glu Ala Thr Met Ser Lys Arg
 115 120 125

Ala Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

Trp Asn Lys Asp Glu His Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160

Met Gln Asp Ser Met Tyr Leu Glu Asn Phe Trp Glu Leu Ser Leu Glu
 165 170 175

Glu Arg Phe Lys Arg Met His Asp Phe Glu Phe Cys Ile Thr Gln Asp
 180 185 190

Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Leu Gly Lys Asp Ile
 195 200 205

Leu Val Gln Glu Ser Met Thr Thr Asn Arg Thr Gly Ile Arg Trp Leu
 210 215 220

Lys Lys His Leu Glu Pro Arg Gly Phe Arg Val His Pro Val His Phe
 225 230 235 240

Pro Leu Asp Phe Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

Arg Pro Gly Leu Ile Leu Thr Asn Pro Glu Arg Pro Ile Arg Glu Glu
 260 265 270

Glu Glu Lys Ile Phe Lys Glu Asn Gly Trp Glu Leu Ile Thr Val Pro
 275 280 285

Gln Pro Thr Cys Ser Asn Asp Glu Met Pro Met Phe Cys Gln Ser Ser
 290 295 300

Lys Trp Leu Ser Met Asn Val Leu Ser Ile Ser Pro Thr Lys Val Ile
 305 310 315 320

Cys Glu Glu Arg Glu Lys Pro Leu Gln Glu Leu Leu Asp Lys His Gly
 325 330 335

Phe Glu Val Phe Pro Leu Pro Phe Arg His Val Phe Glu Phe Gly Gly
 340 345 350

Ser Phe His Cys Ala Thr Trp Asp Ile Arg Arg Lys Gly Glu Cys Glu
 355 360 365

Asp Tyr Leu Pro Asn Leu Asn Tyr Gln Pro Ile Cys Gly
 370 375 380

<210> 5
 <211> 30
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 5
 ccgttaactg ccgagacaat cgcataaagg 30

<210> 6
 <211> 34
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 6
 ctctaggct aatattctta cctcggttg ttgg 34

<210> 7
 <211> 27
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 7
 ccgttaacac ggctggcaag gaactta 27

<210> 8
 <211> 41
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 8
 gccctaggtc aatattacag gccataaact aatgcttattc g 41

<210> 9
 <211> 46
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 9
 ggaaacagct atgacatgat tacgcgggta tcgcggaatc cgtatg 46

<210> 10
 <211> 25
 <212> ADN
 <213> PsodcarAB-LA-R

<400> 10
 ttaagcgttt tgtgcaactc cgtct 25

<210> 11
 <211> 50

<212> ADN
 <213> oligonucleotit tổng hợp

 <400> 11
 agacggagtt gcacaaaacg cttaaaccct acttagctgc caattattcc 50

 <210> 12
 <211> 50
 <212> ADN
 <213> oligonucleotit tổng hợp

 <400> 12
 ggtaggtggt ggtgtcttta ctcattggga aaaaatcctt tcgtaggttt 50

 <210> 13
 <211> 25
 <212> ADN
 <213> oligonucleotit tổng hợp

 <400> 13
 atgagtaaag acaccaccac ctacc 25

 <210> 14
 <211> 46
 <212> ADN
 <213> oligonucleotit tổng hợp

 <400> 14
 gttgtaaaac gacggccagt gccaccggtg atgtggttct tcactg 46

 <210> 15
 <211> 1176
 <212> ADN
 <213> *Cylindrospermopsis raciborskii* AWT205

 <220>
 <221> CDS
 <222> (1)..(1176)
 <223> mã số đăng ký Ngân hàng gen EU140798.1

 <400> 15
 atg caa aca aga att gta aat agc tgg aat gag tgg gat gaa cta aag 48
 Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
 1 5 10 15

 gag atg gtt gtc ggg att gca gat ggt gct tat ttt gaa cca act gag 96
 Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu
 20 25 30

 cca ggt aac cgc cct gct tta cgc gat aag aac att gcc aaa atg ttc 144
 Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe
 35 40 45

 tct ttt ccc agg ggt ccg aaa aag caa gag gta aca gag aaa gct aat 192
 Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn
 50 55 60

 gag gag ttg aat ggg ctg gta gcg ctt cta gaa tca cag ggc gta act 240
 Glu Glu Leu Asn Gly Leu Val Ala Leu Leu Glu Ser Gln Gly Val Thr
 65 70 75 80

 gta cgc cgc cca gag aaa cat aac ttt ggc ctg tct gtg aag aca cca 288
 Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro
 85 90 95

 ttc ttt gag gta gag aat caa tat tgt gcg gtc tgc cca cgt gat gtt 336
 Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val

100	105	110		
atg atc acc ttt ggg aac gaa att ctc gaa gca act atg tca cgg cgg			384	
Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg				
115	120	125		
tca cgc ttc ttt gag tat tta ccc tat cgc aaa cta gtc tat gaa tat			432	
Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr				
130	135	140		
tgg cat aaa gat cca gat atg atc tgg aat gct gcg cct aaa ccg act			480	
Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr				
145	150	155	160	
atg caa aat gcc atg tac cgc gaa gat ttc tgg gag tgt ccg atg gaa			528	
Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu				
	165	170	175	
gat cga ttt gag agt atg cat gat ttt gag ttc tgc gtc acc cag gat			576	
Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp				
	180	185	190	
gag gtg att ttt gac gca gca gac tgt agc cgc ttt ggc cgt gat att			624	
Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile				
	195	200	205	
ttt gtg cag gag tca atg acg act aat cgt gca ggg att cgc tgg ctc			672	
Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu				
	210	215	220	
aaa cgg cat tta gag ccg cgt cgc ttc cgc gtg cat gat att cac ttc			720	
Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe				
	225	230	235	240
cca cta gat att ttc cca tcc cac att gat tgt act ttt gtc ccc tta			768	
Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu				
	245	250	255	
gca cct ggg gtt gtg tta gtg aat cca gat cgc ccc atc aaa gag ggt			816	
Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly				
	260	265	270	
gaa gag aaa ctc ttc atg gat aac ggt tgg caa ttc atc gaa gca ccc			864	
Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro				
	275	280	285	
ctc ccc act tcc acc gac gat gag atg cct atg ttc tgc cag tcc agt			912	
Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser				
	290	295	300	
aag tgg ttg gcg atg aat gtg tta agc att tcc ccc aag aag gtc atc			960	
Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile				
	305	310	315	320
tgt gaa gag caa gag cat ccg ctt cat gag ttg cta gat aaa cac ggc			1008	
Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly				
	325	330	335	
ttt gag gtc tat cca att ccc ttt cgc aat gtc ttt gag ttt ggc ggt			1056	
Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly				
	340	345	350	
tcg ctc cat tgt gcc acc tgg gat atc cat cgc acg gga acc tgt gag			1104	
Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu				
	355	360	365	
gat tac ttc cct aaa cta aac tat acg ccg gta act gca tca acc aat			1152	
Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn				
	370	375	380	
ggc gtt tct cgc ttc atc att tag			1176	
Gly Val Ser Arg Phe Ile Ile				

385

390

<210> 16
 <211> 391
 <212> PRT
 <213> *Cylindrospermopsis raciborskii* AWT205

<400> 16

Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
 1 5 10 15

Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu
 20 25 30

Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe
 35 40 45

Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn
 50 55 60

Glu Glu Leu Asn Gly Leu Val Ala Leu Leu Glu Ser Gln Gly Val Thr
 65 70 75 80

Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro
 85 90 95

Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val
 100 105 110

Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg
 115 120 125

Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140

Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160

Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu
 165 170 175

Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp
 180 185 190

Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile
 195 200 205

Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu
 210 215 220

Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe
 225 230 235 240

Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly
260 265 270

Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro
275 280 285

Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser
290 295 300

Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile
305 310 315 320

Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly
325 330 335

Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly
340 345 350

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
355 360 365

Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
370 375 380

Gly Val Ser Arg Phe Ile Ile
385 390

<210> 17
<211> 1272
<212> ADN
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1272)
<223> mã hóa cho NP_001473.1

<220>
<221> transit_peptit
<222> (1)..(111)
<223> peptit tín hiệu định vị ty thể

<400> 17	
atg ctg cgg gtg cgg tgt ctg cgc ggc ggg agc cgc ggc gcc gag gcg	48
Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala	
1 5 10 15	
gtg cac tac atc gga tct cgg ctt gga cga acc ttg aca gga tgg gtg	96
Val His Tyr Ile Gly Ser Arg Leu Gly Arg Thr Leu Thr Gly Trp Val	
20 25 30	
cag cga act ttc cag agc acc cag gca gct acg gct tcc tcc cgg aac	144
Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Arg Asn	
35 40 45	
tcc tgt gca gct gac gac aaa gcc act gag cct ctg ccc aag gac tgc	192
Ser Cys Ala Ala Asp Asp Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys	
50 55 60	
cct gtc tct tct tac aac gaa tgg gac ccc tta gag gaa gtg ata gtg	240

62

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 355 360 365

cgt gtt atg gtg gat gcc aat gaa gtt cca att caa aag atg ttt gaa 1152
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 370 375 380

aag ctg ggt atc act acc att aaa gtt aac att cgt aat gcc aat tcc 1200
 Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 385 390 395 400

ctg gga gga ggc ttc cat tgc tgg acc tgc gat gtc cgg cgc cga ggc 1248
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 405 410 415

acc tta cag tcc tac ttg gac tga 1272
 Thr Leu Gln Ser Tyr Leu Asp
 420

<210> 18
 <211> 423
 <212> PRT
 <213> Homo sapiens

<400> 18

Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala
 1 5 10 15

Val His Tyr Ile Gly Ser Arg Leu Gly Arg Thr Leu Thr Gly Trp Val
 20 25 30

Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Arg Asn
 35 40 45

Ser Cys Ala Ala Asp Asp Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
 50 55 60

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 65 70 75 80

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
 85 90 95

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly
 100 105 110

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 115 120 125

Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro
 130 135 140

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160

Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 165 170 175

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu

180	185	190
Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala		
195	200	205
Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn		
210	215	220
Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala		
225	230	235
Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala		
	245	250
Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val		
	260	265
Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp		
	275	280
Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile		
	290	300
Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro		
305	310	315
Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr		
	325	330
Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp		
	340	345
Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys		
	355	360
Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu		
	370	375
Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser		
385	390	395
Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly		
	405	410
Thr Leu Gln Ser Tyr Leu Asp		
	420	
<210> 19		
<211> 7		
<212> PRT		
<213> Escherichia coli		
<400> 19		
Met Thr Asp Val Thr Ile Lys		
1	5	

<210> 20
 <211> 375
 <212> PRT
 <213> Homo sapiens

<400> 20

Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
 1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
 35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly
 50 55 60

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 65 70 75 80

Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro
 85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110

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

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn
 165 170 175

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255

Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro
 260 265 270

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285

Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp
 290 295 300

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335

Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365

Thr Leu Gln Ser Tyr Leu Asp
 370 375

<210> 21
 <211> 423
 <212> PRT
 <213> Rattus norvegicus
 <400> 21

Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala
 1 5 10 15

Val His Tyr Ile Gly Ser Arg Leu Gly Gly Ser Leu Thr Gly Trp Val
 20 25 30

Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Gln Asn
 35 40 45

Ser Cys Ala Ala Glu Asp Lys Ala Thr His Pro Leu Pro Lys Asp Cys
 50 55 60

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 65 70 75 80

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 85 90 95

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
 100 105 110

Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
 115 120 125

Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
 130 135 140
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160
 Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly
 165 170 175
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 180 185 190
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 195 200 205
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
 210 215 220
 Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 225 230 235 240
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 245 250 255
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 260 265 270
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 275 280 285
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 290 295 300
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 305 310 315 320
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 325 330 335
 Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
 340 345 350
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 355 360 365
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 370 375 380
 Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 385 390 395 400
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 405 410 415

Thr Leu Gln Ser Tyr Phe Asp
420

<210> 22
<211> 375
<212> PRT
<213> Rattus norvegicus

<400> 22

Met Thr Asp Val Thr Ile Lys Ala Thr His Pro Leu Pro Lys Asp Cys
1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
50 55 60

Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
65 70 75 80

Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
100 105 110

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

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
165 170 175

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
210 215 220

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
225 230 235 240

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile

245 250 255
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285
 Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335
 Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365
 Thr Leu Gln Ser Tyr Phe Asp
 370 375

 <210> 23
 <211> 423
 <212> PRT
 <213> Galeopterus variegatus

 <400> 23
 Met Leu Arg Val Arg Cys Leu Arg Gly Gly Ser Arg Gly Ala Glu Ala
 1 5 10 15
 Val His Tyr Ile Gly Ser Arg Leu Gly Gly Ser Leu Thr Gly Trp Val
 20 25 30
 Gln Arg Thr Phe Gln Ser Thr Gln Ala Ala Thr Ala Ser Ser Gln Asn
 35 40 45
 Ser Cys Ala Ala Glu Asp Lys Ala Thr His Pro Leu Pro Lys Asp Cys
 50 55 60
 Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 65 70 75 80
 Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 85 90 95
 Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
 100 105 110
 Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
 115 120 125

Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
 130 135 140

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 145 150 155 160

Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly
 165 170 175

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 180 185 190

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 195 200 205

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
 210 215 220

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 225 230 235 240

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 245 250 255

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 260 265 270

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 275 280 285

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 290 295 300

Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 305 310 315 320

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 325 330 335

Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
 340 345 350

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 355 360 365

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 370 375 380

Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 385 390 395 400

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 405 410 415

Thr Leu Gln Ser Tyr Phe Asp
420

<210> 24
<211> 375
<212> PRT
<213> Galeopterus variegatus

<400> 24

Met Thr Asp Val Thr Ile Lys Ala Thr Asp Pro Leu Pro Lys Asp Cys
1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys His Gly Gly
50 55 60

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
65 70 75 80

Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro
85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
100 105 110

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

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu
130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
165 170 175

Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
210 215 220

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
225 230 235 240

Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
245 250 255

Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
260 265 270

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
275 280 285

Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
290 295 300

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
305 310 315 320

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Gly
325 330 335

Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
340 345 350

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
355 360 365

Thr Leu Gln Ser Tyr Phe Asp
370 375

<210> 25
<211> 1278
<212> ADN
<213> *Cylindrospermopsis raciborskii* AWT205

<220>
<221> RBS
<222> (49)..(52)
<223> Trình tự Shine-Dalgarno

<220>
<221> CDS
<222> (58)..(1230)

<400> 25
cgtctctgtg gataactgag cggataagtt cctagtacgc gtgcgagcag gaagaac 57

atg cag acc cgt att gtt aat agc tgg aat gaa tgg gat gag ctg aaa 105
Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
1 5 10 15

gaa atg gtt gtt ggt att gca gat ggt gcc tat ttt gaa ccg acc gaa 153
Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu
20 25 30

ccg ggt aat cgt ccg gca ctg cgt gat aaa aac att gca aaa atg ttt 201
Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe
35 40 45

agc ttt ccg cgt ggt ccg aaa aaa caa gaa gtt acc gaa aaa gcc aac 249
Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn
50 55 60

gaa gaa ctg aat ggt ctg gtt gca ctg ctg gaa agc cag ggt gtt acc 297

73

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
 355 360 365
 gat tat ttt ccg aaa ctg aat tat aca ccg gtt acc gca agc acc aat 1209
 Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
 370 375 380
 ggt gtg agc cgt ttt atc att taataagacg tccgcgaggg ccgtgttgcc 1260
 Gly Val Ser Arg Phe Ile Ile
 385 390
 gggtttcttca gagagacg 1278

 <210> 26
 <211> 391
 <212> PRT
 <213> *Cylindrospermopsis raciborskii* AWT205

 <400> 26
 Met Gln Thr Arg Ile Val Asn Ser Trp Asn Glu Trp Asp Glu Leu Lys
 1 5 10 15
 Glu Met Val Val Gly Ile Ala Asp Gly Ala Tyr Phe Glu Pro Thr Glu
 20 25 30
 Pro Gly Asn Arg Pro Ala Leu Arg Asp Lys Asn Ile Ala Lys Met Phe
 35 40 45
 Ser Phe Pro Arg Gly Pro Lys Lys Gln Glu Val Thr Glu Lys Ala Asn
 50 55 60
 Glu Glu Leu Asn Gly Leu Val Ala Leu Leu Glu Ser Gln Gly Val Thr
 65 70 75 80
 Val Arg Arg Pro Glu Lys His Asn Phe Gly Leu Ser Val Lys Thr Pro
 85 90 95
 Phe Phe Glu Val Glu Asn Gln Tyr Cys Ala Val Cys Pro Arg Asp Val
 100 105 110
 Met Ile Thr Phe Gly Asn Glu Ile Leu Glu Ala Thr Met Ser Arg Arg
 115 120 125
 Ser Arg Phe Phe Glu Tyr Leu Pro Tyr Arg Lys Leu Val Tyr Glu Tyr
 130 135 140
 Trp His Lys Asp Pro Asp Met Ile Trp Asn Ala Ala Pro Lys Pro Thr
 145 150 155 160
 Met Gln Asn Ala Met Tyr Arg Glu Asp Phe Trp Glu Cys Pro Met Glu
 165 170 175
 Asp Arg Phe Glu Ser Met His Asp Phe Glu Phe Cys Val Thr Gln Asp
 180 185 190
 Glu Val Ile Phe Asp Ala Ala Asp Cys Ser Arg Phe Gly Arg Asp Ile
 195 200 205

Phe Val Gln Glu Ser Met Thr Thr Asn Arg Ala Gly Ile Arg Trp Leu
 210 215 220

Lys Arg His Leu Glu Pro Arg Arg Phe Arg Val His Asp Ile His Phe
 225 230 235 240

Pro Leu Asp Ile Phe Pro Ser His Ile Asp Cys Thr Phe Val Pro Leu
 245 250 255

Ala Pro Gly Val Val Leu Val Asn Pro Asp Arg Pro Ile Lys Glu Gly
 260 265 270

Glu Glu Lys Leu Phe Met Asp Asn Gly Trp Gln Phe Ile Glu Ala Pro
 275 280 285

Leu Pro Thr Ser Thr Asp Asp Glu Met Pro Met Phe Cys Gln Ser Ser
 290 295 300

Lys Trp Leu Ala Met Asn Val Leu Ser Ile Ser Pro Lys Lys Val Ile
 305 310 315 320

Cys Glu Glu Gln Glu His Pro Leu His Glu Leu Leu Asp Lys His Gly
 325 330 335

Phe Glu Val Tyr Pro Ile Pro Phe Arg Asn Val Phe Glu Phe Gly Gly
 340 345 350

Ser Leu His Cys Ala Thr Trp Asp Ile His Arg Thr Gly Thr Cys Glu
 355 360 365

Asp Tyr Phe Pro Lys Leu Asn Tyr Thr Pro Val Thr Ala Ser Thr Asn
 370 375 380

Gly Val Ser Arg Phe Ile Ile
 385 390

<210> 27
 <211> 1313
 <212> ADN
 <213> Homo sapiens

<220>
 <221> RBS
 <222> (55)..(58)
 <223> Trình tự Shine-Dalgarno

<220>
 <221> CDS
 <222> (62)..(1189)

<400> 27
 ggtctcaaga tatgaccatg attacgccaa gcttgcatgc ctgcaggaaa ggagaggatt 60
 g atg aca gat gta acg att aaa gcg acg gag ccg ctg ccg aag gat tgc 109
 Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
 1 5 10 15
 ccg gtt agc tct tat aat gag tgg gat ccg ttg gag gag gtt atc gtg 157

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val	
20 25 30	
ggc cgc gcg gag aac gcc tgt gtt cct cca ttc act atc gag gtg aag	205
Gly Arg Ala Glu Asn Ala Cys Val Pro Phe Thr Ile Glu Val Lys	
35 40 45	
gcc aac act tat gaa aag tac tgg ccg ttc tac cag aag caa gga ggt	253
Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly	
50 55 60	
cat tac ttc cct aag gat cat ctg aaa aaa gcg gta gcc gaa att gag	301
His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu	
65 70 75 80	
gag atg tgc aac atc ctg aaa acg gaa ggc gtg acc gtc cgt cgc ccg	349
Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro	
85 90 95	
gat ccc att gac tgg tca ttg aag tat aaa acc ccg gac ttc gaa agc	397
Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser	
100 105 110	
acg ggt ctt tat tca gcc atg ccg cgt gat atc ctg atc gtt gtt ggt	445
Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly	
115 120 125	
aat gaa att att gaa gcc cct atg gcc tgg ccg agc cgc ttc ttc gag	493
Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu	
130 135 140	
tac cgt gct tat cgg tct att atc aaa gat tac ttc cat cgt ggt gca	541
Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala	
145 150 155 160	
aaa tgg acg act gct cct aag cca acc atg gca gac gag ttg tac aac	589
Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn	
165 170 175	
cag gat tac cct atc cat tca gtc gaa gac cgc cac aag ctg gca gca	637
Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala	
180 185 190	
cag ggt aaa ttt gtc acg act gag ttc gaa ccc tgc ttc gat gct gcc	685
Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala	
195 200 205	
gac ttt att cgt gcc gga ccg gat att ttc gca caa cgc tcg caa gta	733
Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val	
210 215 220	
act aac tac ctc gcc att gag tgg atg cgt ccg cat ctg gcc cct gat	781
Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp	
225 230 235 240	
tac cgt gtc cac atc att agc ttc aaa gac ccg aac ccg atg cac att	829
Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile	
245 250 255	
gac gcg act ttt aat atc atc gga ccc ggt att gtg ctg tcg aat cct	877
Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro	
260 265 270	
gac cgc ccg tgt cac cag att gac ctt ttt aag aaa gct gga tgg acg	925
Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr	
275 280 285	
att atc acc cct cca act ccg att atc cct gat gac cac cca ctc tgg	973
Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp	
290 295 300	
atg tcc tct aaa tgg ctt tcc atg aac gtt ctg atg ctt gac gag aag	1021

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 cgg gtt atg gtt gac gcc aac gaa gtc ccc att cag aaa atg ttt gaa 1069
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335
 aag ctt ggc atc act acg att aag gtc aat atc cgt aac gcg aac agc 1117
 Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350
 ttg ggc ggt ggt ttt cat tgc tgg acg tgc gat gtg cgc cgc cgg gga 1165
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365
 acc ctt cag tcc tat ctc gac taa taagacgtct aaaaaaaaaac cccgccctg 1219
 Thr Leu Gln Ser Tyr Leu Asp
 370 375
 acagggcggg gtttttttta gtcgactcta gaggatcccc gggtaccgag cttegaattca 1279
 ctggccgtcg ttttacagcc aagcttgga gacc 1313

 <210> 28
 <211> 375
 <212> PRT
 <213> Homo sapiens

 <400> 28
 Met Thr Asp Val Thr Ile Lys Ala Thr Glu Pro Leu Pro Lys Asp Cys
 1 5 10 15
 Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30
 Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Ile Glu Val Lys
 35 40 45
 Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Gln Gly Gly
 50 55 60
 His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 65 70 75 80
 Glu Met Cys Asn Ile Leu Lys Thr Glu Gly Val Thr Val Arg Arg Pro
 85 90 95
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110
 Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
 115 120 125
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 130 135 140
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asn

165 170 175
 Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Ile Val Leu Ser Asn Pro
 260 265 270
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285
 Ile Ile Thr Pro Pro Thr Pro Ile Ile Pro Asp Asp His Pro Leu Trp
 290 295 300
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335
 Lys Leu Gly Ile Thr Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365
 Thr Leu Gln Ser Tyr Leu Asp
 370 375

 <210> 29
 <211> 1313
 <212> ADN
 <213> Rattus norvegicus

 <220>
 <221> RBS
 <222> (55)..(58)

 <220>
 <221> CDS
 <222> (62)..(1189)

 <400> 29
 ggtctcaaga tatgaccatg attacgccaa gcttgcatgc ctgcaggaaa ggagaggatt

g atg aca gat gta acg att aaa gct act cat ccg ctt cct aaa gat tgc 109
 Met Thr Asp Val Thr Ile Lys Ala Thr His Pro Leu Pro Lys Asp Cys
 1 5 10 15
 cct gtg tca agc tat aac gaa tgg gac cct ttg gag gag gtg atc gtg 157
 Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30
 ggc cgt gct gag aac gct tgt gta ccc ccc ttt act gtc gaa gta aaa 205
 Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 35 40 45
 gcc aac act tac gag aag tac tgg ccg ttt tac cag aaa aat ggt ggc 253
 Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
 50 55 60
 ctt tat ttc ccc aaa gat cac ctg aag aaa gcc gtt gca gag gtc gaa 301
 Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
 65 70 75 80
 gag atg tgc aat att ctg agc atg gaa ggt gtg act gtg aaa cgc cca 349
 Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
 85 90 95
 gat ccg att gac tgg tca ctt aag tat aag acg ccc gac ttt gag tcg 397
 Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110
 act gga ctt tac agc gcg atg ccg cgc gac att ctg atg gta gtg gga 445
 Thr Gly Leu Tyr Ser Ala Met Pro Arg Asp Ile Leu Met Val Val Gly
 115 120 125
 aat gag atc att gag gcg ccc atg gcg tgg cgg tcc cgc ttt ttt gag 493
 Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 130 135 140
 tat cgc gcg tac cgg tcg att atc aag gat tac ttt cac cgg ggc gcg 541
 Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
 145 150 155 160
 aag tgg acc acc gct cca aag cct acg atg gct gac gag ctt tac gat 589
 Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
 165 170 175
 caa gat tac ccc att cat tcg gtt gaa gac cgt cac aaa ttg gcg gca 637
 Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
 180 185 190
 caa gga aaa ttc gtt acg act gag ttc gag cct tgc ttt gac gcc gcc 685
 Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
 195 200 205
 gat ttt att cgg gct ggc cgc gac atc ttc gca cag cgt tca caa gta 733
 Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
 210 215 220
 acc aac tac ttg gga atc gag tgg atg cgt cgg cac ctt gcc cca gat 781
 Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
 225 230 235 240
 tac cgg gtc cat atc att tca ttt aaa gat ccc aac ccg atg cac att 829
 Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
 245 250 255
 gac gct acc ttc aat att att gga cca gga ctg gta ctg tca aac cct 877
 Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
 260 265 270
 gat cgg ccc tgt cac cag att gat ctg ttc aag aag gct ggt tgg acg 925
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285

att gtc acg cca ccg act cct gta atc ccg gat gat cat ccc ctc tgg 973
 Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300

atg tct tcc aaa tgg ctc tct atg aac gta ctg atg ctc gat gaa aaa 1021
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

cgc gtg atg gta gat gcg aat gag gtc cca att caa aaa atg ttc gag 1069
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
 325 330 335

aag ctg ggc atc tct act att aag gtg aac att cgg aat gca aac tcg 1117
 Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350

ctc gga gga ggt ttc cac tgt tgg act tgt gac gta cgg cgg cgt gga 1165
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365

act ctc cag agc tac ttc gat taa taagacgtct aaaaaaaaaac cccgccctg 1219
 Thr Leu Gln Ser Tyr Phe Asp
 370 375

acagggcggg gtttttttta gtcgactcta gaggatcccc gggtaccgag ctogaattca 1279
 ctggccgctg ttttacagcc aagcttgga gacc 1313

<210> 30
 <211> 375
 <212> PRT
 <213> Rattus norvegicus

<400> 30

Met Thr Asp Val Thr Ile Lys Ala Thr His Pro Leu Pro Lys Asp Cys
 1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys Asn Gly Gly
 50 55 60

Leu Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Val Glu
 65 70 75 80

Glu Met Cys Asn Ile Leu Ser Met Glu Gly Val Thr Val Lys Arg Pro
 85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110

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

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ser Arg Phe Phe Glu
 130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asp Glu Leu Tyr Asp
165 170 175

Gln Asp Tyr Pro Ile His Ser Val Glu Asp Arg His Lys Leu Ala Ala
180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
210 215 220

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
225 230 235 240

Tyr Arg Val His Ile Ile Ser Phe Lys Asp Pro Asn Pro Met His Ile
245 250 255

Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
260 265 270

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
275 280 285

Ile Val Thr Pro Pro Thr Pro Val Ile Pro Asp Asp His Pro Leu Trp
290 295 300

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
305 310 315 320

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Glu
325 330 335

Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
340 345 350

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
355 360 365

Thr Leu Gln Ser Tyr Phe Asp
370 375

<210> 31
<211> 1313
<212> ADN
<213> Galeopterus variegatus

<220>
<221> RBS
<222> (55)..(58)

<220>
<221> CDS

<222> (62)..(1189)

<400> 31

```

ggtctcaaga tatgaccatg attacgccaa gcttgcacgc ctgcaggaaa ggagaggatt      60

g atg aca gat gta acg att aaa gcg acc gat cct ttg cct aaa gat tgc      109
  Met Thr Asp Val Thr Ile Lys Ala Thr Asp Pro Leu Pro Lys Asp Cys
    1             5             10             15

ccc gtg tcc tcg tat aat gaa tgg gac ccg ctc gag gag gta atc gtt      157
Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
      20             25             30

ggt cgg gct gaa aat gcc tgc gtt ccg ccc ttc act gtt gaa gtc aaa      205
Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
    35             40             45

gcg aat act tat gaa aag tac tgg ccg ttc tat caa aaa cac ggt ggc      253
Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys His Gly Gly
    50             55             60

cac tat ttc ccg aaa gac cac ctg aaa aaa gct gtt gct gag att gaa      301
His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
    65             70             75             80

gaa atg tgt aac atc ctg aag atg gaa ggc gtt acg gtt cgt cgg cca      349
Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro
      85             90             95

gac cca atc gat tgg tca ctg aag tac aag acc ccc gac ttt gag tcg      397
Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
    100            105            110

act gga ctc tac gga gct atg ccc cgt gat atc ctc atc gtt gtt ggc      445
Thr Gly Leu Tyr Gly Ala Met Pro Arg Asp Ile Leu Ile Val Val Gly
    115            120            125

aat gaa atc atc gaa gcc cct atg gcc tgg cgg gca cgc ttc ttc gaa      493
Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu
    130            135            140

tac cgg gct tat cgc agc atc atc aag gat tat ttt cac cgg gga gcc      541
Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
    145            150            155            160

aaa tgg acc acg gcc ccg aaa ccc act atg gct aac gag ttg tat gat      589
Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
    165            170            175

caa gac tat ccg atc tat acg gtc gag gac cgg cat aaa ctt gca gcg      637
Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
    180            185            190

caa ggt aaa ttt gtg act acg gag ttc gaa cct tgt ttc gat gcc gct      685
Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
    195            200            205

gat ttt att cgt gca gga cgc gat atc ttt gcg caa cgg agc caa gta      733
Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
    210            215            220

act aat tac ctc ggt atc gag tgg atg cgg cgt cat ctg gca cca gat      781
Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
    225            230            235            240

tac cgt gtt cac atc gtt tcg ttc aaa gat ccg aac ccg atg cac att      829
Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
    245            250            255

gat gcc acc ttc aac atc att ggt ccc ggc ctc gtg ctc agc aat cca      877
Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
    260            265            270

```

gac cgg cct tgc cac caa att gac ttg ttc aaa aaa gct ggt tgg acc 925
 Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
 275 280 285

att att act cct cca att ccc gtg att cca gac gac cac ccg ctt tgg 973
 Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
 290 295 300

atg agc tca aaa tgg ttg agc atg aat gtc ctt atg ctt gat gag aag 1021
 Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
 305 310 315 320

cgt gtc atg gtt gat gct aac gag gta cca atc caa aag atg ttc ggc 1069
 Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Gly
 325 330 335

aag ctg ggt att tca acg atc aaa gtc aac att cgg aac gcc aat tca 1117
 Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
 340 345 350

ctg ggc ggt ggc ttt cac tgc tgg act tgt gat gtt cgt cgg cgg ggc 1165
 Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
 355 360 365

acg ctt caa agc tat ttc gac taa taagacgtct aaaaaaaaaac cccgccctg 1219
 Thr Leu Gln Ser Tyr Phe Asp
 370 375

acagggcggg gtttttttta gtcgactcta gaggatcccc gggtaccgag ctogaattca 1279
 ctggccgtcg ttttacagcc aagcttggga gacc 1313

<210> 32
 <211> 375
 <212> PRT
 <213> Galeopterus variegatus

<400> 32

Met Thr Asp Val Thr Ile Lys Ala Thr Asp Pro Leu Pro Lys Asp Cys
 1 5 10 15

Pro Val Ser Ser Tyr Asn Glu Trp Asp Pro Leu Glu Glu Val Ile Val
 20 25 30

Gly Arg Ala Glu Asn Ala Cys Val Pro Pro Phe Thr Val Glu Val Lys
 35 40 45

Ala Asn Thr Tyr Glu Lys Tyr Trp Pro Phe Tyr Gln Lys His Gly Gly
 50 55 60

His Tyr Phe Pro Lys Asp His Leu Lys Lys Ala Val Ala Glu Ile Glu
 65 70 75 80

Glu Met Cys Asn Ile Leu Lys Met Glu Gly Val Thr Val Arg Arg Pro
 85 90 95

Asp Pro Ile Asp Trp Ser Leu Lys Tyr Lys Thr Pro Asp Phe Glu Ser
 100 105 110

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

Asn Glu Ile Ile Glu Ala Pro Met Ala Trp Arg Ala Arg Phe Phe Glu
130 135 140

Tyr Arg Ala Tyr Arg Ser Ile Ile Lys Asp Tyr Phe His Arg Gly Ala
145 150 155 160

Lys Trp Thr Thr Ala Pro Lys Pro Thr Met Ala Asn Glu Leu Tyr Asp
165 170 175

Gln Asp Tyr Pro Ile Tyr Thr Val Glu Asp Arg His Lys Leu Ala Ala
180 185 190

Gln Gly Lys Phe Val Thr Thr Glu Phe Glu Pro Cys Phe Asp Ala Ala
195 200 205

Asp Phe Ile Arg Ala Gly Arg Asp Ile Phe Ala Gln Arg Ser Gln Val
210 215 220

Thr Asn Tyr Leu Gly Ile Glu Trp Met Arg Arg His Leu Ala Pro Asp
225 230 235 240

Tyr Arg Val His Ile Val Ser Phe Lys Asp Pro Asn Pro Met His Ile
245 250 255

Asp Ala Thr Phe Asn Ile Ile Gly Pro Gly Leu Val Leu Ser Asn Pro
260 265 270

Asp Arg Pro Cys His Gln Ile Asp Leu Phe Lys Lys Ala Gly Trp Thr
275 280 285

Ile Ile Thr Pro Pro Ile Pro Val Ile Pro Asp Asp His Pro Leu Trp
290 295 300

Met Ser Ser Lys Trp Leu Ser Met Asn Val Leu Met Leu Asp Glu Lys
305 310 315 320

Arg Val Met Val Asp Ala Asn Glu Val Pro Ile Gln Lys Met Phe Gly
325 330 335

Lys Leu Gly Ile Ser Thr Ile Lys Val Asn Ile Arg Asn Ala Asn Ser
340 345 350

Leu Gly Gly Gly Phe His Cys Trp Thr Cys Asp Val Arg Arg Arg Gly
355 360 365

Thr Leu Gln Ser Tyr Phe Asp
370 375

<210> 33
<211> 1211
<212> ADN
<213> ADN tổng hợp

<400> 33
tctagagatc cgcgggggcc caggaggggg gatctggcat ttttgggagg tgtgaaatga

60

gtgagaagat cgtcaactcg tggaaacgaat gggatgagct cgaggagatg gtggtcggca 120
 ttgcggacta tgccagcttc gaaccgaaag agccaggcaa ccatcccaaa ctgcgcaacc 180
 agaacctggc cgaatcatc cccttcccaa gcggcccaaa ggaccgaag gtgctggaga 240
 aagcgaacga agagctgaat gggctggctt acctgctgaa ggaccacgat gtgatcgtgc 300
 gtcgtcccgga gaagatcgac ttaccaaga gcctgaaaac cccgtatttc gaggttgcca 360
 accagtactg cggcgtttgt cctcgcgacg tgatgatcac gtttggaac gaaatcatgg 420
 aagcgaccat gtccaaacgt gcacgcttct tcgaatacct cccctatcgg aagctggctt 480
 acgagtactg gaacaaggac gagcacatga tctggaacgc agccccgaaa cggaccatgc 540
 aggatagcat gtacctggaa aacttctggg agctctcgtt ggaagaacgc ttcaagcgga 600
 tgcaagactt cgaattctgc atcaccaag acgaggtgat cttcgatgcc gccgattgct 660
 cccgcttggg taaggacatc ctggtgcagg aaagcatgac caccaatcgc actggcatcc 720
 gctggctgaa gaagcatctc gaaccacgcg gctttcgcgt ccatccggtg cacttccgt 780
 tggacttctt cctagccac atcgactgca cgttcgtacc gttgcgtccg ggtctgatcc 840
 tgaccaatcc ggaacgcccg attcgcgagg aagaggagaa gatcttcaag gagaatggct 900
 gggagctgat caccgtaccg cagcctacct gctcgaacga cgagatgccc atgttctgcc 960
 agagctcgaa atggctgtcc atgaacgtcc tgagcattag tcccaccaag gtgatctgcg 1020
 aagaacggga aaagccgctg caagaactgc tggacaagca cgggttcgaa gtctttccct 1080
 tgcccttccg ccattgtgtt gagttcgggt gcagctttca ctgtgccacg tgggatattc 1140
 gccgcaaggg cgagtgcgag gactacctgc cgaacctgaa ctaccagccg atttgcggct 1200
 gacgcggatc c 1211

<210> 34
 <211> 2028
 <212> ADN
 <213> Escherichia coli

<400> 34
 ttaatctttc tgcaattga gatgacgcca ctggctgggc gtcattcccg tttcccggt 60
 aaacaccacc gaaaatagt tactatcttc aaagccacat tcggtcgaaa tatcactgat 120
 taacaggcgg ctatgctgga gaagatattg cgcattgacac actctgacct gtcgcagata 180
 ttgattgatg gtcattccag tctgctggcg aaattgctga cgcaaaacgc gctcactgca 240
 cgatgctca tcacaaaatt tatccagcgc aaagggactt tttaggctag ccgccagccg 300
 ggtaatcagc ttatccagca acgtttcgtt ggatgttggc ggcaacgaat cactggtgta 360
 acgatggcga tttagcaaca tcaccaactg cccgaacagc aactcagcca tttcgttagc 420
 aaacggcaca tgctgactac tttcatgctc aagctgaccg ataacctgcc gcgcctgcgc 480
 catcccatg ctacctaacg gccagtgtgg ttgccctgcg ctggcggtta atcccggaat 540
 cggccctgc cagtcaagat tcagcttcag acgtccggg caataaataa tattctgcaa 600
 aaccagatcg ttaacggaag cgtaggagtg tttatcgtca gcatgaatgt aaaagagatc 660
 gccacgggta atgcgataag ggcgatcgtt gagtacatgc aggccattac cgcgccagac 720

aatcaccagc tcacaaaaat catgtgtatg ttcagcaaag acatcttgcg gataacggtc 780
agccacagcg actgcctgct ggtecgctggc aaaaaaatca tctttgagaa gttttaactg 840
atgcgccacc gtggctacct cggccagaga acgaagttga ttattcgcaa tatggcgtae 900
aaatacgttg agaagattcg cgttattgca gaaagccatc ccgtccctgg cgaatatcac 960
gcgggtgacca gttaaactct cggcgaaaaa gcgtcgaaaa gtggttactg tcgctgaatc 1020
cacagcgata ggcgatgtca gtaacgctgg cctcgctgtg gcgtagcaga tgcggggctt 1080
tcatcagtcg caggcgggtc aggtatcgct gaggcgtcag tcccgtttgc tgcttaagct 1140
gccgatgtag cgtacgcagt gaaagagaaa attgatccgc cacggcatcc caattcacct 1200
catcggcaaa atggctctcc agccaggcca gaagcaagtt gagacgtgat gcgctgtttt 1260
ccaggttctc ctgcaaaactg cttttacgca gcaagagcag taattgcata aacaagatct 1320
cgcgactggc ggtecgagggt aaatcatttt ccccttctg ctgttccatc tgtgcaacca 1380
gctgtcgcac ctgctgcaat acgctgtggg taacgcgcca gtgagacgga tactgcccac 1440
ccagctcttg tggcagcaac tgattcagcc cggcgagaaa ctgaaatcga tccggcgagc 1500
gatacagcac attggtcaga cacagattat cggatgttgc atacagatgc cgatcatgat 1560
cgcgtagcaa acagaccgtg ccaccgggtg tggatatagg ctgcccatta aacacatgaa 1620
taccctgtcc atgttcgaca atcacaatct catgaaaatc atgatgatgt tcaggaaaat 1680
ccgcctgcgg gagccggggg tctatcgcca cggacgcgtt accagacgga aaaaaatcca 1740
cactatgtaa tacggtcata ctggcctcct gatgtcgtca acacggcgaa atagtaatca 1800
cgaggtcagg ttcttacctt aaattttcga cggaaaaacca cgtaaaaaac gtcgattttt 1860
caagatacag cgtgaatttt caggaaatgc ggtgagcatc acatcaccac aattcagcaa 1920
attgtgaaca tcatcacgtt catctttccc tgggtgccaa tggcccattt tctgtcagt 1980
aacgagaagg tcgcaaatc aggcgctttt tagactgggc gtaatgaa 2028

<210> 35
<211> 107
<212> ADN
<213> Escherichia coli

<400> 35
caaataaaac gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg 60
gtgaacgctc tctgagtag gacaaatccg ccgggagcgg atttgaa 107

<210> 36
<211> 24
<212> ADN
<213> oligonucleotit tổng hợp

<400> 36
tctagagatc cgcggggggc cagg 24

<210> 37
<211> 46
<212> ADN
<213> oligonucleotit tổng hợp

<400> 37
agttccctac tctcgcgtgc tcgaggatc cgcgtcagcc gcaaat 46

<210> 38
 <211> 10235
 <212> ADN
 <213> khác

<400> 38
 tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg 60
 tcttttaaaaa ggccgtaata tccagctgaa cggctctggt ataggtacat tgagcaactg 120
 actgaaatgc ctcaaaatgt tctttacgat gccattgga tatatcaacg gtggtatata 180
 cagtgaatttt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa 240
 atacgcccgg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcaa aagttggccc agggcttccc ggtatcaaca gggacaccag 360
 gatttattta ttctgcgaag tgatcttcg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagt 480
 gcttctgttt ctatcagctg tccctcctgt tcagctactg acgggggtgt gcgtaacggc 540
 aaaagcaccg ccggacatca gcgctagcgg agtggtataact ggcttactat gttggcactg 600
 atgaggggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgta 660
 gcagaatatg tgatacagga tatattccgc ttctcgcgc actgactcgc tacgctcgg 720
 cgttcgactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc 780
 caggaagata cttaacaggg aagtgcaggg gccgcggcaa agccgttttt ccataggctc 840
 cgccccctg acaagcatca cgaaatctga cgctcaaact agtggtggcg aaacccgaca 900
 ggactataaa gataccaggc gtttccccct ggcggtccc tcgtgcgcgc tccgttcct 960
 gccttccggt ttaccggtgt cattccgctg ttatggcgc gtttgtctca ttccagcct 1020
 gacactcagt tccgggtagg cagttcgcgc caagctggac tgtatgcacg aacccccgt 1080
 tcagtcgcac cgctgcgcct tatccggtta ctatcgtctt gagtccaacc cggaaagaca 1140
 tgcaaaagca ccaactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200
 catgcgccgg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg 1320
 cggttttttc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa tttttctaga ttccagtga atttatctct tcaaatgtag 1440
 cacctgaagt cagccccata cgatataagt tgtaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatgcggtag ttatcacag ttaaattgct aacgcagtca ggcaccgtgt 1560
 atgaaatcta acaatgcgct catcgctatc ctccgcaccg tcacctgga tgctgtaggc 1620
 ataggcttgg ttatgccggt actgccggc ctcttgccgg atatcgtcca ttccgacagc 1680
 atcgcacgtc actatggcgt gctgctagcg ctatatgct tgatgcaatt tctatgcgca 1740
 cccgttctcg gagcactgtc cgaccgcttt ggccgcgcgc cagtctgct cgcttcgcta 1800
 cttggagcca ctatcgacta cgcgatcatg gcgaccacac ccgtcctgtg gatcctctac 1860
 gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg cgctatatc 1920

gccgacatca ccgatgggga agatcgggct cgccacttcg ggctcatgag cgcttgtttc 1980
 ggcgtgggta tggtaggagg ccccgtaggc gggggactgt tgggcgcat ctccttgcac 2040
 gcaccattcc ttgcggcggc ggtgctcaac ggctcaacc tactactggg ctgcttecta 2100
 atgcaggagt cgcataaggg agagcgtcga ccgatgccct tgagagcctt caaccagtc 2160
 agctccttcc ggtgggcgcg gggcatgact atcgtcgccg cacttatgac tgtcttcttt 2220
 atcatgcaac tcttaggaca ggtgccggca gcgctctggg tcattttcgg cgaggaccgc 2280
 tttcgtgga gcgcgacgat gatcggcctg tcgcttgcgg tatteggaat cttgcacgcc 2340
 ctcgctcaag ccttcgtcac tggccccgc accaaacgtt tcggcgagaa gcaggccatt 2400
 atcgccggca tggcgccga cgcgctggc tacgtcttgc tggcgctcgc gacgcgaggc 2460
 tggatggcct tccccattat gattcttctc gcttcggcg gcacgggat gcccgcttg 2520
 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgctc 2580
 gcggtcttta ccagcctaac ttcatcatt ggaccgctga tcgtcacggc gatttatgcc 2640
 gcctcgga gcacatggaa cgggttgga tggattgtag gcgcgcctt ataccttgc 2700
 tgctccccg cgttcgctc cggtagcgg agccgggcca cctcgacctg aatggaagcc 2760
 ggccgacact cgctaaggga ttcaccactc caagaattgg agccaatcaa ttcttgga 2820
 gaactgtgaa tgcgcaaacc aaccttggc agaacatata ctcgcttcc gccatctcca 2880
 gcagccgac gcggcgcatc tcgggcagcg ttgggtcctg gccacgggtg cgcagatcg 2940
 tgctctgtc gttgaggacc cggctaggct ggcggggtt ccttactggt tagcagaatg 3000
 aatcaccgat acgcgagcga acgtgaagc actgtgctg caaacgtct gcgacctgag 3060
 caacaacatg aatggtcttc ggtttccgtg tttcgtaaag tctggaaacg cggaagtccc 3120
 ctacgtgctg ctgaagtgc ccgcaacaga gagggaacc aaccggtgat accacgatac 3180
 tatgactgag agtcaacgcc atgagcggcc tcatttctta ttctgagtta caacagtccg 3240
 caccgctgtc cggtagctcc ttccggtgg gcggggcat gactatcgtc gccgcactta 3300
 tgactgtctt ctttatcatg caactcgtg gacaggtgcc ggcagcgccc aacagtcccc 3360
 cggccacggg gcctgccacc ataccacgc cgaaacaagc gccctgcacc attatgttcc 3420
 ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct gtattaacga 3480
 agcgctaacc gtttttatca ggctctggga ggcagaataa atgatcatat cgtcaattat 3540
 tacctccacg gggagagcct gagcaaacg gcctcaggca ttgagaagc acacggtcac 3600
 actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg ctatttaacg 3660
 acctgcctt gaaccgacga ccgggtcgaa tttgctttcg aatttctgcc attcatccgc 3720
 ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa ctgccttaaa 3780
 aaaattacgc ccgcacctgc cactcatcgc agtactgttg taattcatta agcattctgc 3840
 cgacatggaa gccatcacia acggcatgat gaacctgaat cgccagcggc atcagcacct 3900
 tgtcgcttg cgtataatat ttgccatgg atttaaattt aatctttctg cgaattgaga 3960
 tgacgccact ggctgggct catcccggtt tccgggtaa acaccaccga aaaatagtta 4020
 ctatcttcaa agccacattc ggtcgaaata tctactgatta acaggcggt atgctggaga 4080

agatattgcg catgacacac tctgacctgt cgcagatatt gattgatggt cattccagtc	4140
tgctggcgaa attgctgacg caaaacgcgc tcaactgcacg atgcctcatc acaaaattta	4200
tccagcgcaa agggactttt caggctagcc gccagccggg taatcagctt atccagcaac	4260
gtttcgctgg atgttggcgg caacgaatca ctgggtgaac gatggcgatt cagcaacatc	4320
accaactgcc cgaacagcaa ctgagccatt tcgttagcaa acggcacatg ctgactactt	4380
tcattgctcaa gctgaccgat aacctgccgc gcctgcgcca tccccatgct acctaaagcg	4440
cagtgtgggt gccctgcgct ggcgttaaata cccggaatcg cccctgccca gtcaagattc	4500
agcttcagac gctccgggca ataaataata ttctgcaaaa ccagatcggt aacggaagcg	4560
taggagtgtt tatcgtcagc atgaatgtaa aagagatcgc cacgggtaat gcgataaggg	4620
cgatcggtga gtacatgcag gccattaccg cgccagacaa tcaccagctc acaaaaatca	4680
tgtgtatgtt cagcaagac atcttgcgga taacgggtcag ccacagcgac tgctgtctgg	4740
tcgctggcaa aaaaatcatc ttgagaagt tttaactgat gcgccaccgt ggctacctcg	4800
gccagagaac gaagttgatt attcgcaata tggcgtacaa atacgttgag aagattcgcg	4860
ttattgcaga aagccatccc gtccctggcg aatatcacgc ggtgaccagt taaactctcg	4920
gcgaaaaagc gtcgaaaagt ggttactgtc gctgaatcca cagcgatagg cgatgtcagt	4980
aacgctggcc tcgctgtggc gtagcagatg tcgggctttc atcagtcgca ggcggttcag	5040
gtatcgctga ggcgtcagtc ccgtttgctg cttaagctgc cgatgtagcg tacgcagtga	5100
aagagaaaat tgatccgccca cggcatccca attcacctca tcggcaaaat ggtcctccag	5160
ccaggccaga agcaagtga gacgtgatgc gctgttttcc aggttctcct gcaaaactgct	5220
tttacgcagc aagagcagta attgcataaa caagatctcg cgactggcgg tcgagggtaa	5280
atcattttcc ccttctctgt gttccatctg tgcaaccagc tgctgcacct gctgcaatac	5340
gctgtggtta acgcgccagt gagacggata ctgcccatcc agctcttggt gcagcaactg	5400
attcagcccc gcgagaaact gaaatcgatc cggcgagcga tacagcacat tggtcagaca	5460
cagattatcg gtatgttcat acagatgccg atcatgatcg cgtacgaaac agaccgtgcc	5520
accggtgatg gtatagggct gccattaaa cacatgaata cccgtgccat gttcgacaat	5580
cacaatttca tgaaaatcat gatgatgttc aggaaaatcc gcctgcggga gccggggttc	5640
tatcgccacg gacgcgttac cagacggaaa aaaatccaca ctatgtaata cggtcatact	5700
ggcctcctga tgctgtcaac acggcgaaat agtaatcacg aggtcagggt ttaccttaa	5760
attttcgacg gaaaaccacg taaaaaacgt cgatttttca agatacagcg tgaattttca	5820
ggaaatgcgg tgagcatcac atcaccacaa ttcagcaaat tgtgaacatc atcacgttca	5880
tctttccctg gttgccaatg gccattttc ctgtcagtaa cgagaaggtc gcgaattcag	5940
gcgcttttta gactggctgt aatgaacatt taaatgaatt cccttgggac tctagagatc	6000
cgcgggggcc caggaggggg gatctggcat ttttgggagg tgtgaaatga gtgagaagat	6060
cgtcaactcg tggaacgaat gggatgagct cgaggagatg gtggtcggca ttgcggacta	6120
tgccagcttc gaaccgaaag agccaggcaa ccatcccaaa ctgcgcaacc agaacctggc	6180
cgaaatcatc cccttcccaa gcggcccaaa ggaccggaag gtgctggaga aagcgaacga	6240

agagctgaat gggctggcctt acctgctgaa ggaccacgat gtgatcgtgc gtctgccga 6300
 gaagatcgac ttcaccaaga gcctgaaaac cccgtatttc gaggttgcca accagtactg 6360
 cggcgtttgt cctcgcgacg tgatgatcac gtttggaac gaaatcatgg aagcgaccat 6420
 gtccaaacgt gcacgcttct tcgaatacct cccctatcgg aagctggtct acgagtactg 6480
 gaacaaggac gagcacatga tctggaacgc agccccgaaa ccgaccatgc aggatagcat 6540
 gtacctggaa aacttctggg agctctcgct ggaagaacgc ttcaagcgga tgcacgactt 6600
 cgaattctgc atcacccaag acgaggtgat ctctgatgcc gccgattgct cccgcttggg 6660
 taaggacatc ctggtgcagg aaagcatgac caccaatcgc actggcatcc gctggctgaa 6720
 gaagcatctc gaaccacgcg gctttcgcgt ccattccggtg cacttcccggt tggacttctt 6780
 ccctagccac atcgactgca cgttcgtacc gttgcgtccg ggtctgatcc tgaccaatcc 6840
 ggaacgcccg attcgcgagg aagaggagaa gatcttcaag gagaatggct gggagctgat 6900
 caccgtaccg cagcctacct gctcgaaacga cgagatgccc atgttctgcc agagctcgaa 6960
 atggctgtcc atgaacgtcc tgagcattag tcccaccaag gtgatctgcg aagaacggga 7020
 aaagccgctg caagaactgc tggacaagca cgggttcgaa gtctttccct tgcctttccg 7080
 ccatgtgttt gagttcgggtg gcagctttca ctgtgccacg tgggatattc gccgcaaggg 7140
 cgagtgcgag gactacctgc cgaacctgaa ctaccagccg atttgcggtt gacgcggatc 7200
 cctcgagcac gcgagagtag ggaactgcca ggcataaat aaaacgaaaag gctcagtcga 7260
 aagactgggc ctttcgtttt atctgttgtt tgcggtgaa cgctctcctg agtaggacaa 7320
 atccgccggg agcggatttg aacgatgata agctgtcaaa catgagaatt cttgaagacg 7380
 aaagggcctc gtgtgtacaa acgttcgtca aaagggcgac acaaaattcc tgcaggggcc 7440
 ggcccagcgc cggcggtcga gtggcgacgg cgcggcttgt ccgcgccttg gtagattgcc 7500
 tggccgtagg ccagccattt ttgagcggcc agcggccgcg ataggccgac gcgaagcggc 7560
 ggggcgtagg gagegcagcg accgaagggt aggcgctttt tgcagctctt cggctgtgcg 7620
 ctggccagac agttatgcac aggccaggcg ggttttaaga gttttaataa gttttaaaga 7680
 gttttaggcg gaaaaatcgc cttttttctc ttttatatca gtcacttaca tgtgtgaccg 7740
 gttcccaatg tacggctttg ggttcccaat gtacgggttc cggttcccaa tgtacggctt 7800
 tgggttccca atgtacgtgc tatccacagg aaagagacct ttctgacctt tttccctgc 7860
 tagggcaatt tgcctagca tctgctcgt acattaggaa ccggcggtatg ctctgccctc 7920
 gatcaggttg cggtagcgca tgactaggat cgggccagcc tgccccgct cctccttcaa 7980
 atcgtactcc ggcaggtcat ttgacctgat cagcttgccg acggtgaaac agaacttctt 8040
 gaactctccg gcgctgccac tgcgttcgta gatcgtcttg aacaaccatc tggcttctgc 8100
 cttgcctgcg gcgcggcggtg ccaggcggtg gagaaaacgg ccgatgccg gatcgatcaa 8160
 aaagtaatcg ggtgaaccg tcagcacgtc cgggttcttg ccttctgtga tctcgcggta 8220
 catccaatca actagctcga tctcgatgta ctccggccgc ccggtttcgc tctttacgat 8280
 cttgtagcgg ctaatcaagg cttcaccctc ggataccgtc accaggcggc cgttcttggc 8340
 cttcttcgta cgctgcatgg caacgtgcgt ggtgtttaac cgaatgcagg tttctaccag 8400

```

gtcgtctttc tgccttccgc catcggtcgc ccggcagaac ttgagtacgt ccgcaacgtg      8460
tggacggaac acgcggccgg gcttgtctcc cttcccttcc cggtatcggt tcatggattc      8520
ggtagatgg gaaaccgcca tcagtaccag gtcgtaatcc cacacactgg ccatgccggc      8580
cggccctgcg gaaacctcta cgtgcccgtc tggaagctcg tagcggatca cctcgccagc      8640
tcgtcgggtca cgcttcgaca gacggaaaac ggccacgtcc atgatgctgc gactatcgcg      8700
ggtgccacg tcatagagca tcggaacgaa aaaatctggt tgctcgtcgc ccttgggcgg      8760
cttcctaatac gacggcgcac cggctgccgg cggttgccgg gattccttgc ggattcgatc      8820
agcggccgct tgccacgatt caccggggcg tgcttctgcc tcgatgcgtt gccgctgggc      8880
ggcctgcgcg gccttcaact tctccaccag gtcacacccc agcgcgcgcg cgatttgatc      8940
cgggccggat ggtttgcgac cgctcacgcc gattcctcgg gcttgggggt tccagtgcc      9000
ttgcagggcc ggagacaac ccagccgctt acgctcggcc aaccgccgt tccctcacac      9060
atggggcatt ccacggcgtc ggtgcctggt tgttcttgat tttccatgcc gcctccttta      9120
gccgctaaaa ttcactact catttattca tttgtcatt tactctggta gctgcgcgat      9180
gtattcagat agcagctcgg taatggtctt gccttggcgt accgcgtaca tcttcagctt      9240
ggtgtgatcc tccgccggca actgaaagt gaccgccttc atggctggcg tgtctgccag      9300
gctggccaac gttgcagcct tgctcgtcgc tgcgtcggga cggccggcac ttagcgtggt      9360
tgtgcttttg ctcattttct ctttacctca ttaactcaaa tgagttttga ttaatttca      9420
gcggccagcg cctggacctc gcgggcagcg tcgccctcgg gttctgattc aagaacggtt      9480
gtgcggcgcg cggcagtgcc tgggtagctc acgcgctcgc tgatacggga ctcaagaatg      9540
ggcagctcgt acccgccag cgctcggca acctcaccgc cgatgcgcgt gcctttgatc      9600
gcccgcgaca cgacaaaggc cgcttgtagc cttccatccg tgacctcaat gcgctgctta      9660
accagctcca ccagtgccg ggtggcccat atgtcgtaag ggcttggctg caccggaatc      9720
agcacgaagt cggctgcctt gatcgccgac acagccaagt ccgccgcctg gggcgctccg      9780
tcgatcacta cgaagtcgcg ccggccgatg gccttcacgt cgcggtcaat cgtcgggcgg      9840
tcgatgccga caacggttag cggttgatct tcccgcacgg ccgcccaatc gcgggcaactg      9900
ccctggggat cggaatcgac taacagaaca tcggccccgg cgagttgcag ggcgcgggct      9960
agatggggtg cgatggtcgt cttgcctgac ccgcctttct ggttaagtac agcgataacc      10020
ttcatgcgtt ccccttgctt atttgtttat ttactcatcg catcatatac gcagcgaccg      10080
catgacgcaa gctgttttac tcaaatacac atcacctttt tagacggcgg cgctcgggtt      10140
cttcagcggc caagctggcc ggccaggccg ccagcttggc atcagacaaa ccggccagga      10200
tttcatgcag ccgcacggtt ccggatgagc attca      10235

```

```

<210> 39
<211> 921
<212> ADN
<213> Pseudomonas putida

```

```

<220>
<221> CDS
<222> (1)..(921)

```

<400> 39
atg agc gct agg cac ttt ctc tcc ctg ctg gac ttc acc acc gac gaa 48
Met Ser Ala Arg His Phe Leu Ser Leu Leu Asp Phe Thr Thr Asp Glu
1 5 10 15

ttg ctc ggg gtg atc cgc cac ggc atc gag ctg aag gac ctg cgc aag 96
Leu Leu Gly Val Ile Arg His Gly Ile Glu Leu Lys Asp Leu Arg Lys
20 25 30

cga ggc gtg ctg ttc gaa ccg ctg aag aac cgt gtg cta ggc atg atc 144
Arg Gly Val Leu Phe Glu Pro Leu Lys Asn Arg Val Leu Gly Met Ile
35 40 45

ttc gaa aag tcc tcg acc cgt acc cgt gtg tcg ttc gag gcc ggc atg 192
Phe Glu Lys Ser Ser Thr Arg Thr Arg Val Ser Phe Glu Ala Gly Met
50 55 60

atc cag ctc ggc ggc cag gcc atc ttc ctg tcg ccc cgc gac acc cag 240
Ile Gln Leu Gly Gly Gln Ala Ile Phe Leu Ser Pro Arg Asp Thr Gln
65 70 75 80

ctg ggc cgc ggc gag cca att ggt gac agc gcc atc gtg ctg tcg agc 288
Leu Gly Arg Gly Glu Pro Ile Gly Asp Ser Ala Ile Val Leu Ser Ser
85 90 95

atg gtt gat gtg gtg atg atc cgg acc cac gcc cac agc acc ctg acc 336
Met Val Asp Val Val Met Ile Arg Thr His Ala His Ser Thr Leu Thr
100 105 110

gag ttc gcc gcc aag tcg cgt gtg ccc gtg atc aac ggc ctg tcc gac 384
Glu Phe Ala Ala Lys Ser Arg Val Pro Val Ile Asn Gly Leu Ser Asp
115 120 125

gaa tcg cac ccg tgc caa ctg ctg gcc gac atg cag acc ttc gtt gaa 432
Glu Ser His Pro Cys Gln Leu Leu Ala Asp Met Gln Thr Phe Val Glu
130 135 140

cac cgc ggc tcg att cag ggc aag acc gtg acc tgg atc ggc gat ggc 480
His Arg Gly Ser Ile Gln Gly Lys Thr Val Thr Trp Ile Gly Asp Gly
145 150 155 160

ttc aac atg tgc aac tcc tat atc gaa gcc gcc agg cag ttc gat ttc 528
Phe Asn Met Cys Asn Ser Tyr Ile Glu Ala Ala Arg Gln Phe Asp Phe
165 170 175

cag ctg cgc atc gcc tgc ccc gaa ggc tat gag ccg gat caa cgc ttc 576
Gln Leu Arg Ile Ala Cys Pro Glu Gly Tyr Glu Pro Asp Gln Arg Phe
180 185 190

atg gca ctg ggc ggc gac cgc gtg cag atc atc cgg gat gcc agg gaa 624
Met Ala Leu Gly Gly Asp Arg Val Gln Ile Ile Arg Asp Ala Arg Glu
195 200 205

gct gtg cgt gat gca cac ctg gtg gtc acc gat gtc tgg act tcc atg 672
Ala Val Arg Asp Ala His Leu Val Val Thr Asp Val Trp Thr Ser Met
210 215 220

ggc cag gag gag gaa act gca cgg cgc ctg gcg cat ttc gcg cct tac 720
Gly Gln Glu Glu Glu Thr Ala Arg Arg Leu Ala His Phe Ala Pro Tyr
225 230 235 240

cag gtc acc cgc gaa ctg ctc gac ctg gct gca ccc gat gtc ctc ttc 768
Gln Val Thr Arg Glu Leu Leu Asp Leu Ala Ala Pro Asp Val Leu Phe
245 250 255

atg cac tgc ctg ccc gcc cac cgt ggc gag gaa atc agc cag gac ctg 816
Met His Cys Leu Pro Ala His Arg Gly Glu Glu Ile Ser Gln Asp Leu
260 265 270

ctc gac gac cca cgt tcg gtc gcc tgg gac gag gct gaa aac cgc ctg 864
Leu Asp Asp Pro Arg Ser Val Ala Trp Asp Glu Ala Glu Asn Arg Leu

```

                275                280                285
cat gca cag aag gcg ctt ctc gaa ttc ctt gta gaa ccg gct tac cac      912
His Ala Gln Lys Ala Leu Leu Glu Phe Leu Val Glu Pro Ala Tyr His
    290                295                300

cac gca tga      921
His Ala
305

<210> 40
<211> 306
<212> PRT
<213> Pseudomonas putida

<400> 40

Met Ser Ala Arg His Phe Leu Ser Leu Leu Asp Phe Thr Thr Asp Glu
1                5                10                15

Leu Leu Gly Val Ile Arg His Gly Ile Glu Leu Lys Asp Leu Arg Lys
    20                25                30

Arg Gly Val Leu Phe Glu Pro Leu Lys Asn Arg Val Leu Gly Met Ile
    35                40                45

Phe Glu Lys Ser Ser Thr Arg Thr Arg Val Ser Phe Glu Ala Gly Met
    50                55                60

Ile Gln Leu Gly Gly Gln Ala Ile Phe Leu Ser Pro Arg Asp Thr Gln
65                70                75                80

Leu Gly Arg Gly Glu Pro Ile Gly Asp Ser Ala Ile Val Leu Ser Ser
    85                90                95

Met Val Asp Val Val Met Ile Arg Thr His Ala His Ser Thr Leu Thr
    100                105                110

Glu Phe Ala Ala Lys Ser Arg Val Pro Val Ile Asn Gly Leu Ser Asp
    115                120                125

Glu Ser His Pro Cys Gln Leu Leu Ala Asp Met Gln Thr Phe Val Glu
    130                135                140

His Arg Gly Ser Ile Gln Gly Lys Thr Val Thr Trp Ile Gly Asp Gly
145                150                155                160

Phe Asn Met Cys Asn Ser Tyr Ile Glu Ala Ala Arg Gln Phe Asp Phe
    165                170                175

Gln Leu Arg Ile Ala Cys Pro Glu Gly Tyr Glu Pro Asp Gln Arg Phe
    180                185                190

Met Ala Leu Gly Gly Asp Arg Val Gln Ile Ile Arg Asp Ala Arg Glu
    195                200                205

Ala Val Arg Asp Ala His Leu Val Val Thr Asp Val Trp Thr Ser Met
    210                215                220

```

Gly Gln Glu Glu Glu Thr Ala Arg Arg Leu Ala His Phe Ala Pro Tyr
225 230 235 240

Gln Val Thr Arg Glu Leu Leu Asp Leu Ala Ala Pro Asp Val Leu Phe
245 250 255

Met His Cys Leu Pro Ala His Arg Gly Glu Glu Ile Ser Gln Asp Leu
260 265 270

Leu Asp Asp Pro Arg Ser Val Ala Trp Asp Glu Ala Glu Asn Arg Leu
275 280 285

His Ala Gln Lys Ala Leu Leu Glu Phe Leu Val Glu Pro Ala Tyr His
290 295 300

His Ala
305

<210> 41
<211> 1218
<212> ADN
<213> Pseudomonas putida

<220>
<221> CDS
<222> (1)..(1218)

<400> 41
atg gcg gac gta aaa aag gtc gta ctg gcg tat tcc ggc ggc ctt gat 48
Met Ala Asp Val Lys Lys Val Val Leu Ala Tyr Ser Gly Gly Leu Asp
1 5 10 15
act tcg gtg att ctc aag tgg ctg cag gat acc tac aac tgc gaa gtg 96
Thr Ser Val Ile Leu Lys Trp Leu Gln Asp Thr Tyr Asn Cys Glu Val
20 25 30
gtg acc ttc acc gct gac ctg ggg cag ggc gaa gag gtc gaa ccg gcc 144
Val Thr Phe Thr Ala Asp Leu Gly Gln Gly Glu Glu Val Glu Pro Ala
35 40 45
cgt gcc aag gcc cag gca atg ggc gtt aaa gag atc tac atc gac gac 192
Arg Ala Lys Ala Gln Ala Met Gly Val Lys Glu Ile Tyr Ile Asp Asp
50 55 60
ctg cgc gaa gaa ttc gtg cgt gat ttc gtg ttc ccg atg ttc cgc gcc 240
Leu Arg Glu Glu Phe Val Arg Asp Phe Val Phe Pro Met Phe Arg Ala
65 70 75 80
aac acc gtc tac gaa ggc gag tac ctg ctg ggt act tcc atc gcc cgt 288
Asn Thr Val Tyr Glu Gly Glu Tyr Leu Leu Gly Thr Ser Ile Ala Arg
85 90 95
ccg ctg atc gcc aag cgc ctg atc gaa atc gcc aac gaa acc ggc gct 336
Pro Leu Ile Ala Lys Arg Leu Ile Glu Ile Ala Asn Glu Thr Gly Ala
100 105 110
gac gcc att tcc cat ggc gcc acc ggc aag ggt aac gac cag gtg cgc 384
Asp Ala Ile Ser His Gly Ala Thr Gly Lys Gly Asn Asp Gln Val Arg
115 120 125
ttc gag ctg ggt gcc tat gcc ctg aag cca ggc gtc aag gtc atc gct 432
Phe Glu Leu Gly Ala Tyr Ala Leu Lys Pro Gly Val Lys Val Ile Ala
130 135 140

cca tgg cgc gag tgg gac ctg ctg tcc cgc gaa aag ctg atg gac tac 480
Pro Trp Arg Glu Trp Asp Leu Leu Ser Arg Glu Lys Leu Met Asp Tyr
145 150 155 160

gcc gag aag cac ggc atc ccg atc gag cgc cac ggc aag aag aag tcg 528
Ala Glu Lys His Gly Ile Pro Ile Glu Arg His Gly Lys Lys Lys Ser
165 170 175

ccg tac tcg atg gac gcc aac ctg ctg cac atc tcc tac gag ggc ggt 576
Pro Tyr Ser Met Asp Ala Asn Leu Leu His Ile Ser Tyr Glu Gly Gly
180 185 190

gtc ctg gaa gat acc tgg acc gag cac gaa gaa gac atg tgg cgc tgg 624
Val Leu Glu Asp Thr Trp Thr Glu His Glu Glu Asp Met Trp Arg Trp
195 200 205

agt gtc tcg cct gag aat gcc ccg gac cag gct acc tac atc gag ctg 672
Ser Val Ser Pro Glu Asn Ala Pro Asp Gln Ala Thr Tyr Ile Glu Leu
210 215 220

acc tac cgc aat ggt gac atc gtt gcc atc gac ggc gtc gag aaa tcc 720
Thr Tyr Arg Asn Gly Asp Ile Val Ala Ile Asp Gly Val Glu Lys Ser
225 230 235 240

ccg gcc acc gtc ctg gca gac ctg aac cgt atc ggt ggt gcc aac ggc 768
Pro Ala Thr Val Leu Ala Asp Leu Asn Arg Ile Gly Gly Ala Asn Gly
245 250 255

atc ggc cgt ctg gac atc gtc gaa aac cgt tac gtc ggc atg aag tcg 816
Ile Gly Arg Leu Asp Ile Val Glu Asn Arg Tyr Val Gly Met Lys Ser
260 265 270

cgc ggt tgc tac gaa acg cct ggc ggt acc atc atg ctc aag gca cac 864
Arg Gly Cys Tyr Glu Thr Pro Gly Gly Thr Ile Met Leu Lys Ala His
275 280 285

cgt gcc atc gag tcg atc acc ctg gac cgc gaa gtc gct cac ctg aaa 912
Arg Ala Ile Glu Ser Ile Thr Leu Asp Arg Glu Val Ala His Leu Lys
290 295 300

gat gag ctg atg cca aag tat gcc agc ctg atc tac acc ggc tac tgg 960
Asp Glu Leu Met Pro Lys Tyr Ala Ser Leu Ile Tyr Thr Gly Tyr Trp
305 310 315 320

tgg agc ccg gag cgt ctg atg ctg caa cag atg atc gat gct tcg cag 1008
Trp Ser Pro Glu Arg Leu Met Leu Gln Gln Met Ile Asp Ala Ser Gln
325 330 335

gtc aac gtg aat ggt gtg gtg cgc ctg aaa ctg tac aag ggc aac gtg 1056
Val Asn Val Asn Gly Val Val Arg Leu Lys Leu Tyr Lys Gly Asn Val
340 345 350

acc gtg gtt ggc cgc aag tcg gac gat tcg ctg ttc gat gcc aac atc 1104
Thr Val Val Gly Arg Lys Ser Asp Asp Ser Leu Phe Asp Ala Asn Ile
355 360 365

gcc acc ttt gaa gaa gat ggt ggt gcc tac aac cag gca gat gct gct 1152
Ala Thr Phe Glu Glu Asp Gly Gly Ala Tyr Asn Gln Ala Asp Ala Ala
370 375 380

ggc ttc atc aag ctc aat gca ctg cgt atg cgc att gcc gcc aac aag 1200
Gly Phe Ile Lys Leu Asn Ala Leu Arg Met Arg Ile Ala Ala Asn Lys
385 390 395 400

ggc cgt tcg ctg ctc tga 1218
Gly Arg Ser Leu Leu
405

<210> 42
<211> 405

<212> PRT

<213> Pseudomonas putida

<400> 42

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

Thr Ser Val Ile Leu Lys Trp Leu Gln Asp Thr Tyr Asn Cys Glu Val
 20 25 30

Val Thr Phe Thr Ala Asp Leu Gly Gln Gly Glu Glu Val Glu Pro Ala
 35 40 45

Arg Ala Lys Ala Gln Ala Met Gly Val Lys Glu Ile Tyr Ile Asp Asp
 50 55 60

Leu Arg Glu Glu Phe Val Arg Asp Phe Val Phe Pro Met Phe Arg Ala
 65 70 75 80

Asn Thr Val Tyr Glu Gly Glu Tyr Leu Leu Gly Thr Ser Ile Ala Arg
 85 90 95

Pro Leu Ile Ala Lys Arg Leu Ile Glu Ile Ala Asn Glu Thr Gly Ala
 100 105 110

Asp Ala Ile Ser His Gly Ala Thr Gly Lys Gly Asn Asp Gln Val Arg
 115 120 125

Phe Glu Leu Gly Ala Tyr Ala Leu Lys Pro Gly Val Lys Val Ile Ala
 130 135 140

Pro Trp Arg Glu Trp Asp Leu Leu Ser Arg Glu Lys Leu Met Asp Tyr
 145 150 155 160

Ala Glu Lys His Gly Ile Pro Ile Glu Arg His Gly Lys Lys Lys Ser
 165 170 175

Pro Tyr Ser Met Asp Ala Asn Leu Leu His Ile Ser Tyr Glu Gly Gly
 180 185 190

Val Leu Glu Asp Thr Trp Thr Glu His Glu Glu Asp Met Trp Arg Trp
 195 200 205

Ser Val Ser Pro Glu Asn Ala Pro Asp Gln Ala Thr Tyr Ile Glu Leu
 210 215 220

Thr Tyr Arg Asn Gly Asp Ile Val Ala Ile Asp Gly Val Glu Lys Ser
 225 230 235 240

Pro Ala Thr Val Leu Ala Asp Leu Asn Arg Ile Gly Gly Ala Asn Gly
 245 250 255

Ile Gly Arg Leu Asp Ile Val Glu Asn Arg Tyr Val Gly Met Lys Ser
 260 265 270

Arg Gly Cys Tyr Glu Thr Pro Gly Gly Thr Ile Met Leu Lys Ala His
275 280 285

Arg Ala Ile Glu Ser Ile Thr Leu Asp Arg Glu Val Ala His Leu Lys
290 295 300

Asp Glu Leu Met Pro Lys Tyr Ala Ser Leu Ile Tyr Thr Gly Tyr Trp
305 310 315 320

Trp Ser Pro Glu Arg Leu Met Leu Gln Gln Met Ile Asp Ala Ser Gln
325 330 335

Val Asn Val Asn Gly Val Val Arg Leu Lys Leu Tyr Lys Gly Asn Val
340 345 350

Thr Val Val Gly Arg Lys Ser Asp Asp Ser Leu Phe Asp Ala Asn Ile
355 360 365

Ala Thr Phe Glu Glu Asp Gly Gly Ala Tyr Asn Gln Ala Asp Ala Ala
370 375 380

Gly Phe Ile Lys Leu Asn Ala Leu Arg Met Arg Ile Ala Ala Asn Lys
385 390 395 400

Gly Arg Ser Leu Leu
405

<210> 43
<211> 1407
<212> ADN
<213> Pseudomonas putida

<220>
<221> CDS
<222> (1)..(1407)

<400> 43
atg agt gaa tcc atg agc acc gag aag acc aat cag tcc tgg ggc ggc 48
Met Ser Glu Ser Met Ser Thr Glu Lys Thr Asn Gln Ser Trp Gly Gly
1 5 10 15
cgc ttc agt gag ccc gtc gac gcc ttc gtc gcc cgt ttc acc gcc tcg 96
Arg Phe Ser Glu Pro Val Asp Ala Phe Val Ala Arg Phe Thr Ala Ser
20 25 30
gta gat ttc gac aag cgc ctg tac cgt cac gac atc atg ggt tcg att 144
Val Asp Phe Asp Lys Arg Leu Tyr Arg His Asp Ile Met Gly Ser Ile
35 40 45
gcc cat gcc acc atg ctg gcg cag gtc ggc gtg ctc agt gat gcc gag 192
Ala His Ala Thr Met Leu Ala Gln Val Gly Val Leu Ser Asp Ala Glu
50 55 60
cgc gac acc atc atc gat ggc ctg aaa acc atc cag ggc gag att gaa 240
Arg Asp Thr Ile Ile Asp Gly Leu Lys Thr Ile Gln Gly Glu Ile Glu
65 70 75 80
gcc ggc aac ttc gac tgg cgt gtc gac ctc gaa gac gtg cac atg aac 288
Ala Gly Asn Phe Asp Trp Arg Val Asp Leu Glu Asp Val His Met Asn
85 90 95

atc gaa gca cgc ctg acc gac cgc atc ggc atc acc ggc aag aag ctg Ile Glu Ala Arg Leu Thr Asp Arg Ile Gly Ile Thr Gly Lys Lys Leu 100 105 110	336
cat act ggg cgt agc cgc aac gac cag gtg gcc acc gac atc cgc ctt His Thr Gly Arg Ser Arg Asn Asp Gln Val Ala Thr Asp Ile Arg Leu 115 120 125	384
tgg ctg cgc gac gaa atc gac ctg atc ctg ggc gaa atc acc cgc ctg Trp Leu Arg Asp Glu Ile Asp Leu Ile Leu Gly Glu Ile Thr Arg Leu 130 135 140	432
cag cag ggc ctg ctg gag cag gca gag cgt gaa gcc gaa acc atc atg Gln Gln Gly Leu Leu Glu Gln Ala Glu Arg Glu Ala Glu Thr Ile Met 145 150 155 160	480
cct ggt ttc acc cac ctg cag acg gcg cag ccg gtc acc ttt ggc cac Pro Gly Phe Thr His Leu Gln Thr Ala Pro Val Thr Phe Gly His 165 170 175	528
cac ctg ctg gcg tgg ttc gaa atg ctc agc cgc gac tat gag cgc ctg His Leu Leu Ala Trp Phe Glu Met Leu Ser Arg Asp Tyr Glu Arg Leu 180 185 190	576
gtc gac tgc cgc aag cgc acc aac cgc atg cca ctg ggc agc gcc gcg Val Asp Cys Arg Lys Arg Thr Asn Arg Met Pro Leu Gly Ser Ala Ala 195 200 205	624
ctg gcc ggc acc acc tac ccg atc gac cgt gaa ctg acc tgc aag ctg Leu Ala Gly Thr Thr Tyr Pro Ile Asp Arg Glu Leu Thr Cys Lys Leu 210 215 220	672
ctg gcc ttt gaa gcc gtg gcc ggc aac tcg ctg gat ggc gtg tcg gac Leu Gly Phe Glu Ala Val Ala Gly Asn Ser Leu Asp Gly Val Ser Asp 225 230 235 240	720
cgt gat ttc gcc atc gaa ttc tgc gcc gct gcc agc gtg gcg atg atg Arg Asp Phe Ala Ile Glu Phe Cys Ala Ala Ser Val Ala Met Met 245 250 255	768
cac ctt tcg cgc ttc tcc gaa gag ctg gtg ctg tgg acc agc gcg cag His Leu Ser Arg Phe Ser Glu Glu Leu Val Leu Trp Thr Ser Ala Gln 260 265 270	816
ttc cag ttc atc gac ctt ccg gac cgc ttc tgc act ggc agc tcg atc Phe Gln Phe Ile Asp Leu Pro Asp Arg Phe Cys Thr Gly Ser Ser Ile 275 280 285	864
atg ccg cag aaa aag aac ccg gac gtg cca gag ctg gta cgt ggc aag Met Pro Gln Lys Lys Asn Pro Asp Val Pro Glu Leu Val Arg Gly Lys 290 295 300	912
agc ggc cgc gtg ttc ggc gcc ctg acc ggc ctg ctg acc ctg atg aaa Ser Gly Arg Val Phe Gly Ala Leu Thr Gly Leu Thr Leu Met Lys 305 310 315 320	960
ggc caa ccg ctg gcc tac aac aag gac aac cag gaa gac aag gaa ccg Gly Gln Pro Leu Ala Tyr Asn Lys Asp Asn Gln Glu Asp Lys Glu Pro 325 330 335	1008
ctg ttc gac gcc gcc gat acc ctg cgc gac tcg ctg cgg gcc ttc gct Leu Phe Asp Ala Ala Asp Thr Leu Arg Asp Ser Leu Arg Ala Phe Ala 340 345 350	1056
gac atg atc ccg gcg atc aag ccc aag cac gcc atc atg cgt gaa gcg Asp Met Ile Pro Ala Ile Lys Pro Lys His Ala Ile Met Arg Glu Ala 355 360 365	1104
gcc ctg cgc ggt ttc tcc acc gct acc gac ctg gct gac tat ctg gtt Ala Leu Arg Gly Phe Ser Thr Ala Thr Asp Leu Ala Asp Tyr Leu Val 370 375 380	1152

cgc cgt ggc ctg ccg ttc cgt gac tgc cac gag atc gtt ggc cac gcg 1200
 Arg Arg Gly Leu Pro Phe Arg Asp Cys His Glu Ile Val Gly His Ala
 385 390 395 400

gtg aag tat ggt gtg gac act ggc aag gac ctg gcc gag atg agc ctg 1248
 Val Lys Tyr Gly Val Asp Thr Gly Lys Asp Leu Ala Glu Met Ser Leu
 405 410 415

gac gaa ctg cgc caa ttc agc gac cag atc gag cag gac gtg ttt gcc 1296
 Asp Glu Leu Arg Gln Phe Ser Asp Gln Ile Glu Gln Asp Val Phe Ala
 420 425 430

gtg ctg acg ctg gaa ggc tcg gtg aat gcg cgt gac cac att ggt ggt 1344
 Val Leu Thr Leu Glu Gly Ser Val Asn Ala Arg Asp His Ile Gly Gly
 435 440 445

acg gcg ccg gcg cag gtg cgt gct gcc gtc gtt cgt ggc aag gcc ctg 1392
 Thr Ala Pro Ala Gln Val Arg Ala Ala Val Val Arg Gly Lys Ala Leu
 450 455 460

ttg gcg tct cgc taa 1407
 Leu Ala Ser Arg
 465

<210> 44
 <211> 468
 <212> PRT
 <213> Pseudomonas putida

<400> 44

Met Ser Glu Ser Met Ser Thr Glu Lys Thr Asn Gln Ser Trp Gly Gly
 1 5 10 15

Arg Phe Ser Glu Pro Val Asp Ala Phe Val Ala Arg Phe Thr Ala Ser
 20 25 30

Val Asp Phe Asp Lys Arg Leu Tyr Arg His Asp Ile Met Gly Ser Ile
 35 40 45

Ala His Ala Thr Met Leu Ala Gln Val Gly Val Leu Ser Asp Ala Glu
 50 55 60

Arg Asp Thr Ile Ile Asp Gly Leu Lys Thr Ile Gln Gly Glu Ile Glu
 65 70 75 80

Ala Gly Asn Phe Asp Trp Arg Val Asp Leu Glu Asp Val His Met Asn
 85 90 95

Ile Glu Ala Arg Leu Thr Asp Arg Ile Gly Ile Thr Gly Lys Lys Leu
 100 105 110

His Thr Gly Arg Ser Arg Asn Asp Gln Val Ala Thr Asp Ile Arg Leu
 115 120 125

Trp Leu Arg Asp Glu Ile Asp Leu Ile Leu Gly Glu Ile Thr Arg Leu
 130 135 140

Gln Gln Gly Leu Leu Glu Gln Ala Glu Arg Glu Ala Glu Thr Ile Met
 145 150 155 160

Pro Gly Phe Thr His Leu Gln Thr Ala Gln Pro Val Thr Phe Gly His
 165 170 175
 His Leu Leu Ala Trp Phe Glu Met Leu Ser Arg Asp Tyr Glu Arg Leu
 180 185 190
 Val Asp Cys Arg Lys Arg Thr Asn Arg Met Pro Leu Gly Ser Ala Ala
 195 200 205
 Leu Ala Gly Thr Thr Tyr Pro Ile Asp Arg Glu Leu Thr Cys Lys Leu
 210 215 220
 Leu Gly Phe Glu Ala Val Ala Gly Asn Ser Leu Asp Gly Val Ser Asp
 225 230 235 240
 Arg Asp Phe Ala Ile Glu Phe Cys Ala Ala Ala Ser Val Ala Met Met
 245 250 255
 His Leu Ser Arg Phe Ser Glu Glu Leu Val Leu Trp Thr Ser Ala Gln
 260 265 270
 Phe Gln Phe Ile Asp Leu Pro Asp Arg Phe Cys Thr Gly Ser Ser Ile
 275 280 285
 Met Pro Gln Lys Lys Asn Pro Asp Val Pro Glu Leu Val Arg Gly Lys
 290 295 300
 Ser Gly Arg Val Phe Gly Ala Leu Thr Gly Leu Leu Thr Leu Met Lys
 305 310 315 320
 Gly Gln Pro Leu Ala Tyr Asn Lys Asp Asn Gln Glu Asp Lys Glu Pro
 325 330 335
 Leu Phe Asp Ala Ala Asp Thr Leu Arg Asp Ser Leu Arg Ala Phe Ala
 340 345 350
 Asp Met Ile Pro Ala Ile Lys Pro Lys His Ala Ile Met Arg Glu Ala
 355 360 365
 Ala Leu Arg Gly Phe Ser Thr Ala Thr Asp Leu Ala Asp Tyr Leu Val
 370 375 380
 Arg Arg Gly Leu Pro Phe Arg Asp Cys His Glu Ile Val Gly His Ala
 385 390 395 400
 Val Lys Tyr Gly Val Asp Thr Gly Lys Asp Leu Ala Glu Met Ser Leu
 405 410 415
 Asp Glu Leu Arg Gln Phe Ser Asp Gln Ile Glu Gln Asp Val Phe Ala
 420 425 430
 Val Leu Thr Leu Glu Gly Ser Val Asn Ala Arg Asp His Ile Gly Gly
 435 440 445

Thr Ala Pro Ala Gln Val Arg Ala Ala Val Val Arg Gly Lys Ala Leu
 450 455 460

Leu Ala Ser Arg
 465

<210> 45
 <211> 3701
 <212> ADN
 <213> Pseudomonas putida

<400> 45
 tctagagatc cgcgggggcc cgacctgcat accgttgacg ataaggtagt cactcatgag 60
 cgctaggcac tttctctccc tgctggactt caccaccgac gaattgctcg gggtagtccg 120
 ccacggcatc gagctgaagg acctgcgcaa gcgaggcgtg ctgttcgaac cgctgaagaa 180
 ccgtgtgcta ggcgatgatc tcgaaaagtc ctgcacccgt acccgtgtgt cgttcgaggc 240
 cggcatgatc cagctcggcg gccaggccat ctctctgtcg ccccgcgaca ccagctggg 300
 ccgcggcgag ccaattgggt acagcgccat cgtgctgtcg agcatgggtg atgtgggtgat 360
 gatccggacc caccgccaca gcacctgac cgagtctgcc gccaaagtcg gtgtgcccg 420
 gatcaacggc ctgtccgacg aatcgacccc gtgccaactg ctggccgaca tgcagacctt 480
 cgttgaacac cgcggctcga ttcagggcaa gacctgacc tggatcgcg atggttcaa 540
 catgtgcaac tcctatatcg aagccgccag gcagttcgat ttccagctgc gcatcgctg 600
 cccgaaggc tatgagccg atcaacgctt catggcactg ggcggcgacc gcgtgcagat 660
 catccgggat gccagggaag ctgtgcgtga tgcacacctg gtggtcaccg atgtctggac 720
 ttccatgggt caggaggagg aaactgcacg gcgcctggcg catttcgcg cttaccaggt 780
 caccgcgaa ctgctcgacc tggctgcacc cgatgtctc ttcatgact gcctgcccgc 840
 ccaccgtggc gaggaaatca gccaggacct gctcgacgac ccacgttcgg tcgcctggga 900
 cgaggctgaa aaccgcctgc atgcacagaa ggcgttctc gaattccttg tagaacggc 960
 ttaccaccac gcatgagta accgtacaac cccgtggagt gatggcatgg cggacgtaaa 1020
 aaaggtcgta ctggcgatt ccggcgccct tgatacttcg gtgattctca agtggctgca 1080
 ggatacctac aactgcgaag tggtagacct caccgctgac ctggggcagg gcgaagaggt 1140
 cgaaccggcc cgtgccagg cccaggcaat gggcgtaaa gagatctaca tcgacgacct 1200
 gcgcgaagaa ttcgtgcgtg atttcgtgtt cccgatgtc cgcgccaaca ccgtctacga 1260
 aggcgagtac ctgctgggta ctccatcgc cgtccgctg atcgccaagc gcctgatcga 1320
 aatcgccaac gaaaccggcg ctgacgcat ttccatggc gccaccggca agggtaacga 1380
 ccaggtcgc ttcgagctgg gtgcctatgc cctgaagcca ggcgtcaagg tcctcgtcc 1440
 atggcgcgag tgggacctgc tgtcccgcga aaagctgatg gactacgccg agaagcacgg 1500
 catcccgatc gagcgccacg gcaagaagaa gtgcgctac tcgatggagc ccaacctgct 1560
 gcacatctcc tacgagggcg gtgtcctgga agatacctgg accgagcacg aagaagacat 1620
 gtggcgctgg agtgtctcgc ctgagaatgc cccggaccag gctacctaca tcgagctgac 1680
 ctaccgcaat ggtgacatcg ttgccatcga cggcgctcga aaatccccgg ccaccgtct 1740

```

ggcagacctg aaccgtatcg gtggtgccaa cggcatcggc cgtctggaca tcgtcgaaaa 1800
ccgttacgtc ggcatgaagt cgcgcgggtg ctacgaaacg cctggcggtg ccatcatgct 1860
caaggcacac cgtgccatcg agtcgatcac cctggaccgc gaagtcgctc acctgaaaga 1920
tgagctgatg ccaaagtatg ccagcctgat ctacaccggc tactggtgga gcccgagcg 1980
tctgatgctg caacagatga tcgatgcttc gcaggtcaac gtgaatggtg tggcgcgct 2040
gaaactgtac aagggaacg tgaccgtggt tggccgcaag tcggacgatt cgctgttcga 2100
tgccaacatc gccacctttg aagaagatgg tggcgcttac aaccaggcag atgctgctgg 2160
cttcatcaag ctcaatgcac tgcgtatgcy cattgcgcgc aacaagggcc gtctgctgct 2220
ctgattgcta tcgacgccac ttttctgttc acgcctgcaa tgagtgaatc catgagcacc 2280
gagaagacca atcagtcctg gggcgccgcg ttcagtgage ccgtcgacgc ctctgctgcc 2340
cgttttaccc cctcggtaga tttcgacaag cgcctgtacc gtcacgacat catgggttcg 2400
attgcccatg ccaccatgct ggcgaggtc ggctgctca gtgatgccga gcgcgacacc 2460
atcatcgatg gcctgaaaac catccagggc gagattgaag ccggcaactt cgactggcgt 2520
gtcgacctcg aagacgtgca catgaacatc gaagcacgcc tgaccgaccg catcggcac 2580
accggcaaga agctgcatac tggcgtagc cgcaacgacc aggtggccac cgacatccgc 2640
ctttggctgc gcgacgaaat cgacctgatc ctggcgcaaa tcacccgcct gcagcagggc 2700
ctgctggagc aggcagagcg tgaagccgaa accatcatgc ctggtttcac ccacctgcag 2760
acggcgagc cggtcacctt tggccaccac ctgctggcgt ggttcgaaat gctcagccgc 2820
gactatgagc gcctggctga ctgcccgaag cgcaccaacc gcatgccact gggcagcgcc 2880
gcgctggcgc gcaccaccta cccgatcgac cgtgaactga cctgcaagct gctgggcttt 2940
gaagccgtgg ccggcaactc gctggatggc gtgtcggacc gtgatttcgc catcgaattc 3000
tgcgcgctg ccagcgtggc gatgatgcac ctttcgcgt tctccgaaga gctggtgctg 3060
tggaccagcg cgcagttcca gttcatcgac cttccggacc gttctgcac tggcagctcg 3120
atcatgcgc agaaaaagaa cccggacgtg ccagagctgg tacgtggcaa gagcgccgc 3180
gtgttcggcg ccctgaccgg cctgctgacc ctgatgaaag gccaaaccgt gccctacaac 3240
aaggacaacc aggaagacaa ggaaccgtg ttcgacgcc cggataccct gcgcgactcg 3300
ctgcgggcct tcgctgacat gatcccgcg atcaagccca agcacgccat catgctgaa 3360
gcggccctgc gcggtttctc caccgctacc gacctggctg actatctggt tcgcgtggc 3420
ctgccgttcc gtgactgcca cgagatcggt ggccacgcgg tgaagtatgg tgtggacact 3480
ggcaaggacc tggccgagat gagcctggac gaactgcgcc aattcagcga ccagatcgag 3540
caggacgtgt ttgcgtgct gacgctggaa ggctcggtga atgcgcgtga ccacattggt 3600
ggtacggcgc cggcgaggt gcgtgctgcc gtcgttcgtg gcaaggccct gttggcgtct 3660
cgtaatccc ccaaggctcg agcacgcgag agtagggaac t 3701

```

<210> 46

<211> 23

<212> ADN

<213> oligonucleotit tổng hợp

<400> 46
 ggatccgcgt cagccgcaaa tcg 23

<210> 47
 <211> 45
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 47
 tttgcggctg acgcggatcc cgacctgcat accgttgcatg ataag 45

<210> 48
 <211> 22
 <212> ADN
 <213> oligonucleotit tổng hợp

<400> 48
 agttccctac tctcgctgc tc 22

<210> 49
 <211> 13891
 <212> ADN
 <213> khác

<400> 49
 tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg 60
 tcttttaaaaa ggccgtaata tccagctgaa cggctctggt ataggtacat tgagcaactg 120
 actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc 180
 cagtgatttt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa 240
 atacgccccg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcca aagttggccc agggcttccc ggtatcaaca gggacaccag 360
 gatttattta ttctgcgaag tgatcttcg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg 480
 gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc 540
 aaaagcacgc cggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg 600
 atgaggggtg cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgta 660
 gcagaatatg tgatacagga tatattccgc ttctcgcgc actgactcgc tacgctcgg 720
 cgttcgactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc 780
 caggaagata cttaacaggg aagtgaagg gccgcggcaa agccgttttt ccataggctc 840
 cgcctccctg acaagcatca cgaaatctga cgctcaaact agtgggtggc aaacccgaca 900
 ggactataaa gataccaggc gtttccccct ggcggtccc tegtgcgcgc tctgttcct 960
 gcctttcgggt ttaccgggtg cattccgctg ttatggcgc gtttgtctca ttccagcct 1020
 gacactcagt tccgggtagg cagttcgcgc caagctggac tgtatgcacg aacccccgt 1080
 tcagtcgcac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggaaagaca 1140
 tgcaaaagca ccaactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200
 catgcgcggg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg 1320

cggttttttc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa tatttctaga tticagtga atttatctct tcaaatgtag 1440
 cacctgaagt cagccccata cgatataagt tgtaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatgcggtag tttatcacag ttaaattgct aacgcagtca ggcaccgtgt 1560
 atgaaatcta acaatgcgct catcgatc ctcggcaccg tcaccctgga tgctgtaggc 1620
 ataggcttgg ttatgccggt actgccgggc ctcttgccggg atatcgcca ttccgacagc 1680
 atcgccagtc actatggcgt gctgctagcg ctatatgcgt tgatgcaatt tctatgcgca 1740
 cccgttctcg gagcactgtc cgaccgcttt ggccgcggcc cagtctgct cgcttcgcta 1800
 cttggagcca ctatcgacta cgcgatcatg gcgaccacac cgtcctgtg gatcctctac 1860
 gccggacgca tcgtggccgg catcaccggc gccacaggtg cggttgctgg cgcctatatc 1920
 gccgacatca ccgatgggga agatcgggct cgcacttcg ggctcatgag cgcttgtttc 1980
 ggcgtgggta tgggtggcagg ccccggtggc gggggactgt tgggcgccat ctcttgcac 2040
 gcaccattcc ttgcggcggc ggtgctcaac ggctcaacc tactactggg ctgcttcta 2100
 atgcaggagt cgcataaggg agagcgatga ccgatgcct tgagagcctt caaccagtc 2160
 agctccttcc ggtgggcgcg gggcatgact atcgcgccg cacttatgac tgtcttctt 2220
 atcatgcaac tcgtaggaca ggtgccggca gcgctctggg tcattttcgg cgaggaccgc 2280
 ttctgctgga gcgcagcat gatcggcctg tcgcttgccg tattcggaat cttgcacgcc 2340
 ctcgctcaag ccttcgtcac tggccccgc accaaacgtt tggcgagaa gcaggccatt 2400
 atcgccggca tggcgccga cgcgctggg tacgtcttgc tggcgctcgc gacgcgaggc 2460
 tggatggcct tccccattat gattcttctc gcttcggcg gcacgggat gccgcgttg 2520
 caggccatgc tgtccaggca ggtagatgac gaccatcagg gacagcttca aggatcgctc 2580
 gcggctctta ccagcctaac ttcatcatt ggaccgctga tcgtcacggc gatttatgcc 2640
 gcctcgcgga gcacatggaa cgggttgga tggattgtag gcgcgcctt ataccttgc 2700
 tgctccccg cgttgctcg cgggtcatgg agccgggcca cctcgacctg aatggaagcc 2760
 ggcggcacct cgctaacgga ttcaccactc caagaattgg agccaatcaa ttcttgcgga 2820
 gaactgtgaa tgcgcaaacc aacccttggc agaacatata catcgctcc gccatctcca 2880
 gcagccgcac gcggcgcatc tcgggcagcg ttgggtcctg gccacgggtg cgcgatgatc 2940
 tgctcctgto gttgaggacc cggctaggct ggcggggttg ccttactggt tagcagaatg 3000
 aatcacggat acgcgagcga acgtgaagcg actgctgctg caaaacgtct gcgacctgag 3060
 caacaacatg aatggtcttc ggtttccgtg ttctgtaaag tctggaaacg cggaagtccc 3120
 ctactgctg ctgaagttgc ccgcaacaga gactggaacc aaccggtgat accacgatac 3180
 tatgactgag agtcaacgcc atgagcggcc tcattttctta ttctgagtta caacagtccg 3240
 caccgctgto cggtagctcc ttccggtggg cgcggggcat gactatcgtc gccgactta 3300
 tgactgtctt ctttatcatg caactcgtag gacaggtgcc ggcagcgccc aacagtcccc 3360
 cggccacggg gcctgccacc ataccacgc cgaaacaagc gccctgcacc attatgttcc 3420
 ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct gtattaacga 3480

agcgctaacc gtttttatca ggctctggga ggcagaataa atgatcatat cgtcaattat	3540
tacctccacg gggagagcct gagcaaaactg gcctcaggca tttgagaagc acacggtcac	3600
actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg ctattttaacg	3660
accctgccct gaaccgacga cggggtcgaa tttgctttcg aatttctgcc attcatccgc	3720
ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa ctgccttaaa	3780
aaaattacgc ccgcctcgc cactcatcgc agtactgttg taattcatta agcattctgc	3840
cgacatggaa gccatcacia acggcatgat gaacctgaat cgccagcggc atcagcacct	3900
tgctgccttg cgtataatat ttgccatgg atttaaattt aatctttctg cgaattgaga	3960
tgacgccact ggctgggcgt catcccgtt tcccgggtaa acaccaccga aaaatagtta	4020
ctatcttcaa agccacattc ggtcgaaata tcaactgatta acaggcggct atgctggaga	4080
agatattgag catgacacac tctgacctgt cgcagatatt gattgatggt cattccagtc	4140
tgctggcgaa attgctgacg caaaacgcgc tcaactgcag atgcctcacc acaaaattta	4200
tccagcgcaa agggactttt caggctagcc gccagccggg taatcagctt atccagcaac	4260
gtttcgtcgg atgttgccg caacgaatca ctggtgtaac gatggcgatt cagcaacatc	4320
accaactgcc cgaacagcaa ctcagccatt tcgttagcaa acggcacatg ctgactactt	4380
tcattgctcaa gctgaccgat aacctgccgc gcctgcgcca tccccatgct acctaaagcg	4440
cagtgtggtt gccctgcgct ggcgttaaata cccggaatcg cccctgcca gtcaagattc	4500
agcttcagac gctccgggca ataaataata ttctgcaaaa ccagatcggt aacggaagcg	4560
taggagtgtt tatcgtcagc atgaatgtaa aagagatcgc caccgggtaat gcgataaggg	4620
cgatcgttga gtacatgcag gccattaccg cgccagacaa tcaccagctc acaaaaatca	4680
tgtgtatgtt cagcaaaagc atcttgcgga taacggtcag ccacagcgac tgctgctgg	4740
tcgctggcaa aaaaatcacc tttgagaagt tttaactgat gcgccaccgt ggctacctcg	4800
gccagagaac gaagttgatt attcgcaata tggcgtaaaa atacgttgag aagattcgcg	4860
ttattgcaga aagccatccc gtccctggcg aatatcacgc ggtgaccagt taaactctcg	4920
gcgaaaaagc gtcgaaaagt ggttactgtc gctgaatcca cagcgatagg cgatgtcagt	4980
aacgctggcc tcgctgtggc gtagcagatg tcgggctttc atcagtcgca ggcggttcag	5040
gtatcgctga ggcgtcagtc ccgtttgctg cttaagctgc cgatgtagcg tacgcagtga	5100
aagagaaaat tgatccgcca cggcatccca attcacctca tcggcaaaat ggtcctccag	5160
ccaggccaga agcaagttga gacgtgatgc gctgttttcc aggttctcct gcaaaactgt	5220
tttacgcagc aagagcagta attgcataaa caagatctcg cgactggcgg tcgagggtaa	5280
atcattttcc ccttctgct gtccatctg tgcaaccagc tgcgcacct gctgcaatac	5340
gctgtgggta acgcgccagt gagacggata ctgccatcc agctcttggt gcagcaactg	5400
attcagcccg gcgagaaaact gaaatcgatc cggcgagcga tacagcacat tggtcagaca	5460
cagattatcg gtatgttcat acagatgccg atcatgatcg cgtacgaaac agaccgtgcc	5520
accggtgatg gtatagggt gccattaaa cacatgaata cccgtgccat gttcgacaat	5580
cacaatttca tgaaaaatcat gatgatgttc aggaaaatcc gcctgcggga gccgggggtc	5640

tatcgccacg gacgcgttac cagacggaaa aaaatccaca ctatgtaata cggtcatact	5700
ggcctcctga tgcgtcaac acggcgaaat agtaatcacg aggtcagggtt cttaccttaa	5760
attttcgacg gaaaaccacg taaaaaacgt cgatttttca agatacagcg tgaattttca	5820
ggaaatgcgg tgagcatcac atcaccacaa ttcagcaaatt tgtgaacatc atcacgttca	5880
tctttccctg gttgccaatg gcccattttc ctgtcagtaa cgagaaggtc gcgaattcag	5940
gcgcttttta gactggtcgt aatgaacatt taaatgaatt cccttgggac tctagagatc	6000
cgcggggggc caggaggggg gatctggcat ttttgggagg tgtgaaatga gtgagaagat	6060
cgtcaactcg tggaacgaat gggatgagct cgaggagatg gtggtcgga ttgcggacta	6120
tgccagcttc gaaccgaaag agccaggcaa ccatcccaa ctgcgcaacc agaactggc	6180
cgaaatcatc cccttcccaa gcggcccaa ggaccgaa gtgctggaga aagcgaacga	6240
agagctgaat gggctggctt acctgctgaa ggaccacgat gtgatcgtgc gtcgtccga	6300
gaagatcgac ttcaccaaga gcctgaaaac cccgtatttc gaggttgcca accagtactg	6360
cggcgtttgt cctcgcgacg tgatgatcac gtttggaac gaaatcatgg aagcgacct	6420
gtccaaacgt gcacgcttct tcgaatacct cccctatcgg aagctggtct acgagtactg	6480
gaacaaggac gagcacatga tctggaacgc agccccgaaa ccgacctgc aggatagcat	6540
gtacctggaa aacttctggg agctctcgct ggaagaacgc ttcaagcgga tgcacgactt	6600
cgaattctgc atcacccaag acgaggtgat cttcgatgcc gccgattgct cccgcttggg	6660
taaggacatc ctggtgcagg aaagcatgac caccaatcgc actggcatcc gctggctgaa	6720
gaagcatctc gaaccacgcg gctttcgcgt ccatccggtg cacttccgtt tggacttctt	6780
ccctagccac atcgactgca cgttcgtacc gttgcgtccg ggtctgatcc tgaccaatcc	6840
ggaacgcccg attcgcgagg aagaggagaa gatcttcaag gagaatggct gggagctgat	6900
caccgtaccg cagcctacct gctcgaaacg cgagatgccc atgttctgcc agagctcgaa	6960
atggctgtcc atgaacgtcc tgagcattag tcccaccaag gtgatctgcg aagaacggga	7020
aaagccgctg caagaactgc tggacaagca cgggttcgaa gtctttccct tgcctttccg	7080
ccatgtgttt gagttcgggtg gcagctttca ctgtgccacg tgggatattc gccgcaaggg	7140
cgagtgcgag gactacctgc cgaacctgaa ctaccagccg atttgcggct gacgcggatc	7200
ccgacctgca taccgttgca gataaggtag tcaactcatga gcgctaggca ctttctctcc	7260
ctgctggact tcaccaccga cgaattgctc ggggtgatcc gccacggcat cgagctgaag	7320
gacctgcgca agcgaggcgt gctgttcgaa ccgctgaaga accgtgtgct aggcattgatc	7380
ttcgaaaagt cctcgaccog tacccgtgtg tcgttcgagg ccggcatgat ccagctcggc	7440
ggccaggcca tcttctctgc gccccgcgac acccagctgg gccgcggcga gccaatgggt	7500
gacagcgcca tcgtgctgtc gagcatgggt gatgtgggtga tgatccggac ccacgccac	7560
agcaccctga ccgagttcgc cgccaagtgc cgtgtgcccg tgatcaacgg cctgtccgac	7620
gaatcgacac cgtgccaaact gctggccgac atgcagacct tcgttgaaca ccgcggctcg	7680
attcagggca agaccgtgac ctggatcggc gatggcttca acatgtgcaa ctccatatc	7740
gaagccgcca ggcagttcga tttccagctg cgcacgcct gccccgaagg ctatgagccg	7800

gatcaacgct tcatggcact gggcgggcgac cgcgtgcaga tcatccggga tgccagggaa	7860
gctgtgcgtg atgcacacct ggtggtcacc gatgtctgga cttccatggg tcaggaggag	7920
gaaactgcac ggcgctggc gcatttcgcg ccttaccagg tcaccgcga actgctcgac	7980
ctggctgcac ccgatgtcct cttcatgcac tgctgcccgc cccaccgtgg cgaggaaatc	8040
agccaggacc tgctcgacga cccacgttcg gtcgctggg acgaggctga aaaccgcctg	8100
catgcacaga aggcgcttct cgaattcctt gtagaaccgg cttaccacca cgcatgagtc	8160
aaccgtacaa ccccgaggag tgatggcatg gcggacgtaa aaaaggctcg actggcgtat	8220
tccggcggcc ttgatacttc ggtgattctc aagtggctgc aggataccta caactgcgaa	8280
gtggtgacct tcaccgctga cctggggcag ggcgaagagg tcgaaccggc ccgtgccaag	8340
gccaggcaa tggcggttaa agagatctac atcgacgacc tgcgcgaaga attcgtgcgt	8400
gatttcgtgt tcccgatgtt ccgcgccaac accgtctacg aaggcgagta cctgctgggt	8460
acttccatcg cccgtccgct gatcgccaag cgcctgatcg aaatcgccaa cgaaaccggc	8520
gctgacgcca tttcccatgg cgccaccggc aagggtaacg accagggtcg cttcgagctg	8580
ggtgcctatg cctgaagcc aggcgtcaag gtcacgctc catggcgcgga gtgggacctg	8640
ctgtcccgcg aaaagctgat ggactacgcc gagaagcacg gcacccgat cgagcgccac	8700
ggcaagaaga agtcgccga ctcgatggac gccaacctgc tgcacatctc ctacgagggc	8760
ggtgtcctgg aagatacctg gaccgagcac gaagaagaca tgtggcgctg gagtgtctcg	8820
cctgagaatg ccccggaaca ggctacctac atcgagctga cctaccgcaa tggtgacatc	8880
gttgccatcg acggcgctga gaaatccccg gccaccgtcc tggcagacct gaaccgtatc	8940
ggtggtgcca acggcatcgg ccgtctggac atcgctgaaa accgttacgt cggcatgaag	9000
tcgcgcggtt gctacgaaac gcctggcggt accatcatgc tcaaggcaca ccgtgccatc	9060
gagtcgatca cctggaccg cgaagtcgct cacctgaaag atgagctgat gccaaagtat	9120
gccagcctga tctacaccgg ctactggtgg agcccgagc gtctgatgct gcaacagatg	9180
atcgatgctt cgcaggtaaa cgtgaatggt gtggtgcgcc tgaaactgta caagggcaac	9240
gtgaccgtgg ttggccgcaa gtcggacgat tcgctgttcg atgccaacat cgccaccttt	9300
gaagaagatg gtggtgccta caaccaggca gatgtgctg gcttcatcaa gctcaatgca	9360
ctgcgtatgc gcattgccgc caacaagggc cgttcgctgc tctgattgct atcgacgcca	9420
ctttttcgtt cacgcctgca atgagtgaat ccatgagcac cgagaagacc aatcagtcct	9480
ggggcgggcg cttcagttag cccgtcgacg ccttcgtcgc ccgtttcacc gcctcggtag	9540
atttcgacaa gcgcctgtac cgtcacgaca tcatgggttc gattgcccat gccaccatgc	9600
tggcgcaggt cggcgctgctc agtcatgccc agcgcgacac catcatcgat ggctgaaaa	9660
ccatccaggc cgagattgaa gccggcaact tcgactggcg tgcgacctc gaagacgtgc	9720
acatgaacat cgaagcacgc ctgaccgacc gcacggcat caccggcaag aagctgcata	9780
ctgggcgtag ccgcaacgac caggtggcca ccgacatccg cctttggctg cgcgacgaaa	9840
tcgaacctgat cctgggcgaa atcaccgcc tgcagcaggc cctgctggag caggcagagc	9900
gtgaagccga aaccatcatg cctggtttca cccacctgca gacggcgag ccggtcacct	9960

ttggccacca cctgctggcg tggttcgaaa tgctcagccg cgactatgag cgcctggctg	10020
actgccgcaa ggcaccaac cgcatgccac tgggcagcgc cgcgctggcc ggcaccacct	10080
acccgatcga ccgtgaactg acctgcaagc tgctgggctt tgaagccgtg gccggcaact	10140
cgctggatgg cgtgtcggac cgtgatttcg ccatcgaatt ctgcgccgct gccagcgtgg	10200
cgatgatgca cctttcgcgc ttctccgaag agctggtgct gtggaccagc gcgcagtcc	10260
agttcatcga ccttcgggac cgcttctgca ctggcagctc gatcatgccg cagaaaaaga	10320
acccggacgt gccagagctg gtacgtggca agagcggccg cgtgttcggc gccctgaccg	10380
gcctgctgac cctgatgaaa ggccaaccgc tggcctacaa caaggacaac caggaagaca	10440
aggaaccgct gttcgacgcc gccgataccc tgcgcgactc gctgcggggc ttogctgaca	10500
tgatcccggc gatcaagccc aagcacgcca tcatgcgtga agcggccctg cgcggtttct	10560
ccaccgctac cgacctggct gactatctgg ttccgcgtgg cctgccgttc cgtgactgcc	10620
acgagatcgt tggccacgcg gtgaagtatg gtgtggacac tggcaaggac ctggccgaga	10680
tgagcctgga cgaactgcgc caattcagcg accagatcga gcaggacgtg tttgccgtgc	10740
tgacgctgga aggctcgggt aatgcgcgtg accacattgg tggtagcgcg ccggcgcagg	10800
tgcgtgctgc cgtcgttcgt ggcaaggccc tgttggcgct tcgctaatac cccaaggctc	10860
gagcacgcca gagtagggaa ctgccaggca tcaataaaaa cgaaaggctc agtcgaaaaga	10920
ctgggccttt cgttttatct gttgtttgtc ggtgaacgct ctctgagta ggacaaatcc	10980
gccgggagcg gatttgaacg atgataagct gtcaaacatg agaattcttg aagacgaaag	11040
ggcctcgtgt gtacaaacgt tcgtcaaaag ggcgacacaa aattcctgca ggggccggcc	11100
cagcgccggc ggtcgagtgg cgacggcgcg gcttgtccgc gccctggtag attgcctggc	11160
cgtaggccag ccatttttga gcggccagcg gcccgcatag gccgacgca agcggcgggg	11220
cgtagggagc gcagcgaccg aagggtaggc gctttttgca gctcttcggc tgtgcgctgg	11280
ccagacagtt atgcacaggc caggcgggtt ttaagagttt taataagttt taaagagttt	11340
taggcggaaa aatcgccctt tttctctttt atatcagtca cttacatgtg tgaccggttc	11400
ccaatgtacg gctttgggtt cccaatgtac gggttccggt tcccaatgta cggtcttggg	11460
ttoccaatgt acgtgctatc cacaggaaag agaccttttc gacctttttc ccctgctagg	11520
gcaatttgcc ctagcatctg ctccgtacat taggaaccgg cggatgcttc gccctcgatc	11580
aggttgcggt agcgcatgac taggatcggg ccagcctgcc ccgcctcctc cttcaaactg	11640
tactccggca ggtcatttga cccgatcagc ttgcgcacgg tgaacagaa cttcttgaac	11700
tctccggcgc tgccactgcg ttogtagatc gtcttgaaca accatctggc ttctgccttg	11760
cctgcggcgc ggcgtgccag gcggtagaga aaacggccga tgccgggac gatcaaaaag	11820
taatcggggt gaaccgtcag cacgtccggg ttcttgctt ctgtgatctc gcggtacatc	11880
caatcaacta gctcgatctc gatgtactcc ggccgccggg ttctgctctt tacgatcttg	11940
tagcggctaa tcaaggcttc accctcggat accgtcacca ggcggcggt cttggccttc	12000
ttcgtacgct gcatggcaac gtgcgtggtg tttaaccgaa tgcaggtttc taccaggctg	12060
tctttctgct ttccgccatc ggctcgccgg cagaacttga gtacgtccgc aacgtgtgga	12120

```

cggaacacgc ggccgggctt gtctcccttc ccttcccggg atcggttcat ggattcggtt 12180
agatgggaaa ccgccatcag taccaggtcg taatcccaca cactggccat gccggccggc 12240
cctgcggaaa cctctacgtg ccggtctgga agctcgtagc ggatcacctc gccagctcgt 12300
cggtcacgct tcgacagacg gaaaacggcc acgtccatga tgctgcgact atcgcggtg 12360
cccacgtcat agagcatcgg aacgaaaaaa tctgggtgct cgtcgccctt gggcggttc 12420
ctaategacg gcgcaccggc tgccggcggt tgccgggatt ctttgcggtat tcgatcagcg 12480
gccgcttgcc acgattcacc gggcggtgct tctgcctcga tgcgttgccg ctggcgggcc 12540
tgccgggctt tcaacttctc caccaggtca tccccagcg ccgcgcgat ttgtaccggg 12600
ccggtatggtt tgccaccgct cacgcgatt cctcggtt gggggttcca gtgccattgc 12660
agggccggca gacaaccag ccgcttacgc ctggccaacc gccggttctt ccacacatgg 12720
ggcattccac ggcgtcggtg cctggttggt cttgattttc catgcgcctt cctttagccg 12780
ctaaaattca tctactcatt tattcatttg ctcatctact ctggtagctg cgcgatgtat 12840
tcagatagca gctcggtaat ggtcttgctt tgccgtaccg cgtacatctt cagcttggtg 12900
tgatcctccg ccggcaactg aaagttgacc cgcttcatgg ctggcggtgc tgccaggctg 12960
gccaacgttg cagccttgct gctgcgtgcg ctccgacggc cggcacttag cgtgtttgtg 13020
cttttgctca ttttctcttt acctcattaa ctcaaatgag ttttgattta atttcagcgg 13080
ccagcgcttg gacctcgagg gcagcgctgc cctcggttc tgattcaaga acggttggtc 13140
cgccggcgcc agtgccctgg tagctcacgc gctgcgtgat acgggactca agaattggca 13200
gctcgtaccc ggccagcgcc tcggcaacct caccgccgat gcgcgtgctt ttgatcgccc 13260
gcgacacgac aaaggccgct tgtagccttc catccgtgac ctcaatgcgc tgcttaacca 13320
gtccaccag gtccggggtg gcccatatgt cgtaagggtt tggctgcacc ggaatcagca 13380
cgaagtcggc tgcccttgatc gcggacacag ccaagtccgc cgcctggggc gctccgtcga 13440
tcaactagaa gtccgcggcg ccgatggcct tcacgtcgcg gtcaatcgtc gggcggtcga 13500
tgccgacaac ggttagcggg tgatcttccc gcacggccgc ccaatcgcg gcactgcctt 13560
ggggatcgga atcgactaac agaacatcgg ccccgcgag ttgcaggcg cggttagat 13620
gggttgcatg gtcgtcttg cctgaccgc ctttctggtt aagtacagcg ataaccttca 13680
tgcggtcccc ttggtatctt gtttatttac tcacgcacatc atatacgag cgaccgatg 13740
acgcaagctg ttttactcaa atacacatca cttttttaga cggcgcgct cggtttcttc 13800
agcggccaag ctggccggcc aggcggccag cttggcatca gacaaaccgg ccaggatttc 13860
atgcagccgc acggttcggg atgagcattc a 13891

```

```

<210> 50
<211> 1137
<212> ADN
<213> Pseudomonas putida

```

```

<220>
<221> CDS
<222> (1) .. (1137)
<400> 50

```

ttg aca aag cca gcc ata ctc gcc ctt gcc gac ggc agt att ttc cgc Leu Thr Lys Pro Ala Ile Leu Ala Leu Ala Asp Gly Ser Ile Phe Arg 1 5 10 15	48
ggg gaa gcc atc ggt gcc gac ggt cag acc gtt ggt gag gtg gta ttc Gly Glu Ala Ile Gly Ala Asp Gly Gln Thr Val Gly Glu Val Val Phe 20 25 30	96
aac acc gct atg acc ggc tac cag gaa atc ctt aca gac cct tcc tac Asn Thr Ala Met Thr Gly Tyr Gln Glu Ile Leu Thr Asp Pro Ser Tyr 35 40 45	144
gcg cag caa atc gtt acc ctg acc tac ccg cac atc ggc aac acc ggt Ala Gln Gln Ile Val Thr Thr Tyr Pro His Ile Gly Asn Thr Gly 50 55 60	192
act acc ccg gaa gac gcc gag tcg agc cgc gtc tgg tcc gct ggc ctg Thr Thr Pro Glu Asp Ala Glu Ser Ser Arg Val Trp Ser Ala Gly Leu 65 70 75 80	240
gtc atc cgt gac ctg ccg ctg ctg gcc agc aac tgg cgt aac acc cag Val Ile Arg Asp Leu Pro Leu Leu Ala Ser Asn Trp Arg Asn Thr Gln 85 90 95	288
tcg ctg cct gag tac ctc aag gcc aac aac gtc gtc gcc atc gcc ggc Ser Leu Pro Glu Tyr Leu Lys Ala Asn Asn Val Val Ala Ile Ala Gly 100 105 110	336
atc gac acc cgt cgc ctg acc cgt atc ctg cgt gaa aag ggc gcc cag Ile Asp Thr Arg Arg Leu Thr Arg Ile Leu Arg Glu Lys Gly Ala Gln 115 120 125	384
aac ggc tgc att ctg gcg ggt gac aac atc agc gaa gaa gct gcc atc Asn Gly Cys Ile Leu Ala Gly Asp Asn Ile Ser Glu Glu Ala Ala Ile 130 135 140	432
gct gct gcc cgc ggc ttc ccg ggc ctg aag ggc atg gac ctg gcc aag Ala Ala Ala Arg Gly Phe Pro Gly Leu Lys Gly Met Asp Leu Ala Lys 145 150 155 160	480
gtc gtc tcc acc aag gaa cgt tac gag tgg cgc tcc agc gtg tgg gag Val Val Ser Thr Lys Glu Arg Tyr Glu Trp Arg Ser Ser Val Trp Glu 165 170 175	528
ctg aaa acc gac agc cac ccg acc atc gac gct gcc gac ctg ccg tac Leu Lys Thr Asp Ser His Pro Thr Ile Asp Ala Ala Asp Leu Pro Tyr 180 185 190	576
cac gtg gtt gcc ttc gac tat ggc gtc aag ctg aac atc ctg cgc atg His Val Val Ala Phe Asp Tyr Gly Val Lys Leu Asn Ile Leu Arg Met 195 200 205	624
ctg gtg gcc cgc ggc tgc cgc gtg acc gtg gta cca gcc cag acc ccg Leu Val Ala Arg Gly Cys Arg Val Thr Val Val Pro Ala Gln Thr Pro 210 215 220	672
gcc agc gaa gta ctg gca ctc aac ccg gac ggc gtg ttc ctg tcc aac Ala Ser Glu Val Leu Ala Leu Asn Pro Asp Gly Val Phe Leu Ser Asn 225 230 235 240	720
ggc cct ggt gac cct gag ccg tgc gac tac gcg atc cag gcg atc aag Gly Pro Gly Asp Pro Glu Pro Cys Asp Tyr Ala Ile Gln Ala Ile Lys 245 250 255	768
gaa atc ctc gaa acc gag atc ccg gta ttc ggc atc tgc ctc ggc cac Glu Ile Leu Glu Thr Glu Ile Pro Val Phe Gly Ile Cys Leu Gly His 260 265 270	816
cag ctg ctg gcc ctg gcg tcc ggc gcc aag acc gtg aaa atg ggc cac Gln Leu Leu Ala Leu Ala Ser Gly Ala Lys Thr Val Lys Met Gly His 275 280 285	864

ggc cac cac ggt gcc aac cac ccg gtc cag gac ctg gat act ggt gtg 912
 Gly His His Gly Ala Asn His Pro Val Gln Asp Leu Asp Thr Gly Val
 290 295 300
 gtc atg atc acc agc cag aac cac ggt ttc gcc gtt gac gag gcg acc 960
 Val Met Ile Thr Ser Gln Asn His Gly Phe Ala Val Asp Glu Ala Thr
 305 310 315 320
 ctg ccg ggc aac gtt cgc gcc att cac aag tcg ctg ttc gac ggc acc 1008
 Leu Pro Gly Asn Val Arg Ala Ile His Lys Ser Leu Phe Asp Gly Thr
 325 330 335
 ctg cag ggt atc gag cgt acc gac aag agc gcg ttc agc ttc cag ggc 1056
 Leu Gln Gly Ile Glu Arg Thr Asp Lys Ser Ala Phe Ser Phe Gln Gly
 340 345 350
 cac cct gaa gcg agc ccg gcc ccg acc gac gtc gcg cct ctg ttc gat 1104
 His Pro Glu Ala Ser Pro Gly Pro Thr Asp Val Ala Pro Leu Phe Asp
 355 360 365
 cgt ttc acc gat gcc atg gcc aag cgc cgc tga 1137
 Arg Phe Thr Asp Ala Met Ala Lys Arg Arg
 370 375

<210> 51
 <211> 378
 <212> PRT
 <213> Pseudomonas putida

<400> 51

Leu Thr Lys Pro Ala Ile Leu Ala Leu Ala Asp Gly Ser Ile Phe Arg
 1 5 10 15

Gly Glu Ala Ile Gly Ala Asp Gly Gln Thr Val Gly Glu Val Val Phe
 20 25 30

Asn Thr Ala Met Thr Gly Tyr Gln Glu Ile Leu Thr Asp Pro Ser Tyr
 35 40 45

Ala Gln Gln Ile Val Thr Leu Thr Tyr Pro His Ile Gly Asn Thr Gly
 50 55 60

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

Val Ile Arg Asp Leu Pro Leu Leu Ala Ser Asn Trp Arg Asn Thr Gln
 85 90 95

Ser Leu Pro Glu Tyr Leu Lys Ala Asn Asn Val Val Ala Ile Ala Gly
 100 105 110

Ile Asp Thr Arg Arg Leu Thr Arg Ile Leu Arg Glu Lys Gly Ala Gln
 115 120 125

Asn Gly Cys Ile Leu Ala Gly Asp Asn Ile Ser Glu Glu Ala Ala Ile
 130 135 140

Ala Ala Ala Arg Gly Phe Pro Gly Leu Lys Gly Met Asp Leu Ala Lys
 145 150 155 160

Val Val Ser Thr Lys Glu Arg Tyr Glu Trp Arg Ser Ser Val Trp Glu
165 170 175

Leu Lys Thr Asp Ser His Pro Thr Ile Asp Ala Ala Asp Leu Pro Tyr
180 185 190

His Val Val Ala Phe Asp Tyr Gly Val Lys Leu Asn Ile Leu Arg Met
195 200 205

Leu Val Ala Arg Gly Cys Arg Val Thr Val Val Pro Ala Gln Thr Pro
210 215 220

Ala Ser Glu Val Leu Ala Leu Asn Pro Asp Gly Val Phe Leu Ser Asn
225 230 235 240

Gly Pro Gly Asp Pro Glu Pro Cys Asp Tyr Ala Ile Gln Ala Ile Lys
245 250 255

Glu Ile Leu Glu Thr Glu Ile Pro Val Phe Gly Ile Cys Leu Gly His
260 265 270

Gln Leu Leu Ala Leu Ala Ser Gly Ala Lys Thr Val Lys Met Gly His
275 280 285

Gly His His Gly Ala Asn His Pro Val Gln Asp Leu Asp Thr Gly Val
290 295 300

Val Met Ile Thr Ser Gln Asn His Gly Phe Ala Val Asp Glu Ala Thr
305 310 315 320

Leu Pro Gly Asn Val Arg Ala Ile His Lys Ser Leu Phe Asp Gly Thr
325 330 335

Leu Gln Gly Ile Glu Arg Thr Asp Lys Ser Ala Phe Ser Phe Gln Gly
340 345 350

His Pro Glu Ala Ser Pro Gly Pro Thr Asp Val Ala Pro Leu Phe Asp
355 360 365

Arg Phe Thr Asp Ala Met Ala Lys Arg Arg
370 375

<210> 52
<211> 3222
<212> ADN
<213> Pseudomonas putida

<220>
<221> CDS
<222> (1)..(3222)

<400> 52
atg cca aaa cgt aca gac atc aaa agc atc ctg att ctc ggc gct ggc
Met Pro Lys Arg Thr Asp Ile Lys Ser Ile Leu Ile Leu Gly Ala Gly
1 5 10 15

48

ccg atc gtg atc ggc cag gcc tgc gaa ttc gac tac tcc ggc gcc cag

96

Pro Ile Val Ile Gly Gln Ala Cys Glu Phe Asp Tyr Ser Gly Ala Gln	
20 25 30	
gcc tgt aaa gcc ctg cgc gag gaa ggt ttc cgc gtc atc ctg gtg aac	144
Ala Cys Lys Ala Leu Arg Glu Glu Gly Phe Arg Val Ile Leu Val Asn	
35 40 45	
tcc aac cca gcc acc atc atg acc gac ccg gcc atg gct gac gcc acc	192
Ser Asn Pro Ala Thr Ile Met Thr Asp Pro Ala Met Ala Asp Ala Thr	
50 55 60	
tac atc gag ccg atc aag tgg caa tcg gtg gcc aag atc atc gag aaa	240
Tyr Ile Glu Pro Ile Lys Trp Gln Ser Val Ala Lys Ile Ile Glu Lys	
65 70 75 80	
gag cgc ccg gac gcc gtc ctg ccg acc atg ggt ggc cag acc gcc ctg	288
Glu Arg Pro Asp Ala Val Leu Pro Thr Met Gly Gly Gln Thr Ala Leu	
85 90 95	
aac tgc gcc ctg gac ctg gag cgc cac ggc gtt ctg gag aag ttc ggc	336
Asn Cys Ala Leu Asp Leu Glu Arg His Gly Val Leu Glu Lys Phe Gly	
100 105 110	
gtg gag atg atc ggt gcc aac gct gac acc atc gac aag gcc gaa gac	384
Val Glu Met Ile Gly Ala Asn Ala Asp Thr Ile Asp Lys Ala Glu Asp	
115 120 125	
cgt tcg cgc ttc gac aag gcc atg aag gac atc ggc ctg gag tgc ccg	432
Arg Ser Arg Phe Asp Lys Ala Met Lys Asp Ile Gly Leu Glu Cys Pro	
130 135 140	
cgc tcc ggt atc gcc cac agc atg gaa gag gcc aat gcg gtc ctc gag	480
Arg Ser Gly Ile Ala His Ser Met Glu Glu Ala Asn Ala Val Leu Glu	
145 150 155 160	
aag ctc ggc ttc ccg tgc atc att cgc ccg tcg ttc acc atg ggc ggc	528
Lys Leu Gly Phe Pro Cys Ile Ile Arg Pro Ser Phe Thr Met Gly Gly	
165 170 175	
acc ggc ggc ggt atc gct tac aac cgt gaa gag ttc gaa gaa atc tgc	576
Thr Gly Gly Gly Ile Ala Tyr Asn Arg Glu Glu Phe Glu Glu Ile Cys	
180 185 190	
acc cgt ggt ctg gac ctg tcg ccg acc aaa gag ctg ctg atc gac gaa	624
Thr Arg Gly Leu Asp Leu Ser Pro Thr Lys Glu Leu Leu Ile Asp Glu	
195 200 205	
tcg ctg atc ggc tgg aag gaa tac gag atg gag gtg gtc cgc gac aag	672
Ser Leu Ile Gly Trp Lys Glu Tyr Glu Met Glu Val Val Arg Asp Lys	
210 215 220	
aag gac aac tgc atc atc gtc tgc tcg atc gag aac ttc gac ccg atg	720
Lys Asp Asn Cys Ile Ile Val Cys Ser Ile Glu Asn Phe Asp Pro Met	
225 230 235 240	
ggt gtg cac acc ggt gac tcg atc act gtt gcc ccg gca cag acc ctg	768
Gly Val His Thr Gly Asp Ser Ile Thr Val Ala Pro Ala Gln Thr Leu	
245 250 255	
acc gac aag gaa tac cag atc atg cgc aac gcc tcg ctg gcg gtg ctg	816
Thr Asp Lys Glu Tyr Gln Ile Met Arg Asn Ala Ser Leu Ala Val Leu	
260 265 270	
cgt gaa atc ggt gtg gaa acc ggc ggt tcc aac gtc cag ttc ggc att	864
Arg Glu Ile Gly Val Glu Thr Gly Gly Ser Asn Val Gln Phe Gly Ile	
275 280 285	
tgc ccg aac acc ggc cgc atg gtt gtc atc gag atg aac ccg cgc gtg	912
Cys Pro Asn Thr Gly Arg Met Val Val Ile Glu Met Asn Pro Arg Val	
290 295 300	
tcg cgt tcg tcc gcc ctg gcc tcc aag gcc acc ggc ttc ccg atc gcc	960

Ser Arg Ser Ser Ala Leu Ala Ser Lys Ala Thr Gly Phe Pro Ile Ala	
305 310 315 320	
aag atc gcc gcc aag ctg gcc att ggt tac acc ctc gac gag ctg cag	1008
Lys Ile Ala Ala Lys Leu Ala Ile Gly Tyr Thr Leu Asp Glu Leu Gln	
325 330 335	
aac gac atc act ggc ggt cgc acc cca gcg tcc ttc gaa cct tcg atc	1056
Asn Asp Ile Thr Gly Gly Arg Thr Pro Ala Ser Phe Glu Pro Ser Ile	
340 345 350	
gac tac gtc gtc acc aag ctg cca cgc ttc gcc ttc gag aaa ttc ccg	1104
Asp Tyr Val Val Thr Lys Leu Pro Arg Phe Ala Phe Glu Lys Phe Pro	
355 360 365	
aaa gcc gac gcc cgc ctg acc acc cag atg aaa tcc gtg ggt gaa gtc	1152
Lys Ala Asp Ala Arg Leu Thr Thr Gln Met Lys Ser Val Gly Glu Val	
370 375 380	
atg gcc atc ggc cgt act ttc cag gaa tcc ctg cag aaa gcc ctg cgc	1200
Met Ala Ile Gly Arg Thr Phe Gln Glu Ser Leu Gln Lys Ala Leu Arg	
385 390 395 400	
ggc ctg gaa gtc ggc gcc tgc ggc ctc gac ccg aaa gtc gac ctg gcc	1248
Gly Leu Glu Val Gly Ala Cys Gly Leu Asp Pro Lys Val Asp Leu Ala	
405 410 415	
agc ccg gaa gcc gcc agc atc ctc aag cgc gaa ctg acc gtg ccg ggt	1296
Ser Pro Glu Ala Ala Ser Ile Leu Lys Arg Glu Leu Thr Val Pro Gly	
420 425 430	
gcc gag cgt atc tgg tac gtg gct gac gcc atg cgt tcg ggc atg acc	1344
Ala Glu Arg Ile Trp Tyr Val Ala Asp Ala Met Arg Ser Gly Met Thr	
435 440 445	
tgc gaa gaa atc ttc aat ctg acc ggc atc gac atg tgg ttc ctg gtg	1392
Cys Glu Glu Ile Phe Asn Leu Thr Gly Ile Asp Met Trp Phe Leu Val	
450 455 460	
cag atg gaa gac ctg atc aag gaa gaa gag aag gtc aag acc ctg gcc	1440
Gln Met Glu Asp Leu Ile Lys Glu Glu Glu Lys Val Lys Thr Leu Ala	
465 470 475 480	
ctg tcg gca atc gac aag gac tac atg ctg cgc ctc aag cgc aag ggc	1488
Leu Ser Ala Ile Asp Lys Asp Tyr Met Leu Arg Leu Lys Arg Lys Gly	
485 490 495	
ttc tcg gac cag cgc ctg gca gta ctg ctg ggt atc acc gac aag aac	1536
Phe Ser Asp Gln Arg Leu Ala Val Leu Leu Gly Ile Thr Asp Lys Asn	
500 505 510	
ctg cgt cgc cac cgc cac aag ctg gaa gtg ttc ccg gtg tac aag cgc	1584
Leu Arg Arg His Arg His Lys Leu Glu Val Phe Pro Val Tyr Lys Arg	
515 520 525	
gtc gac acc tgc gcc gcc gag ttc gcc acc gac acc gcc tac ctg tac	1632
Val Asp Thr Cys Ala Ala Glu Phe Ala Thr Asp Thr Ala Tyr Leu Tyr	
530 535 540	
tcc acc tac gag gaa gag tgc gag gcc aac ccg tcg acc cgc gac aag	1680
Ser Thr Tyr Glu Glu Glu Cys Glu Ala Asn Pro Ser Thr Arg Asp Lys	
545 550 555 560	
atc atg atc ctg ggt ggc ggc ccg aac cgt atc ggc caa ggt atc gag	1728
Ile Met Ile Leu Gly Gly Gly Pro Asn Arg Ile Gly Gln Gly Ile Glu	
565 570 575	
ttc gac tac tgc gta cac gcc gcc ctg gcg ctg cgt gaa gac ggt	1776
Phe Asp Tyr Cys Val His Ala Leu Ala Leu Arg Glu Asp Gly	
580 585 590	
tac gag acc atc atg gtc aac tgc aac ccg gaa acc gtc tcc acc gac	1824

Tyr Glu Thr Ile Met Val Asn Cys Asn Pro Glu Thr Val Ser Thr Asp	
595 600 605	
tac gac act tcc gac cgc ctg tac ttc gag ccg ctg acc ctg gaa gac	1872
Tyr Asp Thr Ser Asp Arg Leu Tyr Phe Glu Pro Leu Thr Leu Glu Asp	
610 615 620	
gtg ctg gaa gtc tgc cgc gtc gag aag ccg aag ggc gtg atc gtt cac	1920
Val Leu Glu Val Cys Arg Val Glu Lys Pro Lys Gly Val Ile Val His	
625 630 635 640	
tac ggc ggc cag acc ccg ctg aag ctg gcc cgc gct ctg gaa gag gct	1968
Tyr Gly Gly Gln Thr Pro Leu Lys Leu Ala Arg Ala Leu Glu Glu Ala	
645 650 655	
ggc gtg ccg atc atc ggt acc agc cct gac gcc atc gac cgc gcc gaa	2016
Gly Val Pro Ile Ile Gly Thr Ser Pro Asp Ala Ile Asp Arg Ala Glu	
660 665 670	
gac cgc gag cgc ttc cag cag atg gtt cag cgc ctg agc ctg ctg cag	2064
Asp Arg Glu Arg Phe Gln Gln Met Val Gln Arg Leu Ser Leu Leu Gln	
675 680 685	
ccg cca aac gcc acc gtg cgc agc gaa gaa gaa gcc atc cgt gct gcg	2112
Pro Pro Asn Ala Thr Val Arg Ser Glu Glu Glu Ala Ile Arg Ala Ala	
690 695 700	
ggc agc atc ggc tac ccg ctg gtc gtg cgt ccg tcc tac gta ctg ggc	2160
Gly Ser Ile Gly Tyr Pro Leu Val Val Arg Pro Ser Tyr Val Leu Gly	
705 710 715 720	
ggc cgt gcc atg gag atc gtc tac gag ctg gac gag ctc aag cgc tac	2208
Gly Arg Ala Met Glu Ile Val Tyr Glu Leu Asp Glu Leu Lys Arg Tyr	
725 730 735	
ctg cgt gaa gcg gta caa gtg tcg aac gac agc ccg gta ctg ctc gac	2256
Leu Arg Glu Ala Val Gln Val Ser Asn Asp Ser Pro Val Leu Leu Asp	
740 745 750	
cac ttc ctc aac tgc gcc atc gag atg gac gtg gat gcg gtg tgc gac	2304
His Phe Leu Asn Cys Ala Ile Glu Met Asp Val Asp Ala Val Cys Asp	
755 760 765	
ggc acc gac gtg gtc atc ggc gcg atc atg cag cac atc gag cag gcc	2352
Gly Thr Asp Val Val Ile Gly Ala Ile Met Gln His Ile Glu Gln Ala	
770 775 780	
ggc gta cac tcc ggc gac tcg gcg tgc tcg ctg cca cct tac tcg ctg	2400
Gly Val His Ser Gly Asp Ser Ala Cys Ser Leu Pro Pro Tyr Ser Leu	
785 790 795 800	
agc aag gaa gtg cag gac gaa gtc cgc gtt cag gtc aag aaa atg gcg	2448
Ser Lys Glu Val Gln Asp Glu Val Arg Val Gln Val Lys Lys Met Ala	
805 810 815	
ctg gag ctg ggt gta gtc ggc ctg atg aac gtg cag ctg gcc ctg cag	2496
Leu Glu Leu Gly Val Val Gly Leu Met Asn Val Gln Leu Ala Leu Gln	
820 825 830	
ggc gac aag atc tac gtg atc gaa gtc aac ccg cgt gcc tcg cgt acc	2544
Gly Asp Lys Ile Tyr Val Ile Glu Val Asn Pro Arg Ala Ser Arg Thr	
835 840 845	
gta ccg ttc gtg tcc aag tgc atc ggc acg tcc ctg gcg atg atc gca	2592
Val Pro Phe Val Ser Lys Cys Ile Gly Thr Ser Leu Ala Met Ile Ala	
850 855 860	
gcc cgt gtc atg gcg ggt aaa acc ctg aaa gag ctg ggc ttc acc cag	2640
Ala Arg Val Met Ala Gly Lys Thr Leu Lys Glu Leu Gly Phe Thr Gln	
865 870 875 880	
gaa atc atc ccg aac ttc tac agc gtg aag gaa gcc gtc ttc ccg ttc	2688

Glu Ile Ile Pro Asn Phe Tyr Ser Val Lys Glu Ala Val Phe Pro Phe
 885 890 895
 gcc aag ttc cca ggg gtt gac ccg atc ctc ggc cct gag atg aaa tcg 2736
 Ala Lys Phe Pro Gly Val Asp Pro Ile Leu Gly Pro Glu Met Lys Ser
 900 905 910
 acc ggt gaa gtg atg ggt gtc ggt gac agc ttc ggt gaa gcc ttc gcc 2784
 Thr Gly Glu Val Met Gly Val Gly Asp Ser Phe Gly Glu Ala Phe Ala
 915 920 925
 aaa gcc cag atg ggt gcc agc gaa gtg ctg ccg act ggc ggt acc gcg 2832
 Lys Ala Gln Met Gly Ala Ser Glu Val Leu Pro Thr Gly Gly Thr Ala
 930 935 940
 ttc atc agc gtg cgc gac gac gac aag cca caa gtg gcc ggc gtt gcc 2880
 Phe Ile Ser Val Arg Asp Asp Asp Lys Pro Gln Val Ala Gly Val Ala
 945 950 955 960
 cgc gac ctg atc gcc ctg ggc ttc gaa gtg gtt gcc act gcc ggc acc 2928
 Arg Asp Leu Ile Ala Leu Gly Phe Glu Val Val Ala Thr Ala Gly Thr
 965 970 975
 gcc aag gtt atc gag gcg gct ggc ctg aaa gtg cgc cgt gtg aac aag 2976
 Ala Lys Val Ile Glu Ala Ala Gly Leu Lys Val Arg Arg Val Asn Lys
 980 985 990
 gtg acc gaa ggt cgc cct cac gtg gtc gac atg atc aag aac gac gaa 3024
 Val Thr Glu Gly Arg Pro His Val Val Asp Met Ile Lys Asn Asp Glu
 995 1000 1005
 gtg tcg ctg atc atc aac acc acc gaa ggt cgc cag tcg atc gcc 3069
 Val Ser Leu Ile Ile Asn Thr Thr Glu Gly Arg Gln Ser Ile Ala
 1010 1015 1020
 gac tcc tac tcg att cgt cgc aat gcg ctg cag cac aag att tac 3114
 Asp Ser Tyr Ser Ile Arg Arg Asn Ala Leu Gln His Lys Ile Tyr
 1025 1030 1035
 tgc acc act acc att gcg gct ggt gaa gcc atc tgc gaa gcg ctg 3159
 Cys Thr Thr Thr Ile Ala Ala Gly Glu Ala Ile Cys Glu Ala Leu
 1040 1045 1050
 aaa ttc ggt ccg gaa aag acc gtt cgt cgc ttg cag gat ctg cat 3204
 Lys Phe Gly Pro Glu Lys Thr Val Arg Arg Leu Gln Asp Leu His
 1055 1060 1065
 gca gga ctg aaa gca tga 3222
 Ala Gly Leu Lys Ala
 1070

<210> 53
 <211> 1073
 <212> PRT
 <213> Pseudomonas putida

<400> 53

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

Pro Ile Val Ile Gly Gln Ala Cys Glu Phe Asp Tyr Ser Gly Ala Gln
 20 25 30

Ala Cys Lys Ala Leu Arg Glu Glu Gly Phe Arg Val Ile Leu Val Asn
 35 40 45

Ser Asn Pro Ala Thr Ile Met Thr Asp Pro Ala Met Ala Asp Ala Thr

50 55 60
 Tyr Ile Glu Pro Ile Lys Trp Gln Ser Val Ala Lys Ile Ile Glu Lys
 65 70 75 80
 Glu Arg Pro Asp Ala Val Leu Pro Thr Met Gly Gly Gln Thr Ala Leu
 85 90 95
 Asn Cys Ala Leu Asp Leu Glu Arg His Gly Val Leu Glu Lys Phe Gly
 100 105 110
 Val Glu Met Ile Gly Ala Asn Ala Asp Thr Ile Asp Lys Ala Glu Asp
 115 120 125
 Arg Ser Arg Phe Asp Lys Ala Met Lys Asp Ile Gly Leu Glu Cys Pro
 130 135 140
 Arg Ser Gly Ile Ala His Ser Met Glu Glu Ala Asn Ala Val Leu Glu
 145 150 155 160
 Lys Leu Gly Phe Pro Cys Ile Ile Arg Pro Ser Phe Thr Met Gly Gly
 165 170 175
 Thr Gly Gly Gly Ile Ala Tyr Asn Arg Glu Glu Phe Glu Glu Ile Cys
 180 185 190
 Thr Arg Gly Leu Asp Leu Ser Pro Thr Lys Glu Leu Leu Ile Asp Glu
 195 200 205
 Ser Leu Ile Gly Trp Lys Glu Tyr Glu Met Glu Val Val Arg Asp Lys
 210 215 220
 Lys Asp Asn Cys Ile Ile Val Cys Ser Ile Glu Asn Phe Asp Pro Met
 225 230 235 240
 Gly Val His Thr Gly Asp Ser Ile Thr Val Ala Pro Ala Gln Thr Leu
 245 250 255
 Thr Asp Lys Glu Tyr Gln Ile Met Arg Asn Ala Ser Leu Ala Val Leu
 260 265 270
 Arg Glu Ile Gly Val Glu Thr Gly Gly Ser Asn Val Gln Phe Gly Ile
 275 280 285
 Cys Pro Asn Thr Gly Arg Met Val Val Ile Glu Met Asn Pro Arg Val
 290 295 300
 Ser Arg Ser Ser Ala Leu Ala Ser Lys Ala Thr Gly Phe Pro Ile Ala
 305 310 315 320
 Lys Ile Ala Ala Lys Leu Ala Ile Gly Tyr Thr Leu Asp Glu Leu Gln
 325 330 335
 Asn Asp Ile Thr Gly Gly Arg Thr Pro Ala Ser Phe Glu Pro Ser Ile

340 345 350
 Asp Tyr Val Val Thr Lys Leu Pro Arg Phe Ala Phe Glu Lys Phe Pro
 355 360 365
 Lys Ala Asp Ala Arg Leu Thr Thr Gln Met Lys Ser Val Gly Glu Val
 370 375 380
 Met Ala Ile Gly Arg Thr Phe Gln Glu Ser Leu Gln Lys Ala Leu Arg
 385 390 395 400
 Gly Leu Glu Val Gly Ala Cys Gly Leu Asp Pro Lys Val Asp Leu Ala
 405 410 415
 Ser Pro Glu Ala Ala Ser Ile Leu Lys Arg Glu Leu Thr Val Pro Gly
 420 425 430
 Ala Glu Arg Ile Trp Tyr Val Ala Asp Ala Met Arg Ser Gly Met Thr
 435 440 445
 Cys Glu Glu Ile Phe Asn Leu Thr Gly Ile Asp Met Trp Phe Leu Val
 450 455 460
 Gln Met Glu Asp Leu Ile Lys Glu Glu Glu Lys Val Lys Thr Leu Ala
 465 470 475 480
 Leu Ser Ala Ile Asp Lys Asp Tyr Met Leu Arg Leu Lys Arg Lys Gly
 485 490 495
 Phe Ser Asp Gln Arg Leu Ala Val Leu Leu Gly Ile Thr Asp Lys Asn
 500 505 510
 Leu Arg Arg His Arg His Lys Leu Glu Val Phe Pro Val Tyr Lys Arg
 515 520 525
 Val Asp Thr Cys Ala Ala Glu Phe Ala Thr Asp Thr Ala Tyr Leu Tyr
 530 535 540
 Ser Thr Tyr Glu Glu Glu Cys Glu Ala Asn Pro Ser Thr Arg Asp Lys
 545 550 555 560
 Ile Met Ile Leu Gly Gly Gly Pro Asn Arg Ile Gly Gln Gly Ile Glu
 565 570 575
 Phe Asp Tyr Cys Cys Val His Ala Ala Leu Ala Leu Arg Glu Asp Gly
 580 585 590
 Tyr Glu Thr Ile Met Val Asn Cys Asn Pro Glu Thr Val Ser Thr Asp
 595 600 605
 Tyr Asp Thr Ser Asp Arg Leu Tyr Phe Glu Pro Leu Thr Leu Glu Asp
 610 615 620
 Val Leu Glu Val Cys Arg Val Glu Lys Pro Lys Gly Val Ile Val His

```

625                630                635                640

Tyr Gly Gly Gln Thr Pro Leu Lys Leu Ala Arg Ala Leu Glu Glu Ala
      645                650                655

Gly Val Pro Ile Ile Gly Thr Ser Pro Asp Ala Ile Asp Arg Ala Glu
      660                665                670

Asp Arg Glu Arg Phe Gln Gln Met Val Gln Arg Leu Ser Leu Leu Gln
      675                680                685

Pro Pro Asn Ala Thr Val Arg Ser Glu Glu Glu Ala Ile Arg Ala Ala
      690                695                700

Gly Ser Ile Gly Tyr Pro Leu Val Val Arg Pro Ser Tyr Val Leu Gly
705                710                715                720

Gly Arg Ala Met Glu Ile Val Tyr Glu Leu Asp Glu Leu Lys Arg Tyr
      725                730                735

Leu Arg Glu Ala Val Gln Val Ser Asn Asp Ser Pro Val Leu Leu Asp
      740                745                750

His Phe Leu Asn Cys Ala Ile Glu Met Asp Val Asp Ala Val Cys Asp
      755                760                765

Gly Thr Asp Val Val Ile Gly Ala Ile Met Gln His Ile Glu Gln Ala
      770                775                780

Gly Val His Ser Gly Asp Ser Ala Cys Ser Leu Pro Pro Tyr Ser Leu
785                790                795                800

Ser Lys Glu Val Gln Asp Glu Val Arg Val Gln Val Lys Lys Met Ala
      805                810                815

Leu Glu Leu Gly Val Val Gly Leu Met Asn Val Gln Leu Ala Leu Gln
      820                825                830

Gly Asp Lys Ile Tyr Val Ile Glu Val Asn Pro Arg Ala Ser Arg Thr
      835                840                845

Val Pro Phe Val Ser Lys Cys Ile Gly Thr Ser Leu Ala Met Ile Ala
      850                855                860

Ala Arg Val Met Ala Gly Lys Thr Leu Lys Glu Leu Gly Phe Thr Gln
865                870                875                880

Glu Ile Ile Pro Asn Phe Tyr Ser Val Lys Glu Ala Val Phe Pro Phe
      885                890                895

Ala Lys Phe Pro Gly Val Asp Pro Ile Leu Gly Pro Glu Met Lys Ser
      900                905                910

Thr Gly Glu Val Met Gly Val Gly Asp Ser Phe Gly Glu Ala Phe Ala

```

915	920	925	
Lys Ala Gln Met Gly Ala Ser Glu Val Leu Pro Thr Gly Gly Thr Ala			
930	935	940	
Phe Ile Ser Val Arg Asp Asp Lys Pro Gln Val Ala Gly Val Ala			
945	950	955	960
Arg Asp Leu Ile Ala Leu Gly Phe Glu Val Val Ala Thr Ala Gly Thr			
	965	970	975
Ala Lys Val Ile Glu Ala Ala Gly Leu Lys Val Arg Arg Val Asn Lys			
	980	985	990
Val Thr Glu Gly Arg Pro His Val Val Asp Met Ile Lys Asn Asp Glu			
	995	1000	1005
Val Ser Leu Ile Ile Asn Thr Thr Glu Gly Arg Gln Ser Ile Ala			
	1010	1015	1020
Asp Ser Tyr Ser Ile Arg Arg Asn Ala Leu Gln His Lys Ile Tyr			
	1025	1030	1035
Cys Thr Thr Thr Ile Ala Ala Gly Glu Ala Ile Cys Glu Ala Leu			
	1040	1045	1050
Lys Phe Gly Pro Glu Lys Thr Val Arg Arg Leu Gln Asp Leu His			
	1055	1060	1065
Ala Gly Leu Lys Ala			
	1070		
<210>	54		
<211>	44		
<212>	ADN		
<213>	oligonucleotit tổng hợp		
<400>	54		
agagcctgag caaactggcc gactgatgtg gtgttggtt cttc			44
<210>	55		
<211>	45		
<212>	ADN		
<213>	oligonucleotit tổng hợp		
<400>	55		
ccgtgtgctt ctcaaagcc tatttggtta tgctcatgct ttcag			45
<210>	56		
<211>	4841		
<212>	ADN		
<213>	Pseudomonas putida		
<400>	56		
agagcctgag caaactggcc gactgatgtg gtgttggtt cttcgcgggc ttgcccgctc			60
ccacaggtgc ggcgttgctc ttgggctggt gcgatccctg tgggggcgca ccagcccgcg			120
cagaagccaa cacagatgtc agtctgtcgc attgatgtgt tttagagaca catcccggt			180

agaccgaaaa cacacttttc tgtaagctac agcttttagtg tgtccactaa aagcgcgcag 240
 cgaattgaat tcagcacatg aaagcggggg gacgtgtcca tacgtcactc cgcttttttg 300
 caacctgcga tcgccctttc atgcttgatt tacgggaggt cttcttgaca aagccagcca 360
 tactcgccct tgccgacggc agtattttcc gcggtgaagc catcggtgcc gacggtcaga 420
 ccgttggtga ggtggtattc aacaccgcta tgaccggcta ccaggaaatc cttacagacc 480
 cttcctacgc gcagcaaatc gttaccctga cctaccgca catcggcaac accggtacta 540
 ccccggaaga cgccgagtcg agccgcgtct ggtccgctgg cctggtcac cgtgacctgc 600
 cgctgctggc cagcaactgg cgtaacaccc agtcgctgcc tgagtacctc aaggccaaca 660
 acgtcgtgc catcgccggc atcgacaccc gtcgectgac ccgtatcctg cgtgaaaagg 720
 gcgcccagaa cggtgcatt ctggcgggtg acaacatcag cgaagaagct gccatcgtg 780
 ctgcccggg cttcccgggc ctgaagggca tggacctggc caaggctgct tccaccaagg 840
 aacgttacga gtggcgctcc agcgtgtggg agctgaaaac cgacagccac ccgaccatcg 900
 acgtcgccga cctgcgctac cacgtggttg ccttcgacta tggcgtcaag ctgaacatcc 960
 tgcgcatgct ggtggccgc ggctgccgcg tgacctggt accagcccag accccggcca 1020
 gcgaagtact ggcaactaac ccggacggcg tgttctgtc caacggccct ggtgacctg 1080
 agccgtgcga ctacgcgac caggcgatca aggaatcct cgaaaccgag atcccgtat 1140
 tcggcatctg cctcgccac cagctgctgg ccctggcgtc cgcgccaaag accgtgaaaa 1200
 tgggccacgg ccaccacggt gccaacacc cggccagga cctggatact ggtgtgttca 1260
 tgatcaccag ccagaaccac ggtttcgccg ttgacgaggc gacctgccg ggcaacgttc 1320
 gcgccattca caagtgcgtg ttgacggca ccctgcaggg tatcgagcgt accgacaaga 1380
 gcgcgttcag cttccagggc caccctgaag cgagccggg cccgaccgac gtcgcgcctc 1440
 tgttcgatcg tttcaccgat gccatggcca agcgcgctg agcatcctgc ttcaaggccc 1500
 cgggcggcca cctgcgcgc cggcagcgc ctgaccaga ttgttcaaag cggcttgccg 1560
 actgaccccg gatattgagtg accaccatgc caaacgtac agacatcaa agcatcctga 1620
 ttctcggcgc tggcccgatc gtgatcggcc aggcctgcga attcgactac tccggcgccc 1680
 aggcctgtaa agccctgcgc gaggaagggt tccgcgtcat cctggtgaac tccaaccag 1740
 ccaccatcat gaccgaccg gccatggctg acgccaccta catcgagccg atcaagtggc 1800
 aatcgggtgc caagatcatc gagaaagagc gcccgagcgc cgtcctgccg accatgggtg 1860
 gccagaccgc cctgaactgc gccctggacc tggagcgcca cggcgttctg gagaagtctg 1920
 gcgtggagat gatcggtgcc aacgtgaca ccatcgacaa ggccgaagac cgttcgcgct 1980
 tcgacaaggc catgaaggac atcggcctgg agtgcccgcg ctccggtatc gccacagca 2040
 tggaagaggc caatgcggtc ctcgagaagc tcggcttccc gtgcatcatt cgcccgctgt 2100
 tcaccatggg cggcaccggc ggcggtatcg cttacaaccg tgaagagttc gaagaaatct 2160
 gcacccgtgg tctggacctg tcgcccacca aagagctgct gatcgacgaa tcgctgatcg 2220
 gctggaagga atacgagatg gaggtggtcc gcgacaagaa ggacaactgc atcatcgtct 2280
 gctcgatcga gaacttcgac ccgatgggtg tgcacaccgg tgactcgatc actgttgccc 2340

cggcacagac cctgaccgac aaggaatacc agatcatgcg caacgcctcg ctggcggtgc	2400
tgcgtgaaat cggtgtggaa accggcggtt ccaacgtcca gttcggcatt tgcccgaaca	2460
ccggccgcat ggtgtgcatc gagatgaacc cgcgcgtgtc gcgttcgtcc gccctggcct	2520
ccaaggccac cggttccccg atcgccaaga tcgccgcaa gctggccatt ggttacccc	2580
tcgacgagct gcagaacgac atcaactggcg gtgcacccc agcgtccttc gaaccttcga	2640
tcgactacgt cgtcaccaag ctgccacgct tcgccttcga gaaattcccc aaagccgacg	2700
ccgcctgac caccagatg aaatccgtgg gtgaagtcac ggccatcggc cgtactttcc	2760
aggaatccct gcagaaagcc ctgcgcggcc tggaagtcgg cgctgcggc ctcgaccgga	2820
aagtcgacct ggccagcccc gaagccgcca gcatcctcaa gcgcgaactg accgtgccgg	2880
gtgcgagcg tatctggtac gtggctgacg ccctgcgttc gggcatgacc tgcgaagaaa	2940
tcttcaatct gaccggcatc gacatgtggt tcctggtgca gatggaagac ctgatcaagg	3000
aagaagagaa ggtcaagacc ctggccctgt cggcaatcga caaggactac atgctgcgcc	3060
tcaagcgcaa gggcttctcg gaccagcgcc tggcagtact gctgggtatc accgacaaga	3120
acctgcgtcg ccaccgccac aagctggaag tgttcccgtt gtacaagcgc gtcgacacct	3180
gcgccgcga gttagccacc gacaccgct acctgtactc cacctacgag gaagagtgcg	3240
aggccaaccc gtcgaccgcg gacaagatca tgatcctggg tggcgccccg aaccgtatcg	3300
gccaaggtat cgagttcgac tactgtcgcg tacacgccgc cctggcgctg cgtgaagacg	3360
gttacgagac catcatggtc aactgcaacc cggaaaccgt ctccaccgac tacgacactt	3420
ccgaccgcct gtacttcgag ccgctgacct tggaagacgt gctggaagtc tgccgcgtcg	3480
agaagccgaa gggcggtgac gttcactacg gcggccagac cccgctgaag ctggcccgcg	3540
ctctggaaga ggctggcggt cggatcatcg gtaccagccc tgacgccatc gaccgcgccg	3600
aagaccgca gcgcttccag cagatggttc agcgcctgag cctgctgcag ccgccaacg	3660
ccaccgtgcg cagcgaagaa gaagccatcc gtgctgcggg cagcatcggc taccgctgg	3720
tcgtgcgtcc gtcctacgta ctgggcggcc gtgccatgga gatcgtctac gagctggacg	3780
agctcaagcg ctacctgctg gaagcggtag aagtgtcgaa cgacagcccc gtactgctcg	3840
accacttcct caactgcgcc atcgagatgg acgtggatgc ggtgtgcgac ggcaccgacg	3900
tggtcatcgg cgcgatcatg cagcacatcg agcaggccgg cgtacactcc ggcgactcgg	3960
cgtgctcgct gccaccttac tcgctgagca aggaagtga ggacgaagtc cgcgttcagg	4020
tcaagaaaat ggcgctggag ctgggtgtag tcggcctgat gaacgtgcag ctggccctgc	4080
agggcgacaa gatctacgtg atcgaagtca acccgcgctg ctcgcgtacc gtaccgttcg	4140
tgtccaagtg catcggcacg tccttggcga tgatcgacgc ccgtgtcatg ggggtaaaa	4200
ccctgaaaga gctgggcttc acccaggaaa tcatcccga cttctacagc gtgaaggaa	4260
ccgtcttccc gttcgccaag tccccaggg ttgacccgat cctcgccctc gagatgaaat	4320
cgaccggtag agtgatgggt gtcggtgaca gttcgggtga agccttcgcc aaagcccaga	4380
tgggtgccag cgaagtgcgt ccgactggcg gtaccgcgtt catcagcgtg cgcgacgacg	4440
acaagccaca agtgcccgcc gttgcccgcg acctgatcgc cctgggcttc gaagtggttg	4500

ccactgccgg caccgccaag gttatcgagg cggctggcct gaaagtgcgc cgtgtgaaca 4560
 aggtgaccga aggtcgccct cacgtggctg acatgatcaa gaacgacgaa gtgtcgctga 4620
 tcatcaacac caccgaaggt cgccagtcga tcgccgactc ctactcgatt cgtcgcaatg 4680
 cgctgcagca caagatttac tgcaccacta ccattgcggc tggatgaagc atctgcgaag 4740
 cgctgaaatt cgttcgggaa aagaccgttc gtcgcttgca ggatctgcat gcaggactga 4800
 aagcatgagc attaccaaat aggcatttga gaagcacacg g 4841

<210> 57
 <211> 18689
 <212> ADN
 <213> khác

<400> 57
 tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg 60
 tcttttaaaaa ggccgtaata tccagctgaa cggctctggt ataggatcat tgagcaactg 120
 actgaaatgc ctcaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc 180
 cagtgatattt tttctccatt ttagcttctt tagctctga aaatctcgat aactcaaaaa 240
 atacgcccgg tagtgatctt atttcattat ggtgaaagt ggaacctctt acgtgccgat 300
 caacgtctca ttttcgcaa aagttggccc agggcttccc ggtatcaaca gggacaccag 360
 gatttatatta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt 420
 cgggtgatgc tgccaacta ctgatttagt gtatgatggt gtttttgagg tgctccagt 480
 gcttctgttt ctatcagctg tccctcctgt tcagctactg acgggggtgt gcgtaacggc 540
 aaaagcaccg ccggacatca gcgctagcgg agtgataact ggcttactat gttggcactg 600
 atgaggggtg cagtgaagt cttcatgtgg caggagaaaa aaggctgcac cgggtgcgtca 660
 gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt 720
 cgttcgactg cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc 780
 caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggctc 840
 cgccccctg acaagcatca cgaaatctga cgctcaaata agtggtggcg aaaccgaca 900
 ggactataaa gataccaggc gtttccccct ggcggctccc tcgtcgctc tctgttctt 960
 gcctttcggt ttaccggtgt cattccgctg ttatggcgcg gtttgtctca ttccacgcct 1020
 gacactcagt tcgggtagg cagttcgctc caagctggac tgtatgcacg aacccccctg 1080
 tcagtccgac cgctgcgctt tatccgtaa ctatcgtctt gagtccaacc cggaaagaca 1140
 tgcaaaagca cactggcag cagccactgg taattgattt agaggagtta gtcttgaagt 1200
 catgcgccg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc 1260
 agttacctcg gttcaaagag ttggtagctc agagaacct cgaaaaaccg cctgcaagg 1320
 cggtttttct gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc 1380
 atcttattaa tcagataaaa ttttctaga tttcagtga atttatctct tcaaatgtag 1440
 cacctgaagt cagccccata cgatataagt tgtaattctc atgtttgaca gcttatcatc 1500
 gataagcttt aatgcggtag tttatcacag ttaaattgct aacgcagtca ggcaccgtgt 1560

atgaaatcta acaatgcgct catogtcac	1620
ctcggcaccg tcaccctgga tgctgtaggc	
ataggettgg ttatgccggt actgccgggc	1680
ctcttgccggg atatcgtcca ttccgacagc	
atcgccagtc actatggcgt gctgctagcg	1740
ctatatgcgt tgatgcaatt tctatgcgca	
cccgttctcg gagcactgtc cgaccgcttt	1800
ggcgcgcc cagtcctgct cgcttcgcta	
cttggagcca ctatcgacta cgcgatcatg	1860
gcgaccacac ccgtcctgtg gatcctctac	
gccggacgca tcgtggccgg catcacccggc	1920
gccacaggtg cggttgctgg cgcctatata	
gccgacatca ccgatgggga agatcgggct	1980
cgcacttcg ggctcatgag cgcttgtttc	
ggcgtgggta tgggtggcagg ccccggtggc	2040
gggggactgt tgggcgccat ctccttgcat	
gcaccattcc ttgcggcggc ggtgctcaac	2100
ggcctcaacc tactactggg ctgcttccta	
atgcaggagt cgcataaggg agagcgtcga	2160
ccgatgccct tgagagcctt caaccagtc	
agctccttcc ggtgggcgcg gggcatgact	2220
atcgtcgcg cacttatgac tgtcttcttt	
atcatgcaac tcgtaggaca ggtgccggca	2280
gcgctctggg tcattttcgg cgaggaccgc	
tttcgctgga gcgcgacgat gatcggcctg	2340
tcgcttgccg tattcggaat cttgcacgcc	
ctcgtcaag ccttcgtcac tggccccgc	2400
accaaactgt tcggcgagaa gcaggccatt	
atcgccggca tggcgccga cgcgctgggc	2460
tacgtcttgc tggcgcttcgc gacgcgaggc	
tggatggcct tccccattat gattcttctc	2520
gcttcggcg gcacgggat gcccgcttg	
caggccatgc tgtccaggca ggtagatgac	2580
gaccatcagg gacagcttca aggatcgctc	
gcggctctta ccagcctaac ttcgatcatt	2640
ggaccgctga tcgtcacggc gatttatgcc	
gcctcggcga gcacatggaa cgggttgga	2700
tggattgtag gcgcgcct ataccttgc	
tgctccccg cgttgcgctg cgggtgcatg	2760
agccgggcca cctcgacctg aatggaagcc	
ggcgccacct cgctaaccga ttcaccactc	2820
caagaattgg agccaatcaa ttcttgcgga	
gaactgtgaa tgcgcaaacc aacccttggc	2880
agaacatata catcgctcc gccatctcca	
gcagccgac gcggcgcatc tcgggcagcg	2940
ttgggtcctg gccacgggtg cgcgatgctg	
tgctcctgtc gttgaggacc cggctaggct	3000
ggcggggttg ccttactggt tagcagaatg	
aatcaccgat acgcgagcga acgtgaagcg	3060
actgctgctg caaaaactct gcgacctgag	
caacaacatg aatggctctt ggtttccgtg	3120
tttcgtaaag tctggaaacg cggaagtccc	
ctacgtgctg ctgaagttgc ccgcaacaga	3180
gagtggaaac aaccggtgat accacgatac	
tatgactgag agtcaacgcc atgagcggcc	3240
tcatttctta ttctgagtta caacagtccg	
caccgctgtc cggtagctcc ttccggtggg	3300
cgcggggcat gactatcgtc gccgcactta	
tgactgtctt ctttatcatg caactcgtag	3360
gacaggtgcc ggcagcgccc aacagtcccc	
cggccacggg gcctgccacc ataccacgc	3420
cgaacaagc gccctgcacc attatgttcc	
ggatctgcat cgcaggatgc tgcgtgctac	3480
cctgtggaac acctacatct gtattaacga	
agcgetaacc gtttttatca ggtcttgga	3540
ggcagaataa atgatcatat cgtcaattat	
tacctocacg gggagagcct gagcaaaactg	3600
gccgactgat gtggtgttgg cttcttcgcg	
ggcttgccc ctcccacagg tgcggcggtg	3660
ctcttgggct ggtgcgatcc ctgtgggggc	
gcaccagccc gcgcagaagc caacacagat	3720
gtcagtcctgt cgcattgatg tgttttagag	

acacatccgc ggtagaccga aaacacactt ttctgtaagc tacagcttta gtgtgtccac	3780
taaaagcgcg cagcgaattg aattcagcac atgaaagcgg ggtgacgtgt ccatacgtca	3840
ctccgctttt ttgcaacctg cgatcgccct ttcatgcttg atttacggga ggtcttcttg	3900
acaaagccag ccatactcgc ccttgccgac ggcagtattt tccgcggtga agccatcggc	3960
gccgacggtc agaccgttgg tgaggtggtg ttcaacaccg ctatgaccgg ctaccaggaa	4020
atccttacag acccttccta cgcgcagcaa atcgttaccc tgacctaccc gcacatcggc	4080
aacaccggta ctaccccgga agacgccgag tcgagccgcg tctggtccgc tggcctggtc	4140
atccgtgacc tgccgctgct ggccagcaac tggcgtaaca cccagtcgct gcctgagtac	4200
ctcaaggcca acaacgtcgt cgccatcgcc ggcacgaca cccgtcgct gaccgtatc	4260
ctgctgaaa agggcgccca gaacggctgc attctggcgg gtgacaacat cagcgaagaa	4320
gctgccatcg ctgctgcccg cggttcccg ggctgaagg gcatggacct ggccaaggtc	4380
gtctccacca aggaacgtta cgagtggcgc tccagcgtgt gggagctgaa aaccgacagc	4440
caccgacca tcgacgtgc cgacctgcg taccagtggt ttgccttcga ctatggcgtc	4500
aagctgaaca tcctgcgcat gctggtggcc cgcggctgcc gcgtgacct ggtaccagcc	4560
cagaccccg ccagcgaagt actggcactc aaccggagc gcgtgttcc gtccaacggc	4620
cctggtgacc ctgagccgtg cgactacgcy atccaggcga tcaaggaaat cctcgaaacc	4680
gagatcccg tattcgcat ctgcctcggc caccagctgc tggccctggc gtccggcgcc	4740
aagaccgtga aaatgggcca cggccaccac ggtgccaacc acccggtcca ggacctggat	4800
actggtgtgg tcatgatcac cagccagaac cacggtttcg ccgttgacga ggcgacctg	4860
ccgggcaacg ttccgcgcat tcacaagtcg ctgttcgacg gcacctgca ggtatcgag	4920
cgtaccgaca agagcgctt cagcttcag ggccaccctg aagcgagccc gggcccgacc	4980
gacgtcgcgc ctctgttcga tcgtttcacc gatgccatgg ccaagcgccg ctgagcatcc	5040
tgttcaagg ccccgggcgc gcacctgcg cgcggcgag cgctgacct agattgttca	5100
aagcggcttg ccgactgacc ccggatttga gtgaccacca tgccaaaacg tacagacatc	5160
aaaagcatcc tgattctcgg cgctggccc atcgtgatcg gccaggcctg cgaattcgac	5220
tactccggcg ccaggcctg taaagccctg cgcgaggaag gtttcgcgt catcctggtg	5280
aactccaacc cagccaccat catgaccgac ccggccatgg ctgacgccac ctacatcgag	5340
ccgatcaagt ggcaatcggg ggccaagatc atcgagaaag agcgcccgga cgcgctcctg	5400
ccgacctgg gtggccagac cgccctgaac tgcgcccctg acctggagcg ccacggcggt	5460
ctggagaagt tcggcgtgga gatgateggg gccaacgctg acaccatcga caaggccgaa	5520
gaccgttcgc gtttcgacaa ggccatgaag gacatcgccc tggagtgcgc gcgctccggt	5580
atcgcccaca gcatggaaga ggccaatgcg gtcctcgaga agctcggtt cccgtgcac	5640
attcgccctg cgttcaccat gggcggcacc ggcggcggtg tcgcttacia cgtgaagag	5700
ttcgaagaaa tctgcaccog tggctctggac ctgtcgcca ccaaagagct gctgatcgac	5760
gaatcgctga tcggctggaa ggaatacgag atggaggtgg tccgcgacaa gaaggacaac	5820
tgcacatcg tctgctgat cgagaacttc gaccgatgg gtgtgcacac cggtgactcg	5880

atcactgttg ccccggcaca gaccctgacc gacaaggaat accagatcat ggcgaacgcc	5940
tcgtggcg tgctgcgtga aatcgggtg gaaaccggcg gttccaacgt ccagttcggc	6000
atttccccga acaccggcgg catgggtgtc atcgagatga acccgcgcggt gtcgcgttcg	6060
tccgccctgg cctccaaggc caccggcttc ccgatcgcca agatcgccgc caagctggcc	6120
attggttaca cctcgcagca gctgcagaac gacatcactg gcggtcgcac ccagcgtcc	6180
ttcgaacctt ccatcgacta cgtcgtcacc aagctgccac gcttcgcctt cgagaaattc	6240
ccgaaagccg acgcccgcct gaccaccag atgaaatccg tgggtgaagt catggccatc	6300
ggcgtactt tccaggaatc cctgcagaaa gccctgcgcg gcttgggaagt cggcgctgc	6360
ggcctcgacc cgaaagtga cctggccagc ccggaagccg ccagcatcct caagcgcgaa	6420
ctgaccgtgc cgggtgccga gcgtatctgg tacgtggctg acgccatgcg ttcgggcatg	6480
acctgcgaag aaattctcaa tctgaccggc atcgacatgt ggttcctggt gcagatggaa	6540
gacctgatca aggaagaaga gaaggtcaag accctggccc tgtcggcaat cgacaaggac	6600
tacatgctgc gctcaagcg caagggcttc tcggaccagc gcttggcagt actgctgggt	6660
atcacgcgaca agaacctgcg tcgccaccgc cacaagctgg aagtgttccc ggtgtacaag	6720
cgcgtcgaca cctgcgcgcg cgagttcgcc accgacaccg cctacctgta ctccacctac	6780
gaggaagagt gcgaggccaa cccgtcgacc cgcgacaaga tcatgatcct ggggtggcggc	6840
ccgaaccgta tcggccaagg tatcgagttc gactactgct gcgtacacgc cgcctggcg	6900
ctgcgtgaag acggttacga gaccatcatg gtcaactgca acccggaac cgtctccacc	6960
gactacgaca cttccgaccg cctgtacttc gagccgctga ccctggaaga cgtgctggaa	7020
gtctgccgcg tcgagaagcc gaagggcgtg atcgttcact acggcgcca gaccccgctg	7080
aagctggccc gcgctctgga agaggtggc gtgcgatca tcggtaccag ccctgacgcc	7140
atcgaccggc ccgaagaccg cgagcgcttc cagcagatgg ttcagcgcct gagcctgctg	7200
cagccgcaa acgccaccgt gcgcagcgaa gaagaagcca tccgtgctgc gggcagcatc	7260
ggctaccgcg tggctgtgcg tccgtcctac gtactggcg gccgtgccat ggagatcgtc	7320
tacgagctgg acgagctcaa gcgctacctg cgtgaagcgg tacaagtgtc gaacgacagc	7380
ccggtactgc tcgaccactt cctcaactgc gccatcgaga tggacgtgga tgcggtgtgc	7440
gacggcaccg acgtggtcat cggcgcgac atgcagcaca tcgagcaggc cggcgtacac	7500
tccggcgact cggcgtgctc gctgccacct tactcgtga gcaaggaagt gcaggacgaa	7560
gtccgcgttc aggtcaagaa aatggcgtg gagctgggtg tagtcggcct gatgaacgtg	7620
cagctggccc tgcaggcgca caagatctac gtgatcgaag tcaaccgcg tgccctcgct	7680
accgtaccgt tcgtgtccaa gtgcacccgc acgtccctgg cgatgatcgc agcccgctgc	7740
atggcgggta aaacctgaa agagctggc ttcaccagc aaatcatccc gaacttctac	7800
agcgtgaagg aagccgtctt cccgttcgcc aagttcccag gggttgacct gatcctcggc	7860
cctgagatga aatcgaccgg tgaagtgatg ggtgtcgggt acagcttcgg tgaagccttc	7920
gccaaagccc agatgggtgc cagcgaagtg ctgccgactg gcggtaccgc gttcatcagc	7980
gtgcgcgacg acgacaagcc acaagtggcc ggcgttgccc gcgacctgat cgcctgggc	8040

ttcgaagtgg ttgccactgc cggcaccgcc aaggttatcg aggcggctgg cctgaaagtg	8100
cgccgtgtga acaaggtgac cgaaggtcgc cctcacgtgg tcgacatgat caagaacgac	8160
gaagtgtcgc tgatcatcaa caccaccgaa ggctgccagt cgategccga ctctactcgc	8220
attcgtcgca atgcgctgca gcacaagatt tactgcacca ctaccattgc ggctggtgaa	8280
gccatctgcg aagcgctgaa attcgggtccg gaaaagaccg ttcgtcgctt gcaggatctg	8340
catgcaggac tgaagcatg agcattacca aataggcatt tgagaagcac acggtcacac	8400
tgcttccggg agtcaataaa ccggtaaacc agcaatagac ataagcggct atttaacgac	8460
cctgccctga accgacgacc gggtcgaatt tgctttcgaa tttctgccat tcatccgctt	8520
attatcactt attcaggcgt agcaccaggc gtttaagggc accaataact gccttaaaaa	8580
aattacgccc cgccctgcca ctcatcgag tactgttgta attcattaag cattctgcgc	8640
acatggaagc catcaciaac ggcatgatga acctgaatcg ccagcggcat cagcaccttg	8700
tcgccttgcg tataatatTT gccatggat ttaaatttaa tctttctgcg aattgagatg	8760
acgccactgg ctgggcgtca tcccggtttc ccgggtaaac accaccgaaa aatagt tact	8820
atcttcaaag ccacattcgg tcgaaatata actgattaac aggcggctat gctggagaag	8880
atattgcgca tgacacactc tgacctgtcg cagatattga ttgatggcca ttccagtctg	8940
ctggcgaaat tgctgacgca aaacgcgctc actgcacgat gcctcatcac aaaatttata	9000
cagcgcaaag ggacttttca ggctagccgc cagccgggta atcagcttat ccagcaacgt	9060
ttcgtcggat gttggcgcca acgaatcact ggtgtaacga tggcgattca gcaacatcac	9120
caactgccc aacagcaact cagccatttc gttagcaaac ggcatatgct gactactttc	9180
atgctcaagc tgaccgataa cctgcgcgcg ctgcgccatc cccatgctac ctaagcgcca	9240
gtgtggttgc cctgcgctgg cgttaaattc ccgaatcgcc ccctgccagt caagattcag	9300
cttcagacgc tccgggcaat aaataatatt ctgcaaaacc agatcgtaa cggaagcgta	9360
ggagtgttta tcgtcagcat gaatgtaaaa gagatcgcca cgggtaatgc gataagggcg	9420
atcgttgagt acatgcaggc cattaccgcg ccagacaatc accagctcac aaaaatcatg	9480
tgtatgttca gaaagacat cttgcggata acggtcagcc acagcgactg cctgctggtc	9540
gctggcaaaa aaatcatctt tgagaagttt taactgatgc gccaccgtgg ctacctcggc	9600
cagagaacga agttgattat tcgcaatatg gcgtacaaat acgttgagaa gattcgcgtt	9660
attgcagaaa gccatccgt cctggcgaa tatcacgcgg tgaccagtta aactctcggc	9720
gaaaaagcgt cgaaggtgg ttactgtcgc tgaatccaca gcgataggcg atgtcagtaa	9780
cgctggcctc gctgtggcgt agcagatgtc gggctttcat cagtcgcagg cggttcaggt	9840
atcgtgagg cgtcagtcct gtttgctgct taagctgccg atgtagcgta cgcagtga	9900
gagaaaattg atccgccacg gcataccaat tcacctcatc ggcaaaatgg tcctccagcc	9960
aggccagaag caagttgaga cgtgatgcgc tgttttccag gttctcctgc aaactgcttt	10020
tacgcagcaa gagcagtaat tgcataaaca agatctcgcg actggcggtc gagggtaaat	10080
cattttcccc ttctgctgt tccatctgtg caaccagctg tcgcacctgc tgcaatacgc	10140
tgtggttaac gcgccagtga gacggatact gccatccag ctcttggtggc agcaactgat	10200

tcagcccggc gagaaactga aatcgatccg gcgagcgata cagcacattg gtcagacaca	10260
gattatcggg atgttcatac agatgccgat catgatcgcg tacgaaacag accgtgccac	10320
cggtgatggg ataggggtgc ccattaaaca catgaatacc cgtgccatgt tcgacaatca	10380
caatttcattg aaaatcatga tgatgttcag gaaaatccgc ctgcgggagc cgggggttcta	10440
tcgccacgga cgcgttacca gacggaaaaa aatccacact atgtaatacg gtcatactgg	10500
cctcctgatg tcgtcaacac ggcgaaatag taatcacgag gtcaggttct taccttaaat	10560
tttcgacgga aaaccacgta aaaaacgtcg atttttcaag atacagcgtg aattttcagg	10620
aaatgcggg agcatcacat caccacaatt cagcaaattg tgaacatcat caggttcac	10680
tttcctcggg tgccaatggc ccattttcct gtcagtaacg agaaggctgc gaattcaggc	10740
gctttttaga ctggtcgtaa tgaacattta aatgaattcc cttgggactc tagagatccg	10800
cggggggcca ggagggggga tctggcattt ttgggagggtg tgaaatgagt gagaagatcg	10860
tcaactcgtg gaacgaatgg gatgagctcg aggagatggg ggtcggcatt gcggactatg	10920
ccagcttcga accgaaagag ccaggcaacc atcccaaact gcgcaaccag aacctggccg	10980
aatcatccc cttccaagc ggcccaaagg acccgaagg gctggagaaa gcgaacgaag	11040
agctgaatgg gctggcttac ctgctgaagg accacgatgt gatcgtgctg cgtcccaga	11100
agatcgactt caccaagagc ctgaaaacc cgtatttcga ggttgccaac cagtactgcg	11160
gcgtttgtcc tcgcgacgtg atgatcacgt ttggcaacga aatcatggaa gcgaccatgt	11220
ccaaacgtgc acgcttcttc gaatacctcc cctatcgga gctggtctac gagtactgga	11280
acaaggacga gcacatgatc tggaacgcag ccccgaaacc gaccatgcag gatagcatgt	11340
acctggaaaa cttctgggag ctctcgctgg aagaacgctt caagcggatg cagcacttcg	11400
aattctgcat caaccaagac gaggtgatct tcgatgccg cgattgctcc cgcttgggta	11460
aggacatcct ggtgcaggaa agcatgacca ccaatcgac tggcatccgc tggtgaaga	11520
agcatctcga accacggcgc ttctcgctcc atccgggtgca cttcccggtg gacttcttcc	11580
ctagccacat cgactgcacg ttctgaccgt tgcgtccggg tctgatcctg accaatccgg	11640
aacgcccgat tcgcgaggaa gaggagaaga tcttcaagga gaatggctgg gagctgatca	11700
ccgtaccgca gcctacctgc tcgaacgacg agatgcccat gttctgccag agctcgaaat	11760
ggctgtccat gaacgtcctg agcattagtc ccaccaagg gattctgcga gaacgggaaa	11820
agccgctgca agaactgctg gacaagcacg ggttcgaagt cttcccttg ctttccgcc	11880
atgtgtttga gttcgggtgc agctttcact gtgccacgtg ggatattcgc cgcaagggcg	11940
agtgcgagga ctacctgccg aacctgaact accagccgat ttgcggctga cgcggatccc	12000
gacctgcata ccgttcgaga taaggtagtc actcatgagc gctaggcact ttctctccct	12060
gctggacttc accaccgacg aattgctcgg ggtgatccgc caccgcatcg agctgaagga	12120
cctgcgcaag cgaggcgtgc tgttcgaacc gctgaagaac cgtgtgctag gcatgatctt	12180
cgaaaagtcc tcgacccgta cccgtgtgtc gttcgaggcc ggcattgatcc agctcggcgg	12240
ccaggccatc ttctgtgcgc cccgcgacac ccagctgggc cgcggcgagc caattggtga	12300
cagcgccatc gtgctgtcga gcatgggtga tgtggtgatg atccggaccc acgcccacag	12360

cacctgacc gagttcgccg ccaagtcgcg tgtgcccgtg atcaacggcc tgtccgacga	12420
atcgcacccg tgccaactgc tggccgacat gcagaccttc gttgaacacc gcggctcgat	12480
tcagggcaag accgtgacct ggatcgcgga tggcttcaac atgtgcaact cctatatcga	12540
agccgccagg cagttcgatt tccagctgcg catcgctgc cccgaaggct atgagccgga	12600
tcaacgcttc atggcactgg gcggcgaccg cgtgcagatc atccgggatg ccaggaagc	12660
tgtgcgtgat gcacacctgg tggtcaccga tgtctggact tccatgggtc aggaggagga	12720
aactgcacgg cgctggcg ctttcgccc ttaccaggtc acccgcgaa tgctcgacct	12780
ggctgcaccc gatgtcctct tcatgcactg cctgccgcc caccgtggcg aggaatcag	12840
ccaggacctg ctgcagacc cacttcggt cgctgggac gaggtgaaa accgctgca	12900
tgcacagaag gcgcttctcg aattccttgt agaaccggct taccaccacg catgagtcaa	12960
ccgtacaacc ccgtggagt atggcatggc ggacgtaaaa aaggctgtac tggcgtattc	13020
cgccggcctt gatacttcgg tgattctcaa gtggctgcag gatactaca actgcgaagt	13080
ggtgaccttc accgtgacc tggggcaggg cgaagaggtc gaaccggccc gtgccaaggc	13140
ccaggcaatg ggcgttaaag agatctacat cgacgacctg cgcaagaat tcgtgcgtga	13200
tttcgtgttc ccgatgttc gcgccaacac cgtctacgaa ggcgagtacc tctgggtac	13260
ttccatcgcc cgtccgctga tcgccaagcg cctgatcgaa atcgccaacg aaaccggcgc	13320
tgacgccatt tccatggcg ccaccggcaa gggtaacgac caggtgcgtc tcgagctggg	13380
tgccatgcc ctgaagccag gcgtcaaggt catcgctcca tggcgcgagt gggacctgct	13440
gtcccgcgaa aagctgatgg actacgccga gaagcacggc atcccgatcg agcgccacgg	13500
caagaagaag tcgccgtact cgatggacgc caacctgctg cacatctcct acgagggcgg	13560
tgctctggaa gatactgga ccgagcacga agaagacatg tggcgctgga gtgtctcgcc	13620
tgagaatgcc ccggaccagg ctacctacat cgagctgacc taccgcaatg gtgacatcgt	13680
tgccatcgac ggcgtcgaga aatccccggc caccgtcctg gcagacctga accgtatcgg	13740
tggtgccaac ggcacggcc gtctggacat cgtcgaaaac cgttacgtcg gcatgaagtc	13800
gcgcggttgc tacgaaacgc ctggcggtac catcatgctc aaggcacacc gtgccatcga	13860
gtcgatcacc ctggaccgcg aagtcgctca cctgaaagat gagctgatgc caaagtatgc	13920
cagcctgac tacaccggct actggtggag cccggagcgt ctgatgctgc aacagatgat	13980
cgatgcttcg caggtcaacg tgaatggtgt ggtgcgcctg aaactgtaca agggcaacgt	14040
gaccgtggtt gcccgcaagt cggacgattc gctgttcgat gccaacatcg ccacctttga	14100
agaagatggt ggtgcctaca accaggcaga tgctgctggc ttcacaaagc tcaatgcact	14160
gcgtatgcgc attgccgcca acaaggcccg ttcgctgctc tgattgctat cgacgccact	14220
ttttcgttca cgctgcaat gagtgaatcc atgagcaccg agaagaccaa tcagtctcgg	14280
ggcggccgct tcagttagcc cgtcgacgcc ttcgtcgccc gtttcaccgc ctcggtagat	14340
ttcgacaagc gcctgtaccg tcacgacatc atgggttoga ttgcccatgc caccatgctg	14400
gcgcaggtcg gcgtgctcag tgatgccgag cgcgacacca tcatcgatgg cctgaaaacc	14460
atccaggcgc agattgaagc cggcaacttc gactggcgtg tcgacctcga agacgtgcac	14520

atgaacatcg aagcacgcct gaccgaccgc atcggcatca ccggcaagaa gctgcatact	14580
ggcgtagcc gcaacgacca ggtggccacc gacatccgcc tttggctgcg cgacgaaatc	14640
gacctgatcc tggcgaaat caccgcctg cagcagggcc tgctggagca ggcagagcgt	14700
gaagcogaaa ccatcatgcc tggtttcacc cacctgcaga cggcgagcc ggtaaccttt	14760
ggccaccacc tgctggcgtg gttcgaaatg ctcagccgcg actatgagcg cctggtogac	14820
tgccgcaagc gcaccaaccg catgccactg ggcagcgccg cgctggccgg caccacctac	14880
ccgatcgacc gtgaactgac ctgcaagctg ctgggctttg aagccgtggc cggcaactcg	14940
ctggatggcg tgctggaccg tgatttcgcc atcgaattct gcgcgctgc cagcgtggcg	15000
atgatgcacc ttctcgctt ctccgaagag ctggtgctgt ggaccagcg gcagttccag	15060
ttcatcgacc ttccggaccg cttctgcact ggcagctcga tcatgccga gaaaaagac	15120
ccggacgtgc cagagctggt acgtggcaag agcggccgcg tgctcgccg cctgaccggc	15180
ctgctgacct tgatgaaagg ccaaccgctg gcctacaaca aggacaacca ggaagacaag	15240
gaaccgctgt tcgacgccg cgataccctg cgcgactcgc tcggggcctt cgctgacatg	15300
atcccgcgga tcaagcccaa gcacgccatc atgcgtgaag cggccctgcg cggtttctcc	15360
accgctaccg acctggctga ctatctggtt cgcctggcc tgccgttccg tgactgccac	15420
gagatcgttg gccacgcggt gaagtatggt gtggacactg gcaaggacct ggcgagatg	15480
agcctggagc aactgcgcc attcagcgac cagatcgagc aggacgtgtt tgccgtgctg	15540
acgctggaag gctcggtgaa tgcgcgtgac cacattggtg gtacggcgcc ggcgcaggtg	15600
cgtgctgcg tcgttcgttg caaggccctg ttggcgtctc gctaataccc caaggctcga	15660
gcacgcgaga gtagggaact gccaggcatc aaataaaacg aaaggctcag tcgaaagact	15720
gggcctttcg ttttatctgt tgtttgtcgg tgaacgctct cctgagtagg acaaatccgc	15780
cgggagcgga tttgaacgat gataagctgt caaacatgag aattcttgaa gacgaaagg	15840
cctcgtgtgt acaaagttc gtcaaaagg cgacacaaaa ttctgcagg ggcgggcca	15900
gcgcggcgcg tcgagtggcg acggcgcgcc ttgtccgcgc cctggtagat tgctggccg	15960
taggccagcc atttttgagc ggccagcgcc cgcgataggc cgacgcgaag cggcgggcg	16020
tagggagcgc agcgaccgaa gggtaggcgc tttttgcagc tcttcggctg tgcgctggcc	16080
agacagttat gcacaggcca ggcgggtttt aagagtttta ataagtttta aagagtttta	16140
ggcggaaaaa tcgccttttt tctcttttat atcagtcact tacatgtgtg accggttccc	16200
aatgtacggc tttgggttcc caatgtacgg gtcccggttc ccaatgtacg gctttgggtt	16260
cccaatgtac gtgctatcca caggaaagag accttttoga cttttttccc ctgctagggc	16320
aatttgccct agcatctgct ccgtacatta ggaaccggcg gatgcttcgc cctcgatcag	16380
gttcggtag cgcatgacta ggatcgggcc agcctgcccc gcctcctcct tcaaatocta	16440
ctccggcagc tcatttgacc cgatcagctt gcgcacggtg aaacagaact tcttgaactc	16500
tccggcgctg ccaactgcgtt cgtagatcgt cttgaacaac catctggctt ctgccttgcc	16560
tgccggcgcg cgtgccaggc ggtagagaaa acggccgatg ccgggatcga tcaaaaagta	16620
atcggggtga accgtcagca cgtccgggtt cttgccttct gtgatctcgc ggtacatoca	16680

atcaactagc	togatctega	tgtactccgg	ccgcccgggt	tcgctcttta	cgatcttgta	16740
gcggctaata	aaggettca	cctcggatac	cgtcaccagg	cgcccggtct	tggccttctt	16800
cgtacgctgc	atggcaacgt	gcgtgggtgt	taaccgaatg	caggtttcta	ccaggtcytc	16860
tttctgcttt	ccgccatcgg	ctcgccggca	gaacttgagt	acgtccgcaa	cgtgtggacg	16920
gaacacgcgg	ccgggcttgt	ctcccttccc	ttcccgggat	cggttcatgg	attcggttag	16980
atgggaaacc	gccatcagta	ccaggtcgta	atcccacaca	ctggccatgc	cgcccgcccc	17040
tgcggaaacc	tctacgtgcc	cgtctggaag	ctcgtagcgg	atcacctcgc	cagctcgtcg	17100
gtcacgcttc	gacagacgga	aaacggccac	gtccatgatg	ctgcgactat	cgcggtgtgc	17160
cacgtcatag	agcatcggaa	cgaaaaaatc	tggttgctcg	tcgcccttgg	gcggcttctt	17220
aatcgacggc	gcaccggctg	ccggcgggtg	ccgggattct	ttcgggattc	gatcagcggc	17280
cgcttgccac	gattcaccgg	ggcgtgcttc	tgcctcgatg	cgttgccgct	ggcgccgctg	17340
cgcgcccttc	aactttctca	ccaggtcata	accagcgcgc	gcgcgatttt	gtaccggggc	17400
ggatggtttg	cgaccgctca	cgccgattcc	tcgggcttgg	gggttccagt	gccattgcag	17460
ggccggcaga	caaccagacc	gcttacgcct	ggccaaccgc	ccgttccctc	acacatgggg	17520
cattccacgg	cgtcgggtgc	tggttgttct	tgattttcca	tgccgcctcc	tttagccgct	17580
aaaattcatc	tactcattta	ttcatttgct	catttactct	ggtagctgcg	cgatgtattc	17640
agatagcagc	tcggtaatgg	tcttgccctg	gcgtaccgcg	tacatcttca	gcttggtgtg	17700
atcctccgcc	ggcaactgaa	agttgaccgc	cttcattggc	ggcgtgtctg	ccaggctggc	17760
caacgttgca	gccttgctgc	tgcgtgcgct	cggacggccg	gcacttagcg	tgtttgtgct	17820
tttgctcatt	ttctctttac	ctcattaact	caaataagtt	ttgatttaac	ttcagcggcc	17880
agcgccctga	cctcgcgggc	agcgtgcgcc	tcgggttctg	attcaagaac	ggttgtgccg	17940
gcggcgccag	tgccctggga	gctcacgcgc	tgcgtgatac	gggactcaag	aatgggcagc	18000
tcgtaccocg	ccagcgcttc	ggcaacctca	ccgccgatgc	gcgtgccttt	gatcgccgcg	18060
gacacgacaa	aggccgcttg	tagccttcca	tccgtgacct	caatgcgctg	cttaaccagc	18120
tccaccaggt	cggcggtggc	ccatatgtcg	taagggcttg	gctgcaccgg	aatcagcacg	18180
aagtcggctg	ccttgatcgc	ggacacagcc	aagtcgcgcg	cctggggcgc	tcgctcgatc	18240
actacgaagt	cgcgcgggcc	gatggccttc	acgtcgcggg	caatcgtcgg	gcggtcgatg	18300
ccgacaacgg	ttagcgggtg	atcttcccgc	acggccgcgc	aatcgcgggc	actgccctgg	18360
ggatcggaat	cgactaacag	aacatcggcc	ccggcgagtt	gcagggcgcg	ggctagatgg	18420
gttgcgatgg	tcgtcttgcc	tgaccgcctt	ttctgggtta	gtacagcgat	aaccttcacg	18480
cgttccccct	gcgtatttgt	ttatttactc	atcgcatcat	atacgacgcg	accgcatgac	18540
gcaagctggt	ttactcaaata	acacatcacc	tttttagacg	gcggcgctcg	gtttcttcag	18600
cggccaagct	ggccggccag	gccgcagct	tggcatcaga	caaaccggcc	aggatttcac	18660
gcagccgcac	ggttcgggat	gagcattca				18689

YÊU CẦU BẢO HỘ

1. Vi sinh vật có hoạt tính enzym có chức năng carbamoylphosphat synthaza gia tăng so với hoạt tính của enzym tương ứng ở vi sinh vật kiểu dại và mang ít nhất một gen dị hợp mã hóa cho protein có hoạt tính của L-arginin:glyxin amidinotransferaza, trong đó hoạt tính tăng lên của enzym có chức năng carbamoylphosphat synthaza đạt được bằng cách biểu hiện quá mức gen mã hóa cho enzym có chức năng carbamoylphosphat synthaza.

2. Vi sinh vật theo điểm 1, trong đó sự biểu hiện quá mức của gen mã hóa cho enzym có chức năng carbamoylphosphat synthaza đạt được bằng cách tăng số lượng bản sao của gen và/hoặc bằng cách liên kết chức năng gen này với vùng khởi động mạnh.

3. Vi sinh vật theo điểm 1 hoặc điểm 2, trong đó hoạt tính của L-arginin:glyxin amidinotransferaza được gia tăng bằng cách biểu hiện quá mức gen mã hóa cho L-arginin:glyxin amidinotransferaza.

4. Vi sinh vật theo điểm bất kỳ trong số các điểm nêu trên, trong đó vi sinh vật này có khả năng tổng hợp L-arginin cải thiện hơn so với khả năng tổng hợp L-arginin của vi sinh vật kiểu dại.

5. Vi sinh vật theo điểm 4, trong đó gen *argR* mã hóa cho protein ức chế đáp ứng arginin *argR* bị làm yếu đi hoặc bị loại bỏ.

6. Vi sinh vật theo điểm bất kỳ trong số các điểm 4 hoặc 5, trong đó ít nhất một hoặc nhiều gen mã hóa cho enzym của con đường sinh tổng hợp L-ornitin và L-arginin, bao gồm *gdh*, *argJ*, *argB*, *argC* và/hoặc *argD* lần lượt mã hóa cho glutamat dehydrogenaza, ornitin axetyltransferaza, axetylglutamat

kinaza, axetylglutamylphosphat reductaza và axetylornitin aminotransferaza, được biểu hiện quá mức.

7. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 1 đến 6, trong đó protein có chức năng L-arginin:glyxin amidinotransferaza có trình tự axit amin tương đồng với trình tự axit amin được chọn từ nhóm bao gồm SEQ ID NO:2, SEQ ID NO:16, SEQ ID NO:24, và SEQ ID NO:20 ít nhất là 70%.

8. Vi sinh vật theo điểm bất kỳ trong số các điểm nêu trên, trong đó vi sinh vật này thuộc giống *Corynebacterium*, giống *Enterobacteriaceae* hoặc giống *Pseudomonas*.

9. Vi sinh vật theo điểm 8, trong đó vi sinh vật này là *Corynebacterium glutamicum*.

10. Vi sinh vật theo điểm 8, trong đó vi sinh vật này là *Escherichia coli*.

11. Vi sinh vật theo điểm 8, trong đó vi sinh vật này là *Pseudomonas putida*.

12. Phương pháp lên men tổng hợp axit guanidino axetic (GAA), trong đó phương pháp này bao gồm các bước a) nuôi cấy vi sinh vật theo điểm bất kỳ trong số các điểm nêu trên trong môi trường lên men, và b) tích lũy GAA trong môi trường này để tạo ra môi trường lên men chứa GAA.

13. Phương pháp theo điểm 12, trong đó phương pháp này còn bao gồm bước tách GAA từ môi trường lên men chứa GAA này.

14. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 1 đến 11, trong đó vi sinh vật này còn mang gen mã hóa cho enzym có hoạt tính guanidinoaxetat N-metyltransferaza.

15. Vi sinh vật theo điểm 14, trong đó gen mã hóa cho enzym có hoạt tính guanidinoaxetat N-metyltransferaza được biểu hiện quá mức.

16. Phương pháp lên men tổng hợp creatin, trong đó phương pháp này bao gồm các bước a) nuôi cấy vi sinh vật theo điểm 14 hoặc điểm 15 trong môi trường lên men, và b) tích lũy creatin trong môi trường này để tạo ra môi trường lên men chứa creatin.

17. Phương pháp theo điểm 16, trong đó phương pháp này còn bao gồm bước tách creatin từ môi trường lên men chứa creatin.