



(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) 
1-0022414

(51)⁷ C12P 7/50, C12N 15/09, C12P 13/00,
13/10, 13/14, 13/24

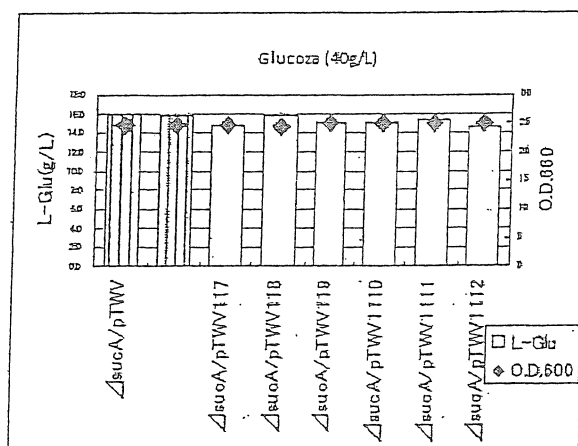
(13) B

- (21) 1-2014-01844 (22) 06.11.2012
(86) PCT/JP2012/078725 06.11.2012 (87) WO2013/069634A1 16.05.2013
(30) 2011-247031 11.11.2011 JP
61/558,685 11.11.2011 US
(45) 25.12.2019 381 (43) 25.09.2014 318
(73) AJINOMOTO CO., INC. (JP)
15-1, Kyobashi 1-chome, Chuo-ku, Tokyo 104-8315 Japan
(72) NISHIO, Yousuke (JP), YAMAMOTO, Youko (JP), YAMADA, Kazuteru (JP),
YOKOTA, Kosuke (JP)
(74) Công ty TNHH một thành viên Sở hữu trí tuệ VCCI (VCCI-IP CO.,LTD)

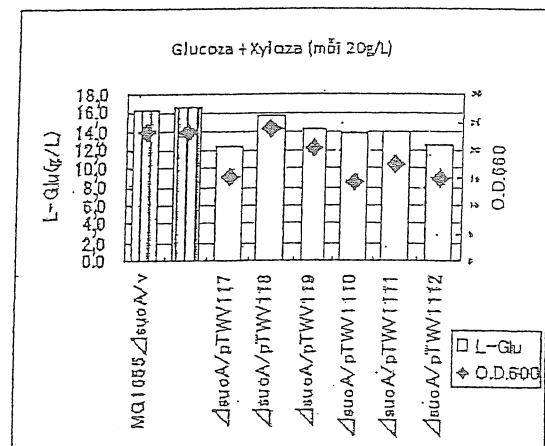
(54) PHƯƠNG PHÁP SẢN XUẤT HỢP CHẤT ĐÍCH BẰNG CÁCH LÊN MEN

(57) Sáng chế đề cập đến phương pháp sản xuất hợp chất đích bằng cách lên men. Bằng cách nuôi cấy vi khuẩn có khả năng sản xuất axit 2-ketoglutaric hoặc dẫn xuất của nó dưới dạng hợp chất đích, và khả năng sản xuất axit xylonic từ xyloza, mà được truyền hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3 - deoxyxylonat dehydrataza và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza, hoặc trong đó các hoạt tính này được làm tăng, trong môi trường nuôi cấy chứa xyloza dùng làm nguồn cacbon để tạo ra và tích lũy hợp chất đích trong môi trường này, và thu gom hợp chất đích từ môi trường, hợp chất đích được tạo ra.

a)



b)



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp sản xuất hợp chất đích như các axit amin dạng L bằng cách lên men sử dụng vi sinh vật, chính xác hơn, phương pháp sản xuất hợp chất đích bằng cách lên men như vậy sử dụng xyloza làm nguyên liệu thô.

Tình trạng kỹ thuật của sáng chế

Các phương pháp sản xuất hợp chất đích như các axit amin dạng L bằng cách lên men sử dụng vi khuẩn bao gồm phương pháp sử dụng vi khuẩn kiểu đại (chủng kiểu đại), phương pháp sử dụng chủng khuyết dưỡng bắt nguồn từ chủng kiểu đại này, phương pháp sử dụng chủng đột biến điều hòa cơ chế trao đổi chất bắt nguồn từ chủng kiểu đại này dưới dạng chủng kháng các dược chất khác nhau, phương pháp sử dụng chủng có các tính chất của cả chủng khuyết dưỡng và thể đột biến điều hòa quá trình trao đổi chất, và v.v..

Ví dụ, axit L-glutamic chủ yếu được sản xuất bằng cách lên men sử dụng vi khuẩn sản xuất axit L-glutamic của vi khuẩn được gọi là coryneform thuộc họ *Brevibacterium*, *Corynebacterium* hoặc *Microbacterium* hoặc chủng đột biến của chúng (đề cập đến, ví dụ, tài liệu phi sáng chế 1). Dùng làm các phương pháp sản xuất axit L-glutamic bằng cách sử dụng các chủng khác, các phương pháp đã biết sử dụng vi sinh vật thuộc họ *Bacillus*, *Streptomyces*, *Penicillium*, hoặc các loài tương tự (đề cập đến, ví dụ, tài liệu sáng chế 1), các phương pháp sử dụng vi sinh vật thuộc họ *Pseudomonas*, *Arthrobacter*, *Serratia*, *Candida*, hoặc các họ tương tự (đề cập đến, ví dụ, tài liệu sáng chế 2), các phương pháp sử dụng vi sinh vật thuộc họ *Bacillus*, hoặc *Aerobacter aerogenes* (hiện tại là họ *Enterobacter aerogenes*), hoặc các họ tương tự (đề cập đến, ví dụ, tài liệu sáng chế 3), các phương pháp sử dụng chủng đột biến của *Escherichia coli* (đề cập đến, ví dụ, tài liệu sáng chế 4), và v.v.. Hơn thế nữa, cũng đã bộc lộ các phương pháp sản xuất axit L-glutamic sử dụng vi sinh vật thuộc họ *Klebsiella*, *Erwinia*, *Pantoea*, hoặc *Enterobacter* (đề cập đến, ví dụ, các tài liệu sáng chế 5 đến 7).

Trong những năm gần đây, các kỹ thuật ADN tái tổ hợp được sử dụng trong phương pháp sản xuất các hợp chất đích bằng phương pháp ly tâm. Ví dụ, khả năng

sản xuất axit amin dạng L của vi khuẩn được cải thiện bằng cách làm tăng sự biểu hiện của gen mã hóa enzym sinh tổng hợp axit amin dạng L (các tài liệu sáng chế 8 và 9), hoặc bằng cách làm tăng sự hấp thu nguồn cacbon vào hệ sinh tổng hợp axit amin dạng L (tài liệu sáng chế 10).

Trong phương pháp sản xuất các chất thông thường bằng cách ly tâm, dùng làm nguồn cacbon, sacarit, tức là, glucoza, fructoza, sucroza, mật đường, dịch thủy phân tinh bột và v.v. đã được sử dụng, nhưng chúng có giá thành tương đối cao, và việc sử dụng nguyên liệu sinh khối thu được từ thực vật và vật liệu tương tự cũng được ưu tiên trong những năm gần đây.

Mặc dù các nguyên liệu thô bao gồm các thành phần ăn được như tinh bột và chất béo và dầu hiện tại chủ yếu được sử dụng dưới dạng nguyên liệu sinh khối, trong tương lai cần phải thay đổi nguyên liệu sinh khối như vậy bằng các nguyên liệu bao gồm các thành phần không ăn được, cụ thể là, xenluloza, hemixenluloza, lignin, và v.v.. Xenluloza và hemixenluloza ở dạng sinh khối không ăn được được chuyển hóa thành các pentoza hoặc hexoza nhờ quá trình tiền xử lý sơ bộ sử dụng nhiệt hoặc axit, và xử lý sacarit hóa sử dụng enzym xenlulaza, và chúng có thể được sử dụng làm các nguyên liệu thô trong quá trình lên men (các tài liệu sáng chế 11 và 12). Đã biết rằng nếu sacarit hỗn hợp của các pentoza hoặc hexoza như vậy được dùng làm các nguyên liệu thô cho quá trình lên men axit amin v.v., *Escherichia coli* ưu tiên đồng hóa glucoza, và kết quả là, hiện tượng tăng sinh trong hai bước (diauxy), sự sinh trưởng bị trì hoãn, v.v. đã được quan sát thấy (các tài liệu sáng chế 2 và 3)

Trong *Escherichia coli*, quá trình đồng hóa xyloza bao gồm xyloza isomerase được mã hóa bởi gen *xylA* và xylulokinaza được mã hóa bởi gen *xylB* là đã biết, và cũng đã biết rằng các axit amin dạng L có thể được tạo ra từ xyloza bằng cách đưa quá trình đó vào *Escherichia coli* hoặc *Corynebacterium glutamicum* (tài liệu sáng chế 5, tài liệu sáng chế 13, tài liệu sáng chế 7).

Cũng đã có thông báo rằng *Caulobacter crescentus* và *Haloferax volcanii* có quá trình chuyển hóa xyloza thành axit 2-ketoglutaric thông qua axit xylonic trong năm bước, không sử dụng quá trình đã biết thông thường như vậy như được mô tả trên đây (tài liệu sáng chế 6). Hơn thế nữa, các ví dụ về sự biểu hiện của quá trình đã biết đó trong *Escherichia coli* là cũng đã biết (tài liệu sáng chế 8, tài liệu sáng chế

14).

Tài liệu trích dẫn

Tài liệu sáng chế

Tài liệu sáng chế 1: Đơn sáng chế Mỹ số 3,220,929

Tài liệu sáng chế 2: Đơn sáng chế Mỹ số 3,563,857

Tài liệu sáng chế 3: Công bố đơn sáng chế Nhật Bản (Kokoku) số 32-9393

Tài liệu sáng chế 4: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định (Kokai) số 5-244970

Tài liệu sáng chế 5: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2000-106869

Tài liệu sáng chế 6: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2000-189169

Tài liệu sáng chế 7: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2000-189175

Tài liệu sáng chế 8: Đơn sáng chế Mỹ số 5,168,056

Tài liệu sáng chế 9: Đơn sáng chế Mỹ số 5,776,736

Tài liệu sáng chế 10: patent Mỹ số 5,906,925

Tài liệu sáng chế 11: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định trên cơ sở đơn PCT (Kohyo) số 9-507386

Tài liệu sáng chế 12: Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định trên cơ sở đơn PCT số 11-506934

Tài liệu sáng chế 13: Patent châu Âu số 1577396

Tài liệu sáng chế 14: patent Mỹ số 7,923,226

Tài liệu phi sáng chế

Tài liệu phi sáng chế 1: Akashi K. et al., Axit amin Fermentation, Japan Scientific Societies Press, pp,195-215, 1986

Tài liệu phi sáng chế 2: Nichols N.N. et al., Appl. Microbiol. Biotechnol.,

2001 Jul, 56(1-2):120-125

Tài liệu phi sáng chế 3: Gonzalez, R., *Biotechnol. Prog.*, 2002 Jan-Feb, 18(1):6-20

Tài liệu phi sáng chế 4: Song S. et al., *J. Bacteriol.*, 1997 Nov, 179 (22):7025-7032

Tài liệu phi sáng chế 5: Tao H. et al., *J. Bacteriol.*, 2001 May, 183 (10):2979-2988

Tài liệu phi sáng chế 6: Stephens, C. et al., *J. Bacteriol.*, 2007 Mar., 189 (5):2181-2185

Tài liệu phi sáng chế 7: Gopinath, V. et al., *Appl. Microbiol. Biotechnol.*, 2011 Jul., 28

Tài liệu phi sáng chế 8: Huaiwei, L et al., *Bioresour Technol.*, 2011 Aug., 22

Bản chất kỹ thuật của sáng chế

Vấn đề được giải quyết bởi sáng chế

Mục đích của sáng chế là đề xuất vi sinh vật mà có thể tạo ra một cách hiệu quả hợp chất đích như axit L-glutamic trong môi trường nuôi cấy chứa xyloza, và phương pháp sản xuất hợp chất đích sử dụng vi sinh vật này.

Các phương pháp thực hiện

Các tác giả sáng chế đã nghiên cứu sự phát triển của vi sinh vật bằng cách nhân giống sử dụng quá trình chuyển hóa xyloza thành axit 2-ketoglutaric thông qua axit xylonic với mục đích phát triển vi khuẩn sản xuất axit amin có hoạt tính đồng hóa pentoza hoặc hexoza bằng cách nhân giống. Kết quả là, họ đã khám phá ra rằng vi sinh vật biểu hiện quá trình như vậy như được mô tả trên đây có thể đồng hóa xyloza một cách hiệu quả, và hoàn thành sáng chế.

Do đó, sáng chế được đề cập như sau.

(1) Phương pháp sản xuất hợp chất đích bao gồm bước nuôi cấy vi khuẩn có khả năng tạo ra hợp chất đích trong môi trường nuôi cấy chứa xyloza dùng làm nguồn cacbon, để tạo ra và tích lũy hợp chất đích trong môi trường này, và thu gom hợp chất đích từ môi trường, trong đó:

hợp chất đích là axit 2-ketoglutaric hoặc dẫn xuất của nó,

vi khuẩn có khả năng sản xuất axit xylonic từ xyloza, và hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza đã được truyền hoặc làm tăng trong vi khuẩn.

(2) Phương pháp như được mô tả trên đây, trong đó các hoạt tính enzym của xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza được truyền hoặc làm tăng trong vi khuẩn bằng cách nhập vào lần lượt các gen mã hóa các enzym này ở dạng có thể biểu hiện vào vi khuẩn.

(3) Phương pháp như được mô tả trên đây, trong đó gen mã hóa mỗi xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza thu được từ vi sinh vật thuộc họ *Caulobacter*, *Escherichia*, *Agrobacterium*, *Herbaspirillum*, *Actinoplanes*, *Cupriavidus*, *Pseudomonas*, *Zobellia*, *Thermobacillus*, *Arthrobacter*, *Azospirillum*, *Halomonas*, *Bacillus*, hoặc *Aspergillus*.

(4) Phương pháp như được mô tả trên đây, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza vì có một trong các đặc điểm sau:

(A) hoạt tính xyloza dehydrogenaza, hoặc hoạt tính xyloza dehydrogenaza và hoạt tính xylonolactonaza được truyền hoặc làm tăng trong vi khuẩn, hoặc

(B) vi khuẩn có hoạt tính glucoza dehydrogenaza mà xúc tác phản ứng sản xuất axit xylonic từ xyloza.

(5) Phương pháp như được mô tả trên đây, trong đó glucoza dehydrogenaza sử dụng pyroloquinolin quinon làm coenzym, và vi khuẩn thể hiện hoạt tính glucoza dehydrogenaza vì nó có khả năng sản xuất pyroloquinolin quinon, hoặc được nuôi cấy trong môi trường nuôi cấy chứa pyroloquinolin quinon.

(6) Phương pháp như được mô tả trên đây, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza vì nó đã được đưa vào gen mã hóa xyloza dehydrogenaza, hoặc các gen mã hóa xyloza dehydrogenaza và xylonolactonaza ở dạng có thể biểu hiện được.

(7) Phương pháp như được mô tả trên đây, trong đó vi khuẩn được biến đổi sao cho hoạt tính của 2-ketoglutarat dehydrogenaza là giảm.

(8) Phương pháp như được mô tả trên đây, trong đó vi khuẩn được tiếp tục biến đổi sao cho hoạt tính của succinat dehydrogenaza là giảm.

- (9) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là vi khuẩn đường ruột hoặc vi khuẩn coryneform.
- (10) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là vi khuẩn thuộc họ *Pantoea*.
- (11) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là *Pantoea ananatis*.
- (12) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là vi khuẩn thuộc họ *Escherichia*.
- (13) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là *Escherichia coli*.
- (14) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là vi khuẩn thuộc họ *Corynebacterium*.
- (15) Phương pháp như được mô tả trên đây, trong đó vi khuẩn này là *Corynebacterium glutamicum*.
- (16) Phương pháp như được mô tả trên đây, trong đó dẫn xuất của axit 2-ketoglutaric là hợp chất được chọn từ nhóm bao gồm axit L-glutamic, L-glutamin, L-arginin, L-xitruilin, L-ornithin, L-prolin, putresxin, và axit γ -aminobutyric.

Mô tả vắn tắt các hình vẽ

Fig. 1 mô tả các đồ thị thể hiện các kết quả của thử nghiệm sự sinh trưởng bổ trợ đối với chủng biểu hiện operon NXA thu được từ *C. crescentus* sử dụng chủng thiếu hụt gen *icd*. E1- α KG, M9- α KG, M9-Xyl, và E1-Xyl là môi trường tối thiểu M9 hoặc môi trường tổng hợp E1 chứa axit 2-ketoglutaric hoặc xyloza dùng làm nguồn cacbon duy nhất.

Fig. 2 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của chủng sản xuất axit L-glutamic *E. coli* biểu hiện operon ccrNXA *E. coli*.

Fig. 3 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của chủng biểu hiện operon ccrNXA sử dụng chủng thiếu hụt quá trình đồng hóa xyloza đặc trưng cho *E. coli* dùng làm vật chủ.

Fig. 4 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của chủng biểu hiện operon ccrNXA sử dụng plasmit loại số bản sao

trung bình.

Fig. 5 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của các chủng biểu hiện gen đồng đẳng *xylD* thu được từ các loại vi sinh vật khác nhau. Atu, Hse, Amis, và Aor lần lượt là *Agrobacterium tumefaciens*, *Herbaspirillum seropedicae*, *Actinoplanes missouriensis*, và *Aspergillus oryzae*.

Fig. 6 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của các chủng biểu hiện gen đồng đẳng *xylX* thu được từ các loại vi sinh vật khác nhau. Art, Atu, Cne, Zga, Tco, và Selo lần lượt là *Arthrobacter globiformis*, *Agrobacterium tumefaciens*, *Cupriavidus necator*, *Zobellia galactanivorans*, *Thermobacillus composti*, và *Pseudomonas elodea*.

Fig. 7 mô tả các đồ thị thể hiện các kết quả của quá trình nuôi cấy để sản xuất axit L-glutamic của các chủng biểu hiện gen đồng đẳng *xylA* thu được từ các loại vi sinh vật khác nhau. Hbo và Abr lần lượt là *Halomonas boliviensis* và *Azospirillum brasilense*.

Mô tả chi tiết sáng chế

Phương pháp theo sáng chế là phương pháp sản xuất hợp chất đích bao gồm bước nuôi cấy vi khuẩn có khả năng tạo ra hợp chất đích trong môi trường nuôi cấy chứa xyloza dùng làm nguồn cacbon để tạo ra và tích lũy hợp chất đích trong môi trường này, và thu gom hợp chất đích từ môi trường, trong đó:

hợp chất đích là axit 2-ketoglutaric hoặc dẫn xuất của nó, và

vi khuẩn có khả năng sản xuất axit xylonic từ xyloza, và hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza đã được truyền hoặc làm tăng trong vi khuẩn.

Hợp chất đích là axit 2-ketoglutaric (axit α -ketoglutaric, α KG) hoặc dẫn xuất của nó. Các ví dụ về dẫn xuất của axit 2-ketoglutaric bao gồm axit L-glutamic, L-glutamin, L-arginin, L-xitruilin, L-ornithin, L-prolin, putresxin, và axit γ -aminobutyric.

"Khả năng sản xuất hợp chất đích" có nghĩa là khả năng của vi khuẩn được sử dụng theo sáng chế để tạo ra hợp chất đích theo cách sao cho hợp chất đích có thể được thu gom từ các tế bào hoặc môi trường, khi được nuôi cấy trong môi trường này, tốt hơn là khả năng sản xuất hợp chất đích với lượng lớn hơn so với lượng có thể thu

được bằng chủng kiểu đại hoặc chủng không được biến đổi được nuôi cấy trong cùng điều kiện. Vi khuẩn có thể có khả năng sản xuất hai hoặc nhiều loại trong các hợp chất đích.

Hợp chất đích bao gồm hợp chất như hợp chất đích ở dạng tự do và/hoặc muối của nó, ví dụ, sulfat, hydroclorua, cacbonat, muối amoni, muối natri, muối kali, và v.v..

<1> Vi khuẩn theo sáng chế

Vi khuẩn theo sáng chế là vi khuẩn có khả năng tạo ra axit xylonic từ xyloza, mà hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza, và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza được truyền, hoặc trong đó các hoạt tính này được làm tăng.

Vi khuẩn theo sáng chế có thể là vi khuẩn vốn không có hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza, và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza, nhưng được truyền hoạt tính enzym này, hoặc vi khuẩn vốn có những hoạt tính enzym này trong đó những hoạt tính enzym này được tăng cường.

Khả năng sản xuất axit xylonic từ xyloza đạt được bằng, ví dụ, một hoặc cả hai cách sau:

- 1) truyền hoặc làm tăng hoạt tính xyloza dehydrogenaza, và
- 2) có hoạt tính glucoza dehydrogenaza mà có thể xúc tác phản ứng sản xuất axit xylonic từ xyloza.

Các ví dụ về các vi sinh vật mà có nghĩa 1) có thể được sử dụng bao gồm vi khuẩn *Escherichia* và vi khuẩn coryneform, và các ví dụ về vi sinh vật có đặc tính 2) bao gồm vi khuẩn *Pantoea*, và v.v..

Đối với 1), bên cạnh hoạt tính xyloza dehydrogenaza, hoạt tính xylonolactonaza cũng có thể được làm tăng.

Vi khuẩn thuộc những chi này sẽ được giải thích dưới đây.

Axit xylonic tạo ra được từ xyloza được chuyển hóa thành axit 2-ketoglutaric bởi xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza. Quá trình mà trong đó xyloza được chuyển hóa thành axit

2-ketoglutaric bởi xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza cũng được gọi là quá trình Weimberg (J. Biol. Chem., 236:629-636). Trong bản mô tả này, quá trình Weimberg và, bên cạnh quá trình Weimberg, quá trình mà trong đó xyloza được chuyển hóa thành axit xylonic bởi xyloza dehydrogenaza và/hoặc xylonolactonaza có thể được đề cập chung đến dưới dạng quá trình NXA (đồng hóa xyloza mới - Novel Xyloza Assimilation).

Cho dù vi khuẩn có quá trình Weimberg, hay quá trình này đã được đưa vào vi khuẩn có thể được xác định bằng cách đo hoạt tính enzym của xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza trong dịch chiết vi khuẩn, hoặc khẳng định khả năng đồng hóa axit xylonic, mà được tích lũy ở chủng không có quá trình Weimberg. Hơn thế nữa, cho dù vi khuẩn có quá trình NXA hoặc quá trình này được đưa vào vi khuẩn có thể được xác định bằng cách đo hoạt tính enzym của xyloza dehydrogenaza và/hoặc xylonolactonaza bên cạnh các enzym nêu trên. Hơn thế nữa, những hoạt tính enzym này cũng có thể được xác định bằng cách đo axit xylonic tạo ra được từ xyloza.

Theo sáng chế, xylonat dehydrataza là enzym mà xúc tác ngược phản ứng sau (EC4.2.1.82), và cũng được gọi là D-xylo-aldonat dehydrataza, D-xylonat dehydrataza, hoặc D-xylonat hydro-lyaza.

Axit D-Xylonic $\rightarrow$ 2-dehydro-3-deoxy-D-xylonat + H₂O

Hoạt tính xylonat dehydrataza có thể được đo bằng cách, ví dụ, trộn dung dịch axit D-xylonic và mẫu thử nghiệm cho phép phản ứng này xảy ra, sau đó kết thúc phản ứng này bằng cách bổ sung dung dịch làm ngừng phản ứng chứa dung dịch nước 1% semicarbazit hydroclorua và dung dịch nước 1,5% natri axetat, và đo độ hấp thụ của dung dịch phản ứng loãng ở 250 nm (Dahms, A.S., et al., Methods Enzymol., 1982, 90 Pt E:302-5).

Theo sáng chế, 2-keto-3-deoxyxylonat dehydrataza (2-keto-3-deoxy-xylonat dehydrataza) là enzym mà xúc tác ngược phản ứng sau (EC4.2.1-).

2-Dehydro-3-deoxy-D-xylonat $\rightarrow$ 2-oxoglutaric semialdehyt + H₂O

Hoạt tính 2-keto-3-deoxyxylonat dehydrataza có thể được đo bằng cách, ví dụ, trộn dung dịch chứa axit 2-keto-3-deoxyxylonic làm cơ chất và mẫu thử nghiệm cho

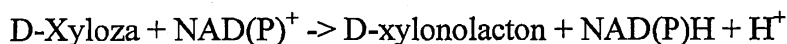
phép phản ứng này xảy ra, và sau đó đo sự suy giảm axit 2-keto-3-deoxyxylonic.

Theo sáng chế, 2-ketoglutaric semialdehyt dehydrogenaza là oxidoreductaza mà xúc tác ngược phản ứng sau (EC1.2.1.26).



Hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza có thể được đo bằng cách, ví dụ, đo sự suy giảm NAD(P). Ví dụ, hoạt tính của enzym này có thể được đo bằng cách bổ sung 2-ketoglutaric semialdehyt vào hỗn hợp của axit pyrophosphoric (pH 8,5), NAD(P), và mẫu thử nghiệm, và đo độ hấp thụ của hỗn hợp phản ứng ở 340 nm (Adams, E., et al., J. Biol. Chem., 1967, 242, 1802-1814).

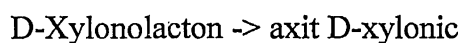
Xyloza dehydrogenaza (D-xyloza-1-dehydrogenaza) là một trong các dismutaza đối với pentoza và axit glucuronic, và là oxidoreductaza mà xúc tác ngược phản ứng sau (EC1.1.1.175).



Hoạt tính D-xyloza-1-dehydrogenaza có thể được đo bằng cách, ví dụ, trộn xyloza, mẫu thử nghiệm, và NAD(P) để cho phép phản ứng xảy ra, và đo độ hấp thụ của hỗn hợp phản ứng ở 340 nm (Stephens, C. et al., J. Bacteriol., 2007, 189(5):181-2185).

Xyloza dehydrogenaza của *Caulobacter crescentus* có thể xúc tác cho phản ứng chuyển hóa D-xyloza thành axit xylonic.

Theo sáng chế, xylonolactonaza là enzym mà xúc tác ngược phản ứng sau (EC3.1.1.68).



Hoạt tính xylonolactonaza có thể được đo bằng cách, ví dụ, trộn xylonolacton và mẫu thử nghiệm cho phép phản ứng này xảy ra, và định lượng xylonolacton còn lại theo phương pháp hydroxamat (Appl. Microbiol. Biotechnol., 29:375-379, 1988; Appl. Microbiol., Biotechnol., 27:333-336, 1988).

Các gen mã hóa các enzym này, xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza, có thể là các gen của vi sinh vật bất kỳ có quá trình Weimberg, và các ví dụ bao gồm, ví dụ, các gen thu được

từ các vi sinh vật như vi khuẩn thuộc họ *Caulobacter*, *Escherichia*, *Agrobacterium*, *Herbaspirillum*, *Actinoplanes*, *Cupriavidus*, *Pseudomonas*, *Zobellia*, *Thermobacillus*, *Arthrobacter*, *Azospirillum*, *Halomonas*, *Bacillus*, hoặc nấm sợi thuộc họ *Aspergillus*.

Các ví dụ về vi khuẩn *Caulobacter* bao gồm *Caulobacter crescentus*.

Dùng làm *Caulobacter crescentus*, chủng CB-15 và chủng CB-13 là đã biết, và được bảo quản ở Viện giống chuẩn Mỹ (Address: P.O. Box 1549, Manassas, VA 20108, United States of America) lần lượt dưới dạng ATCC 19089 và ATCC 33532. Hơn thế nữa, chủng NA-1000 (J. Bacteriol., 192:3678-88, 2010) và chủng K31 cũng có thể được sử dụng.

Các trình tự của hệ gen của *Caulobacter crescentus* CB15, NA1000, và các chủng K31 lần lượt được đăng ký dưới dạng số đăng ký Ngân hàng gen AE005673, CP001340, và CP000927.

Các gen của các enzym của các chủng *Caulobacter crescentus* CB15, NA1000, và K31 được đăng ký ở Ngân hàng gen bằng cách ký hiệu gen sau.

Bảng 1

Gen	Enzym	Số EC	Ký hiệu gen (Ngân hàng gen)		
			CB15	NA1000	K31
<i>xylD</i>	Xylonat dehydrataza	EC4.2.1.82	CC_082 3	CCNA_0086 6	Caul_4000
<i>xylX</i>	2-keto-3-deoxyxylonat dehydrataza	EC:4.2.1	CC_082 2	CCNA_0086 5	
<i>xylA</i>	2-Ketoglutaric semialdehyt dehydrogenaza	EC:1.2.1.26	CC_082 1	CCNA_0086 4	Caul_4001
<i>xylB</i>	Xyloza dehydrogenaza	EC:1.1.1.175	CC_082 0	CCNA_0086 3	Caul_4002
<i>xylC</i>	Xylonolactonaza	EC:3.1.1.68	CC_081 9	CCNA_0086 2	Caul_4003

Ngoài ra, gen xyloza dehydrogenaza và gen 2-ketoglutaric semialdehyt dehydrogenaza của *Caulobacter crescentus* lần lượt có thể được đề cập ở dạng *ccrxylB* và *ccrxylA*.

Hơn thế nữa, các ví dụ về các enzym của quá trình NXA (Quá trình đồng hóa xyloza mới - Novel Xyloza Assimilation) bao gồm, bên cạnh enzym của vi khuẩn *Caulobacter*, ví dụ, xyloza dehydrogenaza của *Hypocrea jecorina* (*Trichoderma reesei*) (FEMS Microbiol. Lett., 277, 249-254, 2007); *ycbD* của *Bacillus subtilis* (2-ketoglutaric semialdehyt dehydrogenaza, loại III); 2-ketoglutaric semialdehyt dehydrogenaza (loại II) của *Pseudomonas putida*; 2-ketoglutaric semialdehyt dehydrogenaza, loại I, loại II, loại III của *Azospirillum brasilense* (J. Bac. Chem., 282, 6685-6695, 2007 đối với các loại này), và các đồng đẳng của chúng.

Cụ thể, đối với gen xylonat dehydrataza, gen *yjhG* và gen *yagF* của vi khuẩn thuộc họ Escherichia như *Escherichia coli* có thể được sử dụng. Gen *yjhG* của *Escherichia coli* được thể hiện trong SEQ ID NO: 34, và gen *yagF* của *Escherichia coli* được thể hiện trong SEQ ID NO: 36. Hơn thế nữa, đối với xylonat dehydrataza, các đồng đẳng gen *xylD* của các vi sinh vật thuộc họ *Agrobacterium*, *Herbaspirillum*, *Actinoplanes*, hoặc *Aspergillus*, như *Agrobacterium tumefaciens*, *Herbaspirillum seropedicae*, *Actinoplanes missouriensis*, và *Aspergillus oryzae*, cũng có thể được sử dụng.

Hơn thế nữa, đối với 2-keto-deoxyxylonat dehydrataza, các đồng đẳng của gen *xylX* của vi khuẩn thuộc họ *Agrobacterium*, *Pseudomonas*, *Zobellia*, *Thermobacillus*, hoặc *Arthrobacter*, như *Agrobacterium tumefaciens*, *Cupriavidus necator*, *Pseudomonas elodea*, *Zobellia galactanivorans*, *Thermobacillus composti*, và *Arthrobacter globiformis*, cũng có thể được sử dụng.

Hơn thế nữa, đối với 2-ketoglutaric semialdehyt dehydrogenaza, các gen của vi khuẩn thuộc họ *Azospirillum*, *Halomonas*, hoặc *Bacillus*, như các đồng đẳng gen *xylA* của *Azospirillum brasilense* và *Halomonas boliviensis*, và *ycbD* of *Bacillus subtilis*, cũng có thể được sử dụng.

Trình tự nucleotit của các gen và các trình tự axit amin được đề cập trên đây được mã hóa bởi chúng được thể hiện trong bảng 11.

Trong *Caulobacter crescentus*, các gen của năm enzym của quá trình NXA

cấu thành cấu trúc operon như được mô tả dưới đây. Trình tự nucleotit của operon này được đăng ký ở Ngân hàng gen với số đăng ký AAK22808_ *Caulobacter crescentus*. Trình tự nucleotit của operon này được thể hiện trong SEQ ID NO: 23. Trình tự axit amin của 2-keto-3-deoxyxylonat dehydrataza, 2-ketoglutaric semialdehyt dehydrogenaza, xyloza dehydrogenaza, và xylonolactonaza được mã hóa bởi operon này được thể hiện trong các SEQ ID NO: 24 đến 27. Hơn thế nữa, trình tự nucleotit của gen *xylD* trong operon này, và các trình tự axit amin của xylonat dehydrataza được mã hóa bởi nó lần lượt được thể hiện trong SEQ ID NOS: 28 và 29. Trình tự nucleotit của SEQ ID NO: 28 tương ứng với các vị trí 5509 đến 7296 của trình tự của SEQ ID NO: 23.

Mặc dù, hai vị trí được gợi ý ở dạng codon khởi đầu của *xylX*, các vị trí từ 1175 đến 1177 được mô tả dưới dạng codon khởi đầu trong SEQ ID NO: 23. Hai codon khởi đầu cũng được gợi ý cho *xylD*, và các vị trí từ 1 đến 3 hoặc các vị trí từ 13 đến 15 của SEQ ID NO: 28 có thể được sử dụng làm codon khởi đầu. Khi các vị trí từ 13 đến 15 được xem là codon khởi đầu, trình tự axit amin của xylonat dehydrataza của SEQ ID NO: 29 bắt đầu từ Leu của vị trí 5.

Mặc dù glucoza dehydrogenaza là enzym gốc mà xúc tác ngược phản ứng dưới đây (EC1.1.1.119), "glucoza dehydrogenaza mà có thể xúc tác phản ứng sản xuất axit xylonic từ xyloza" có nghĩa là enzym mà có thể chuyển hóa D-xyloza thành D-xylonolacton bằng cách sử dụng pyroloquinolin quinon làm coenzym.



Pyroloquinolin quinon có thể được tạo ra bởi khả năng có được ban đầu bởi vi sinh vật, hoặc có thể được bổ sung vào môi trường (Appl. Environ. Microbiol., 2009 May, 75(9) 2784-2791).

Theo sáng chế, đối với vi khuẩn mà tạo ra axit 2-ketoglutaric hoặc dẫn xuất của nó, mong muốn rằng quá trình phân hủy của axit 2-ketoglutaric bị suy giảm hoặc mất đi. Đối với sự suy giảm hoặc mất đi quá trình phân hủy của axit 2-ketoglutaric, mong muốn rằng các hoạt tính hoặc hoạt tính của α -ketoglutarat dehydrogenaza và/hoặc succinat dehydrogenaza giảm đi hoặc mất đi. Theo sáng chế, hoạt tính α -ketoglutarat dehydrogenaza (dưới đây cũng được đề cập dưới dạng " α -KGDH") có nghĩa là hoạt tính xúc tác phản ứng khử carboxyl bằng cách oxy hóa axit α -

ketoglutaric (axit 2-oxoglutaric) để tạo ra succinyl-CoA. Phản ứng được đề cập trên đây được xúc tác bởi ba loại enzym, α -KGDH (E1o, α -ketoglutarat dehydrogenaza, EC:1.2.4.2), dihydrolipoamit S-succinyltransferaza (E2o, EC: 2.3.1.61), và dihydrolipoamit dehydrogenaza (E3, EC:1.8.1.4). Đó là, ba loại phân nhóm này lần lượt xúc tác các phản ứng dưới đây, và hoạt tính xúc tác phản ứng bao gồm sự kết hợp của ba loại phản ứng được gọi là hoạt tính α -KGDH. Hoạt tính α -KGDH này có thể được khẳng định bằng cách đo theo phương pháp của Shioo và các đồng tác giả (Isamu Shioo và Kyoko Ujigawa-Takeda, Agric. Biol. Chem., 44 (8), 1897-1904, 1980).

E1o: 2-oxoglutarat + [succinyltransferaza có gốc dihydrolipoyl] lysin = [succinyltransferaza có gốc dihydrolipoyl] S-succinyl-dihydrolipoyl lysin + CO₂

E2o: CoA + enzym N6-(S-succinyl-dihydrolipoyl) lysin = succinyl-CoA + enzym N6-(dihydrolipoyl) lysin

E3: protein N6-(dihydrolipoyl) lysin + NAD⁺ = protein N6-(lipoyl) lysin + NADH + H⁺

α -KGDH cũng được gọi là oxoglutarat dehydrogenaza hoặc 2-oxoglutarat dehydrogenaza.

Ở vi khuẩn *Enterobacteriaceae* như *Pantoea ananatis*, các phân nhóm protein có ba loại hoạt tính enzym này lần lượt tạo ra phức chất. Các phân nhóm này lần lượt được mã hóa bởi *sucA*, *sucB* và *lpd*, và các gen *sucA* và *sucB* có mặt ở phía sau từ gen succinat dehydrogenaza sắt-lưu huỳnh protein (*sdhB*) (patent Mỹ số 6,331,419). Mặc dù các gen này được mô tả dưới dạng các gen của *Enterobacter agglomerans* AJ13355 trong patent nêu trên, chúng này sau đó được phân loại lại vào *Pantoea ananatis*.

Để làm gen mã hóa α -KGDH của vi khuẩn đường ruột, trình tự các nucleotit của gen *sucA*, gen *sucB* và gen *sucC* có mặt ở vùng phía sau và các trình tự axit amin của các phân nhóm của *Pantoea ananatis* được bộc lộ trong Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 2100957 A1. Hơn thế nữa, các gen *sucA*, *sucB* và *sucC* mã hóa α -KGDH của *Escherichia coli* được công khai cho công chúng lần lượt với số đăng ký Ngân hàng gen NP_415254 và NP_415255.

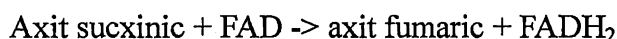
Trong vi khuẩn coryneform, phân nhóm E1o được mã hóa bởi gen *odhA* (được đăng ký dưới dạng NCgl1084 có số đăng ký NC_003450, mà cũng được gọi là gen *sucA*), và phân nhóm E3 được mã hóa bởi gen *lpd* (số đăng ký Y16642). Mặt khác,

ước tính rằng phân nhóm E2o được mã hóa bởi gen *odhA* cùng với phân nhóm E1o làm protein hai chức năng (Usuda et al., Microbiology, 142, 3347-3354, 1996), hoặc được mã hóa bởi gen được đăng ký dưới dạng NCgl2126 có số đăng ký NC_003450, mà khác so với gen *odhA*. Do đó, theo sáng chế, mặc dù gen *odhA* là gen mã hóa phân nhóm E1o, nó cũng mã hóa E2o.

Trình tự nucleotit của gen *odhA* của *Brevibacterium lactofermentum* và trình tự axit amin của phân nhóm E1o nhờ đó được mã hóa (WO2006/028298), trình tự nucleotit của NCgl2126 có số đăng ký NC_003450 nêu trên và trình tự axit amin của phân nhóm E2o nhờ đó được mã hóa, cũng như trình tự nucleotit của NCgl1084 có số đăng ký NC_003450 nêu trên và trình tự axit amin của phân nhóm E1o nhờ đó được mã hóa được bộc lộ trong Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 2100957 A1.

Theo sáng chế, các gen mã hóa mỗi phân nhóm α -KGDH, và cụm gen chứa chúng có thể được gọi chung là "gen mã hóa α -KGDH".

Succinat dehydrogenaza (dưới đây cũng được đề cập dưới dạng "SDH") là enzym của EC:1.3.99.1, mà xúc tác ngược phản ứng sau. Theo sáng chế, hoạt tính SDH có nghĩa là hoạt tính xúc tác cho phản ứng này.



SDH bao gồm ba hoặc bốn cấu trúc phân nhóm phụ thuộc vào loại vi sinh vật, và hoạt tính của chúng có thể được làm giảm hoặc mất đi bằng cách biến đổi ít nhất một trong các protein này sao cho nó không thực hiện chức năng thông thường. Cụ thể, SDH bao gồm các phân nhóm sau (tên của các gen mã hóa các phân nhóm được mô tả trong ngoặc đơn), và protein neo màng được mã hóa riêng rẽ bởi *sdhC* hoặc bởi *sdhC* và *sdhD* phụ thuộc vào các họ.

SDHA: phân nhóm flavoprotein (*sdhA*)

SDHB: phân nhóm Fe-S protein (*sdhB*)

SDHC: protein neo màng (*sdhC*)

SDHD: protein neo màng (*sdhD*)

Hơn thế nữa, phức chất phân nhóm SDH có thể có hoạt tính của cả SDH và fumarat reductaza. Ví dụ, phức chất phân nhóm SDH của vi khuẩn coryneform có hoạt

tính của cả SDH và fumarat reductaza (WO2005/021770).

Hoạt tính SDH có thể được khẳng định bằng cách đo sự giảm của 2,6-dicloindophenol (DCIP) dùng làm chỉ số chỉ dẫn. Phương pháp cụ thể được mô tả trong Tatsuki Kurokawa và Junshi Sakamoto, Arch. Microbiol., (2005) 183:317-324.

Theo sáng chế, các gen mã hóa các phân nhóm SDH, và operon chứa chúng có thể được gọi chung là "các gen mã hóa SDH."

Để làm gen mã hóa SDH của vi khuẩn đường ruột, các trình tự nucleotit của các gen của *Pantoea ananatis* này và các trình tự axit amin của các phân nhóm này được bộc lộ trong WO2008/075483.

Để làm các gen mã hóa SDH của vi khuẩn coryneform, ví dụ, đã bộc lộ các trình tự của operon *sdh* *Corynebacterium glutamicum* (số đăng ký NCgl0359 (*sdhC*) NCgl0360 (*sdhA*) NCgl0361 (*sdhB*)), và operon *sdh* *Brevibacterium flavum* (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2005-095169, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 1672077 A1, WO2008/075483).

Để làm giảm hoặc mất hoạt tính của α -KGDH và SDH, các phương pháp được mô tả dưới đây để làm giảm hoạt tính của enzym mà xúc tác phản ứng phân nhánh từ quá trình sinh tổng hợp axit L-glutamic, và tạo ra các hợp chất khác có thể được sử dụng.

Hơn thế nữa, trong vi sinh vật theo sáng chế, hoạt tính của enzym mà kết hợp xyloza vào các tế bào có thể được làm tăng hơn nữa.

Các ví dụ về enzym mà kết hợp xyloza vào tế bào bao gồm D-xyloza permeaza, và các ví dụ về gen mã hóa D-xyloza permeaza bao gồm gen *xylE*. Trình tự nucleotit của gen *xylE* của *Escherichia coli* mã hóa D-xyloza permeaza, và trình tự axit amin được mã hóa bởi gen này lần lượt được thể hiện trong các SEQ ID NO: 30 và 31.

Hơn thế nữa, mong muốn rằng, trong vi sinh vật theo sáng chế, xyloza isomeraza (*xylA*) và xyluloza kinaza (*xylB*) là giảm. Gen xyloza isomeraza (*xylA*) và gen xyluloza kinaza (*xylB*) của *Escherichia coli* lần lượt được bộc lộ dưới dạng NC000913.1 gi:16131436 và 16131435.

Theo sáng chế, axit xylonic có thể được tích lũy trong môi trường này, và cụ

thể là trong *Escherichia coli*, mong muốn rằng hoạt tính của xylonat dehydrataza được làm tăng hơn nữa. Ví dụ, tốt hơn là làm tăng để thể hiện hoạt tính của chúng với lượng 10 $\mu\text{mol}/\text{min}/\text{mg}$ protein hoặc cao hơn, tốt hơn nữa là 15 $\mu\text{mol}/\text{min}/\text{mg}$ protein hoặc cao hơn, tốt vẫn hơn nữa là 17 $\mu\text{mol}/\text{min}/\text{mg}$ protein hoặc cao hơn.

Các phương pháp truyền hoạt tính của enzym đích vào vi sinh vật hoặc làm tăng hoạt tính của enzym đích của vi sinh vật sẽ được giải thích dưới đây.

Khi vi sinh vật không có hoạt tính ban đầu của enzym đích, hoạt tính của enzym đích có thể được truyền cho vi sinh vật bằng cách đưa gen của enzym đích vào vi sinh vật đó. Hơn thế nữa, khi vi sinh vật có hoạt tính của enzym đích, bằng cách đưa gen enzym đích ngoại, làm tăng số bản sao của gen của enzym đích nội sinh, hoặc biến đổi trình tự kiểm soát sự biểu hiện như yếu tố khởi đầu của gen của enzym đích để làm tăng sự biểu hiện của gen, hoạt tính của enzym đích có thể được làm tăng. Theo sáng chế, việc thể hiện "đưa gen của enzym đích" có nghĩa là không chỉ đưa gen của enzym đích vào vi sinh vật mà ban đầu không có hoạt tính của enzym đích, mà còn đưa gen enzym đích ngoại vào vi sinh vật có hoạt tính của enzym đích, và đưa gen của enzym đích nội sinh vào vi sinh vật có hoạt tính của enzym đích để làm tăng sự biểu hiện của gen của enzym đích nội sinh đó.

Để đưa gen của enzym đích, ví dụ, gen của enzym đích được tách dòng thành vector thích hợp, và vi sinh vật chủ được biến nạp với vector thu được này.

Các ví dụ về vector được sử dụng để biến nạp bao gồm plasmit mà có thể tự sao chép trong vi sinh vật được chọn. Các ví dụ về plasmit có thể tự sao chép trong vi sinh vật thuộc về họ *Enterobacteriaceae* bao gồm pUC19, pUC18, pBR322, RSF1010, pHSG299, pHSG298, pHSG399, pHSG398, pSTV28, pSTV29, pTWV228, pTWV229 (các vector pHSG, pSTV và pTWV là có sẵn từ Takara Bio), pMW119, pMW118, pMW219, pMW218 (các vector pMW có sẵn từ Nippon Gene), và v.v.. Hơn thế nữa, các plasmit cho vi khuẩn coryneform bao gồm pAM330 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-67699), pHM1519 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-77895), pSFK6 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2000-262288), pVK7 (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2003/0175912), pAJ655, pAJ611, pAJ1844 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-192900), pCG1 (Đơn yêu cầu cấp bằng sáng

chế Nhật Bản chưa thẩm định số 57-134500), pCG2 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-35197), pCG4, pCG11 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-183799), pHK4 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 5-7491), và v.v..

Các ví dụ về các phương pháp biến nạp bao gồm xử lý các tế bào nhận bằng canxi clorua để làm tăng khả năng thẩm thấu đối với ADN, mà đã được báo cáo đối với *Escherichia coli* K-12 (Mandel, M. và Higa, A., J. Mol. Biol., 1970, 53:159-162), tạo ra các tế bào khả biến từ các tế bào mà ở pha sinh trưởng, sau đó bằng cách biến nạp bằng ADN, mà đã được báo cáo đối với *Bacillus subtilis* (Duncan, C.H., Wilson, G.A. và Young, F.E., 1977, Gene, 1:153-167), và v.v.. Theo cách khác, phương pháp làm cho các tế bào nhận ADN vào các tế bào nguyên sinh hoặc tế bào trần, mà có thể dễ dàng hấp thu ADN tái tổ hợp, sau đó bằng cách đưa ADN tái tổ hợp này vào các tế bào nhận ADN, mà có thể sử dụng cho *Bacillus subtilis*, xạ khuẩn và nấm men (Chang, S. và Choen, S.N., 1979, Mol. Gen. Genet., 168:111-115; Bibb, M.J., Ward, J.M. và Hopwood, O.A., 1978, Nature, 274:398-400; Hinnen, A., Hicks, J.B. và Fink, G.R., 1978, Proc. Natl. Sci., USA, 75:1929-1933) cũng có thể được sử dụng. Ngoài ra, quá trình biến nạp của các vi sinh vật cũng có thể được tiến hành bằng phương pháp xung điện (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2-207791).

Gen của enzym đích cũng có thể được đưa vào bằng cách đưa gen vào nhiễm sắc thể của vật chủ vi sinh vật. Gen của enzym đích có thể được đưa vào trong nhiễm sắc thể của vi sinh vật bằng phương pháp đưa ngẫu nhiên nó vào nhiễm sắc thể sử dụng đoạn di động hoặc Mini-Mu (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2-109985, patent Mỹ số 5,882,888, Công bố patent châu Âu số 805867 B1), hoặc bằng phương pháp tái tổ hợp đồng nhất sử dụng trình tự có trên ADN nhiễm sắc thể trong nhiều bản sao ở dạng đích. Dùng làm trình tự có trên ADN nhiễm sắc thể trong nhiều bản sao, ADN lặp, và được lặp đảo ở vị trí đầu của yếu tố di động có thể được sử dụng. Theo cách khác, bằng cách sử dụng phương pháp hòa nhập điều khiển bởi Red (WO2005/010175), cũng có thể đưa gen đích vào nhiễm sắc thể. Hơn thế nữa, gen đích cũng có thể được đưa vào nhiễm sắc thể bằng cách tải nạp sử dụng thể thực khuẩn như thể thực khuẩn P1, hoặc bằng cách sử dụng vector vận chuyển liên kết. Hơn thế nữa, cũng có thể đưa gen của enzym đích sử dụng gen không cần thiết cho quá trình sản xuất chất đích làm đích, như được mô tả trong WO03/040373. Một hoặc

nhiều bản sao của gen của enzym đích có thể được đưa vào trong trình tự đích bằng các phương pháp như được mô tả trên đây.

Sự truyền gen đích trên nhiễm sắc thể có thể được khẳng định bằng phương pháp lai Southern sử dụng đoạn dò có trình tự bổ sung với gen đích hoặc một phần của nó.

Mặc dù đầy đủ nếu số bản sao của gen đích được đưa vào là không nhỏ hơn 1, số bản sao tốt hơn là 2 hoặc lớn hơn, tốt hơn nữa là 3 hoặc lớn hơn, vẫn tốt hơn nữa là 5 hoặc lớn hơn. Đối với gen xylonat dehydrataza dùng làm gen đích, cụ thể, tốt hơn là đưa 2 hoặc nhiều bản sao của gen này.

Hơn thế nữa, hoạt tính của gen của enzym đích có thể được tối ưu hóa bằng cách sử dụng kết hợp sự thay thế hoặc đột biến của trình tự kiểm soát sự biểu hiện như yếu tố khởi đầu của gen của enzym đích như được mô tả dưới đây. Cụ thể, tốt hơn là gen xylonat dehydrataza được biểu hiện quá mức bởi sự thay thế hoặc đột biến của trình tự kiểm soát sự biểu hiện thay vì hoặc cùng với sự tăng số bản sao được đề cập trên đây.

Các ví dụ về phương pháp làm tăng sự biểu hiện của gen của enzym đích bao gồm thay thế trình tự kiểm soát sự biểu hiện như yếu tố khởi đầu của gen của enzym đích bằng trình tự có độ mạnh thích hợp trên ADN nhiễm sắc thể hoặc plasmit để làm tăng sự biểu hiện của gen này. Ví dụ, đoạn khởi đầu thr, đoạn khởi đầu lac, đoạn khởi đầu trp, đoạn khởi đầu trc, đoạn khởi đầu pL, đoạn khởi đầu tac, và v.v. là đã biết dưới dạng trình tự thường được sử dụng. Hơn thế nữa, các biến thể của đoạn khởi đầu tac được sử dụng trong các ví dụ được mô tả trên đây (đoạn khởi đầu PtacA, đoạn khởi đầu PtacB) cũng có thể được sử dụng. Các phương pháp đánh giá độ mạnh của các đoạn khởi đầu và các đoạn khởi đầu mạnh được mô tả trong tài liệu của Goldstein và Doi (Goldstein, M.A. và Doi R.H., 1995, Prokaryotic promoters in biotechnology, Biotechnol. Annu. Rev., 1, 105-128), và v.v..

Hơn thế nữa, cũng có thể đưa sự thay thế nucleotit đối với một vài nucleotit vào vùng đoạn khởi đầu của gen để làm biến đổi nó thành đoạn khởi đầu có độ mạnh thích hợp, như được bộc lộ trong Công bố quốc tế WO00/18935. Sự thay thế trình tự kiểm soát sự biểu hiện có thể được tiến hành theo cách tương tự như, ví dụ, phương pháp thay thế gen sử dụng plasmit nhạy nhiệt độ. Các ví dụ về các vectơ có yếu tố gốc

sao chép nhảy nhiệt độ và có thể sử dụng cho *Escherichia coli* và *Pantoea ananatis* bao gồm, ví dụ, plasmid nhảy nhiệt độ pMAN997 được mô tả trong Công bố quốc tế WO99/03988, các dẫn xuất của nó, và v.v.. Hơn thế nữa, sự thay thế trình tự kiểm soát sự biểu hiện cũng có thể được tiến hành bằng phương pháp sử dụng ADN mạch thẳng như phương pháp được gọi là “phương pháp hòa nhập được điều khiển bởi Red” sử dụng recombinaza Red của thể thực khuẩn λ (Datsenko, K.A. và Wanner, B.L., 2000, Proc. Natl. Acad. Sci. USA. 97:6640-6645), và phương pháp sử dụng sự kết hợp của phương pháp hòa nhập được điều khiển bởi Red và hệ phân cắt thể thực khuẩn λ (Cho, E.H., Gumpert, R.I., Gardner, J.F., J. Bacteriol. 184: 5200-5203 (2002)) (đề cập đến WO2005/010175). Sự biến đổi trình tự kiểm soát sự biểu hiện có thể được kết hợp với việc làm tăng số bản sao của gen.

Hơn thế nữa, đã biết rằng sự thay thế của một vài nucleotit trong đoạn đệm giữa vị trí liên kết ribosom (RBS) và codon khởi đầu sự dịch mã, đặc biệt trình tự ở ngay phía trước tính từ codon khởi đầu, ảnh hưởng lớn đến hiệu quả dịch mã mRNA, và do đó trình tự này có thể được biến đổi để cải thiện lượng dịch mã.

Khi gen đích được đưa vào plasmid khuếch đại hoặc nhiễm sắc thể nêu trên, đoạn khởi đầu bất kỳ có thể được sử dụng để biểu hiện gen này miễn sao đoạn khởi đầu được chọn thực hiện chức năng ở vi sinh vật được sử dụng. Đoạn khởi đầu có thể là đoạn khởi đầu nguyên thể đối với gen được sử dụng, hoặc đoạn khởi đầu được biến đổi. Sự biểu hiện của gen cũng có thể được kiểm soát bằng cách lựa chọn thích hợp đoạn khởi đầu mà thực hiện chức năng trong vi sinh vật được sử dụng, hoặc bằng cách làm cho các vùng -35 và -10 của đoạn khởi đầu gần hơn với trình tự tương đồng.

Cho dù hoạt tính của enzym đích có thể được tăng cường hay không, có thể khẳng định được bằng cách so sánh các hoạt tính của enzym đích của chủng được biến đổi và chủng gốc hoặc không được biến đổi. Nếu hoạt tính của enzym đích của chủng được biến đổi được làm tăng so với chủng gốc hoặc chủng không được biến đổi này, hoạt tính của enzym đích được làm tăng. Hơn thế nữa, khi chủng gốc không có hoạt tính của enzym đích, nếu hoạt tính của enzym đích có thể được xác định ở chủng được biến đổi, hoạt tính của enzym đích được làm tăng.

Gen của enzym đích có thể thu được bằng PCR sử dụng các oligonucleotit điều chế được dựa trên thông tin trình tự được đề cập trên đây hoặc thông tin trình tự

của gen hoặc protein đã biết đối với vi sinh vật ở dạng đoạn khởi đầu, hoặc quá trình lai sử dụng oligonucleotit điều chế được dựa trên thông tin trình tự được đề cập trên đây dưới dạng đoạn dò từ ADN nhiễm sắc thể hoặc thư viện ADN nhiễm sắc thể của vi sinh vật có enzym đích.

Hơn thế nữa, enzym đích và gen mã hóa nó có thể là chất đồng đẳng hoặc thể biến đổi nhân tạo của nó, hoặc protein có sự đột biến bảo tồn, hoặc gen mã hóa nó miễn sao hoạt tính enzym được duy trì.

Chất đồng đẳng, thể biến đổi nhân tạo của nó, hoặc protein có sự đột biến bảo tồn hoặc các gen mã hóa chúng được đề cập dưới dạng biến thể bảo tồn.

Biến thể bảo tồn của enzym đích có thể là, ví dụ, protein có trình tự axit amin nêu trên của mỗi enzym, nhưng bao gồm sự thay thế, khuyết đoạn, xen đoạn, bổ sung hoặc hiện tượng tương tự của một hoặc một vài gốc axit amin ở một hoặc một vài vị trí.

Mặc dù số lượng “một hoặc một vài” gốc axit amin có thể khác nhau phụ thuộc vào vị trí trong cấu trúc ba chiều hoặc các loại gốc axit amin của protein, cụ thể, tốt hơn là 1 đến 20, tốt hơn nữa là 1 đến 10, vẫn tốt hơn nữa là 1 đến 5. Đột biến bảo tồn thường là sự thay thế bảo tồn. Sự thay thế bảo tồn là đột biến trong đó đột biến xảy ra lẫn nhau trong số Phe, Trp, và Tyr, nếu vị trí thay thế là axit amin thơm; trong số Leu, Ile và Val, nếu vị trí thay thế là axit amin kỵ nước; giữa Gln và Asn, nếu vị trí thay thế là axit amin phân cực; trong số Lys, Arg và His, nếu vị trí thay thế là axit amin bazơ; giữa Asp và Glu, nếu vị trí thay thế là axit amin axit; và giữa Ser và Thr, nếu vị trí thay thế là axit amin có nhóm hydroxyl. Những sự thay thế được xem là sự thay thế bảo tồn bao gồm, cụ thể, sự thay thế của Ser hoặc Thr cho Ala, sự thay thế của Gln, His hoặc Lys cho Arg, sự thay thế của Glu, Gln, Lys, His hoặc Asp cho Asn, sự thay thế của Asn, Glu hoặc Gln cho Asp, sự thay thế của Ser hoặc Ala cho Cys, sự thay thế của Asn, Glu, Lys, His, Asp hoặc Arg cho Gln, sự thay thế của Gly, Asn, Gln, Lys hoặc Asp cho Glu, sự thay thế của Pro cho Gly, sự thay thế của Asn, Lys, Gln, Arg hoặc Tyr cho His, sự thay thế của Leu, Met, Val hoặc Phe cho Ile, sự thay thế của Ile, Met, Val hoặc Phe cho Leu, sự thay thế của Asn, Glu, Gln, His hoặc Arg cho Lys, sự thay thế của Ile, Leu, Val hoặc Phe cho Met, sự thay thế của Trp, Tyr, Met, Ile hoặc Leu cho Phe, sự thay thế của Thr hoặc Ala cho Ser, sự thay thế của Ser hoặc Ala cho Thr, sự thay thế của Phe hoặc Tyr cho Trp, sự thay thế của His, Phe hoặc Trp cho Tyr, và sự

thay thế của Met, Ile hoặc Leu cho Val. Sự thay thế axit amin, khuyết đoạn, xen đoạn, bổ sung, đảo đoạn được đề cập trên đây hoặc hiện tượng tương tự có thể là kết quả của sự đột biến xảy ra trong tự nhiên hoặc sự biến đổi do sự khác nhau về mặt cá thể hoặc sự khác nhau của các họ vi sinh vật mà từ đó các gen được tạo dẫn xuất (thể đột biến hoặc biến thể). Các protein như vậy có thể thu được bằng cách, ví dụ, biến đổi trình tự nucleotit của gen của enzym đích kiểu đại bằng cách gây đột biến đặc hiệu vị trí sao cho gốc axit amin ở các vị trí đặc hiệu của protein được mã hóa bao gồm sự thay thế, khuyết, chèn, hoặc bổ sung gốc axit amin.

Hơn thế nữa, protein như vậy có sự đột biến bảo tồn như được mô tả trên đây có thể là protein thể hiện độ đồng nhất bằng, ví dụ, 80% hoặc lớn hơn, tốt hơn là 90% hoặc lớn hơn, tốt hơn nữa là 95% hoặc lớn hơn, vẫn tốt hơn nữa là 97% hoặc lớn hơn, tốt hơn nữa là 98% hoặc lớn hơn, đặc biệt tốt hơn nữa là 99% hoặc lớn hơn, so với trình tự axit amin đầy đủ, và có chức năng tương đương của chức năng của protein kiểu dạng. Trong bản mô tả này, "độ đồng nhất" có nghĩa là "sự giống nhau".

Miễn sao gen của enzym đích kiểu đại mã hóa trình tự axit amin này như được mô tả trên đây, không chỉ giới hạn ở các gen của *Caulobacter crescentus*, *Haloferax volcanii*, và tương tự, cũng có thể là các gen bất kỳ có codon tương đương với codon bất kỳ.

Gen loại kiểu đại cũng có thể là ADN mà có thể lai với trình tự nucleotit bổ trợ với trình tự nucleotit của mỗi gen enzym, hoặc đoạn dò mà có thể điều chế được từ trình tự bổ trợ này dưới các điều kiện nghiêm ngặt, và mã hóa protein có các chức năng tương đương với chức năng của enzym đích kiểu đại. "Điều kiện nghiêm ngặt" đề cập đến điều kiện mà xảy ra sự lai đặc hiệu, và lai không đặc hiệu là không được tiến hành. Các ví dụ về điều kiện nghiêm ngặt bao gồm điều kiện mà dưới đó các ADN có độ đồng nhất cao lai với nhau, ví dụ, các ADN có độ đồng nhất không nhỏ hơn 80%, tốt hơn là độ đồng nhất không thấp hơn 90%, tốt hơn nữa là độ đồng nhất không thấp hơn 95%, vẫn tốt hơn nữa là độ đồng nhất không thấp hơn 97%, tốt hơn nữa là độ đồng nhất không nhỏ hơn 98%, đặc biệt tốt hơn nữa là độ đồng nhất không thấp hơn 99%, lai với nhau, và các ADN độ đồng nhất nhỏ hơn giá trị nêu trên không lai với nhau, hoặc các điều kiện tương ứng với việc làm sạch của phương pháp lai Southern thông thường, tức là, các điều kiện làm sạch một lần, tốt hơn là 2 hoặc 3 lần, ở nồng độ muối và nhiệt độ 1 x SSC, 0,1% SDS ở 60°C, tốt hơn là 0,1 x SSC, 0,1% SDS ở 60°C,

tốt hơn nữa là 0,1 x SSC, 0,1% SDS ở 68°C.

Để làm đoạn dò, một phần trình tự mà là bổ sung với gen của enzym đích cũng có thể được sử dụng. Đoạn dò như vậy có thể được điều chế bằng PCR sử dụng các oligonucleotit điều chế được dựa trên trình tự gen đã biết dùng làm đoạn mồi và đoạn ADN chứa trình tự nucleotit dùng làm mẫu. Ví dụ, khi đoạn ADN có độ dài khoảng 300 bp được dùng làm đoạn dò này, các điều kiện làm sạch của phương pháp lai này có thể là, ví dụ, 50°C, 2 x SSC và 0,1% SDS.

Việc mô tả được đề cập trên đây liên quan đến các biến thể bảo tồn của các protein được đề cập trên đây và các gen mã hóa chúng được sử dụng tương tự cho các gen khác được mô tả dưới đây đối với vi khuẩn mà tạo ra các hợp chất đích.

Vi sinh vật được sử dụng trong sáng chế có thể vốn đã có khả năng sản xuất hợp chất đích, hoặc khả năng này có thể được truyền bằng cách nhân giống sử dụng phương pháp đột biến, kỹ thuật ADN tái tổ hợp hoặc các phương pháp tương tự.

Vi sinh vật được sử dụng theo sáng chế bao gồm, nhưng không chỉ giới hạn ở, vi khuẩn thuộc họ *Enterobacteriaceae* như vi khuẩn thuộc các chi *Escherichia*, *Pantoea*, và *Enterobacter*, vi khuẩn coryneform như *Corynebacterium glutamicum* và *Brevibacterium lactofermentum*, và vi khuẩn *Bacillus* như *Bacillus subtilis*.

Theo sáng chế, vi khuẩn coryneform bao gồm các vi khuẩn này được phân loại vào họ *Brevibacterium*, nhưng hiện tại được phân loại vào họ *Corynebacterium* (Int. J. Syst. Bacteriol., 41, 255 (1981)), và bao gồm vi khuẩn thuộc họ *Brevibacterium* liên quan mật thiết với họ *Corynebacterium*. Các ví dụ về vi khuẩn coryneform như vậy được liệt kê dưới đây.

Corynebacterium axetoacidophilum

Corynebacterium axetoglutamicum

Corynebacterium alkanolyticum

Corynebacterium callunae

Corynebacterium glutamicum

Corynebacterium lilium

Corynebacterium melassecola

Corynebacterium thermoaminogenes (*Corynebacterium efficiens*)

Corynebacterium herculis

Brevibacterium divaricatum

Brevibacterium flavum

Brevibacterium immariophilum

Brevibacterium lactofermentum (*Corynebacterium glutamicum*)

Brevibacterium roseum

Brevibacterium saccharolyticum

Brevibacterium thiogenitalis

Corynebacterium amoniacgenes

Brevibacterium album

Brevibacterium cerinum

Microbacterium amoniacphilum

Các ví dụ cụ thể về vi khuẩn bao gồm các vi khuẩn sau.

Corynebacterium axetoacidophilum ATCC 13870

Corynebacterium axetoglutamicum ATCC 15806

Corynebacterium alkanolyticum ATCC 21511

Corynebacterium callunae ATCC 15991

Corynebacterium glutamicum ATCC 13020, ATCC 13032, ATCC 13060

Corynebacterium lilium ATCC 15990

Corynebacterium melassecola ATCC 17965

Corynebacterium thermoaminogenes AJ12340 (FERM BP-1539)

Corynebacterium herculis ATCC 13868

Brevibacterium divaricatum ATCC 14020

Brevibacterium flavum ATCC 13826, ATCC 14067

Brevibacterium immariophilum ATCC 14068

Brevibacterium lactofermentum ATCC 13869 (*Corynebacterium glutamicum* ATCC 13869)

Brevibacterium roseum ATCC 13825

Brevibacterium saccharolyticum ATCC 14066

Brevibacterium thiogenitalis ATCC 19240

Brevibacterium amoniacgenes ATCC 6871, ATCC 6872

Brevibacterium album ATCC 15111

Brevibacterium cerinum ATCC 15112

Microbacterium amoniacphilum ATCC 15354

Các chủng này có sẵn từ Viện giống chuẩn Mỹ (ATCC) (địa chỉ: P.O. Box 1549, Manassas, VA 20108, 1, Hoa kỳ). Đó là, số đăng ký được chỉ định cho mỗi chủng. Các chủng có thể được sắp xếp theo thứ tự sử dụng số đăng ký này (<http://www.atcc.org/>). Số đăng ký được chỉ định cho các chủng được liệt kê trong danh mục của ATCC. Chủng AJ12340 được lưu giữ ngày 27 tháng 10 năm 1987 ở Viện khoa học sinh học và kỹ thuật con người quốc gia thuộc Ủy ban Khoa học và công nghệ Công nghiệp (hiện tại là cơ quan hoạt động độc lập, Viện công nghệ và đánh giá quốc gia, Cơ quan lưu giữ sinh vật độc quyền quốc tế, Tsukuba Central 6, 1-1, Higashi 1-Chome, Tsukuba-shi, Ibaraki-ken, 305-8566, Japan), với số đăng ký FERM BP-1539 dựa trên Hiệp ước Budapest.

Vi sinh vật thuộc họ *Enterobacteriaceae* được sử dụng theo sáng chế bao gồm, nhưng không chỉ giới hạn ở, vi khuẩn thuộc chi *Escherichia*, *Enterobacter*, *Pantoea*, *Klebsiella*, *Serratia*, *Erwinia*, *Salmonella*, *Morganella* hoặc chi tương tự và có khả năng sản xuất hợp chất đích. Đặc biệt, vi khuẩn thuộc họ *Enterobacteriaceae* theo phân loại được thể hiện trong dữ liệu NCBI (Trung tâm thông tin công nghệ sinh học quốc gia) (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?mode=Tree&id=1236&lvl=3&keep=1&srchmode=1&unlock>) có thể được sử dụng. Trong số các vi khuẩn của họ *Enterobacteriaceae*, vi khuẩn thuộc họ *Escherichia*, *Enterobacter*, hoặc *Pantoea* tốt hơn là được sử dụng làm chủng gốc.

Vi khuẩn *Escherichia* mà có thể được sử dụng làm chủng gốc bao gồm, nhưng

không chỉ giới hạn ở, vi khuẩn *Escherichia* được thông báo bởi Neidhardt và các đồng tác giả (Neidhardt, F.C. et al., *Escherichia coli* và *Salmonella Typhimurium*, Hiệp hội vi sinh vật Mỹ, Washington D.C., 1029 Bảng 1), như *Escherichia coli*. Các ví dụ cụ thể về *Escherichia coli* bao gồm chủng *Escherichia coli* W3110 (ATCC 27325), và chủng MG1655 (ATCC 47076), mà thu được từ chủng *Escherichia coli* K12 (kiểu gốc), và v.v..

Cụ thể, *Pantoea bacteria*, *Erwinia bacteria* và *Enterobacter bacteria* được phân loại là γ -proteobacteria, và chủng rất gần về mặt danh pháp với kiểu còn lại (J. Gen. Appl. Microbiol., Dec. 1997, 43(6), 355-361; International Journal of Systematic Bacteriology, Oct. 1997, pp,1061-1067). Trong những năm gần đây, một số vi khuẩn thuộc họ *Enterobacter* được phân loại lại dưới dạng *Pantoea agglomerans*, *Pantoea dispersa*, hoặc họ tương tự, dựa trên cơ sở các thử nghiệm lai AND-ADN, v.v.. (International Journal of Systematic Bacteriology, July 1989, 39(3).p,337-345). Hơn thế nữa, một số vi khuẩn thuộc họ *Erwinia* đã được phân loại lại dưới dạng *Pantoea ananas* hoặc *Pantoea stewartii* (đề cập đến: International Journal of Systematic Bacteriology, Jan. 1993, 43(1), pp,162-173). Ngoài ra, sau đó *Pantoea ananas* cũng đã được phân loại lại dưới dạng *Pantoea ananatis*.

Các ví dụ về *Enterobacter bacteria* bao gồm *Enterobacter agglomerans* (hiện tại được phân loại lại dưới dạng *Pantoea ananatis* v.v.), *Enterobacter aerogenes*, và v.v.. Cụ thể, các chủng được minh họa trong Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 952221 có thể được sử dụng. Chủng tiêu biểu của họ *Enterobacter* là *Enterobacter agglomerans* ATCC 12287 (hiện tại được phân loại lại dưới dạng *Pantoea ananatis*).

Các chủng tiêu biểu của vi khuẩn *Pantoea* bao gồm *Pantoea ananatis*, *Pantoea stewartii*, *Pantoea agglomerans*, và *Pantoea citrea*. Các ví dụ cụ thể bao gồm các chủng sau:

Pantoea ananatis AJ13355 (FERM BP-6614, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 0952221)

Pantoea ananatis AJ13356 (FERM BP-6615, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 0952221)

Mặc dù các chủng này được mô tả dưới dạng *Enterobacter agglomerans*

trong Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 0952221, hiện tại chúng được phân loại ở dạng *Pantoea ananatis* dựa trên sự phân tích trình tự nucleotit của 16S rRNA v.v., như được mô tả trên đây.

Chủ *Pantoea ananatis* AJ13355 được phân lập từ đất ở vùng Iwata-shi, Shizuoka, Japan dưới dạng chủng mà có thể tăng sinh trong môi trường nuôi cấy chứa axit L-glutamic và nguồn cacbon ở độ pH thấp. Chủng SC17 được chọn làm chủng đột biến sản xuất chất đích có độ nhớt thấp từ chủng AJ13355 (patent Mỹ số 6,596,517). Chủng *Pantoea ananatis* AJ13355 được lưu giữ ở Viện khoa học sinh học và kỹ thuật con người quốc gia, Cơ quan khoa học và công nghệ công nghiệp, Bộ thương mại quốc tế và công nghiệp (hiện tại là, Viện công nghệ và đánh giá quốc gia, Cơ quan lưu giữ sinh vật độc quyền quốc tế, địa chỉ: Tsukuba Central 6, 1-1, Higashi 1-Chome, Tsukuba-shi, Ibaraki-ken, 305-8566, Japan) ngày 19 tháng 2 năm 1998 và đăng ký dưới số đăng ký FERM P-16644. Việc lưu giữ này sau đó được chuyển thành lưu giữ quốc tế dưới sự quy định của Hiệp ước Budapest ngày 1 tháng 11 năm 1999 và được chỉ định với số đăng ký FERM BP-6614. Chủng *Pantoea ananatis* SC17 được cung cấp số riêng AJ416, và được lưu giữ ngày 4 tháng 2 năm 2009 tại Viện công nghệ và đánh giá quốc gia, Cơ quan lưu giữ sinh vật độc quyền quốc tế (địa chỉ: Tsukuba Central 6, 1-1, Higashi 1-Chome, Tsukuba-shi, Ibaraki-ken, 305-8566, Japan), và được chỉ định với số đăng ký FERM BP-11091.

Các ví dụ về vi khuẩn *Pantoea ananatis* sản xuất axit L-glutamic còn bao gồm các chủng SC17sucA/RSFCPG+pSTVCB, AJ13601, NP106, và NA1. Chủng SC17sucA/RSFCPG+pSTVCB thu được bằng cách đưa plasmit RSFCPG chứa gen xitrat syntaza (*gltA*), gen phosphoenolpyruvat carboxylaza (*prpC*), và gen glutamat dehydrogenaza (*gdhA*) thu được từ *Escherichia coli*, và plasmit pSTVCB chứa gen xitrat syntaza (*gltA*) thu được từ *Brevibacterium lactofermentum*, vào chủng SC17sucA, mà là chủng khuyết gen *sucA* thu được từ chủng SC17 (patent Mỹ số 6,596,517). Chủng AJ13601 được chọn từ chủng SC17sucA/RSFCPG+pSTVCB vì khả năng của nó kháng axit L-glutamic ở nồng độ cao, độ pH thấp. Hơn thế nữa, chủng NP106 thu được từ chủng AJ13601 bằng cách loại bỏ plasmit RSFCPG+pSTVCB (WO2010/027045). Chủng AJ13601 được lưu giữ tại Viện công nghệ và đánh giá quốc gia, Cơ quan lưu giữ sinh vật độc quyền quốc tế (Tsukuba Central 6, 1-1, Higashi 1-Chome, Tsukuba-shi, Ibaraki-ken, Japan, postal code: 305-

8566) ngày 18 tháng 8 năm 1999, và được chỉ định số đăng ký FERM P-17516. Sau đó, việc lưu giữ này được chuyển sang lưu giữ quốc tế dưới sự quy định của Hiệp ước Budapest ngày 6 tháng 7 năm 2000, và được chỉ định số đăng ký FERM BP-7207. Chủng này được xác định nguồn gốc là *Enterobacter agglomerans* khi nó được phân lập, và được lưu giữ dưới dạng *Enterobacter agglomerans*. Tuy nhiên, gần đây đã được phân loại lại là *Pantoea ananatis* dựa trên sự phân tích trình tự của 16S rRNA và v.v..

Chủng NA1 là chủng tương ứng với chủng NP106 có RSFPPG (WO2008/020654) mà trong đó gen *gltA* của RSFCPPG được mô tả trên đây được thay thế bằng gen metyl xitrat syntaza (*prpC*) (WO2010/027045).

Các ví dụ về vi khuẩn *Erwinia* bao gồm *Erwinia amylovora* và *Erwinia carotovora*, và các ví dụ về vi khuẩn *Klebsiella* bao gồm *Klebsiella planticola*. Các ví dụ cụ thể bao gồm các chủng sau:

Erwinia amylovora ATCC 15580

Erwinia carotovora ATCC 15713

Klebsiella planticola AJ13399 (FERM BP-6600, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 955368)

Klebsiella planticola AJ13410 (FERM BP-6617, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 955368).

Dưới đây, các phương pháp truyền khả năng sản xuất hợp chất đích cho các vi sinh vật như được mô tả trên đây, hoặc các phương pháp làm tăng khả năng sản xuất hợp chất đích của các vi sinh vật như vậy được mô tả.

Để truyền khả năng sản xuất hợp chất đích, các phương pháp thường được sử dụng trong việc nhân giống vi khuẩn coryneform hoặc vi khuẩn thuộc loài *Escherichia* (xem "Lên men axit amin", Gakkai Shuppan Center (Ltd.), chỉnh sửa lần 1, công bố ngày 30 tháng 5 năm 1986, trang 77-100) có thể được sử dụng. Các phương pháp này bao gồm thu nhận thể đột biến khuyết dưỡng, chủng kháng chất tương tự hợp chất đích, hoặc thể đột biến điều hòa quá trình trao đổi chất, cấu trúc chủng tái tổ hợp trong đó sự biểu hiện của enzym sinh tổng hợp hợp chất đích được làm tăng, và v.v.. Trong quá trình nhân giống vi khuẩn sản xuất hợp chất đích, các đặc tính được truyền như đột

biến khuyết dưỡng, khả năng kháng chất tương tự, hoặc đột biến điều hòa quá trình trao đổi chất. Sự biểu hiện của enzym sinh tổng hợp hợp chất đích có thể được làm tăng riêng rẽ hoặc kết hợp của hai hoặc nhiều enzym. Hơn thế nữa, các đặc tính truyền như đột biến khuyết dưỡng, khả năng kháng chất tương tự, hoặc đột biến điều hòa quá trình trao đổi chất có thể được kết hợp với sự gia tăng các enzym sinh tổng hợp.

Chúng đột biến khuyết dưỡng, chúng kháng chất tương tự, hoặc chúng đột biến điều hòa cơ chế trao đổi chất có khả năng sản xuất hợp chất đích có thể thu được bằng cách đưa chủng gốc hoặc chủng kiểu đại trải qua quá trình đột biến thông thường, như tiếp xúc với tia X hoặc chiếu tia UV, hoặc xử lý với chất gây đột biến như N-metyl-N'-nitro-N-nitrosoguanidin, v.v., và sau đó lựa chọn các chủng mà thể hiện thể hiện tính khuyết dưỡng, khả năng kháng chất tương tự, hoặc đột biến điều hòa quá trình trao đổi chất và mà cũng có khả năng sản xuất hợp chất đích. Hơn thế nữa, vi khuẩn sản xuất hợp chất đích cũng có thể thu được bằng cách làm tăng hoạt tính của enzym sinh tổng hợp của hợp chất đích bởi quá trình tái tổ hợp gen.

Dưới đây, các ví dụ về phương pháp truyền khả năng sản xuất hợp chất đích và các vi sinh vật mà khả năng sản xuất hợp chất đích được truyền sẽ được giải thích.

Các ví dụ về phương pháp truyền hoặc làm tăng khả năng sản xuất hợp chất đích bằng cách nhân giống bao gồm, ví dụ, phương pháp biến đổi vi sinh vật sao cho sự biểu hiện của gen mã hóa enzym bao gồm trong quá trình sinh tổng hợp của hợp chất đích được làm tăng. Ví dụ, các ví dụ về enzym bao gồm trong quá trình sinh tổng hợp axit L-glutamic bao gồm glutamat dehydrogenaza (*gdhA*), glutamin synthetaza (*glnA*), glutamat synthetaza (*gltBD*), aconitat hydrataza (*acnA*, *acnB*), xitrat syntaza (*gltA*), phosphoenolpyruvat carboxylaza (*ppc*), pyruvat carboxylaza, pyruvat dehydrogenaza (*aceEF*, *lpdA*), pyruvat kinaza (*pykA*, *pykF*), phosphoenolpyruvat syntaza (*ppsA*), enolaza (*eno*), phosphoglyxeromutaza (*pgmA*, *pgmI*), phosphoglyxerat kinaza (*pgk*), glyxeraldehyt-3-phosphat dehydrogenaza (*gapA*), trioza phosphat isomeraza (*tpiA*), fructoza bisphosphat aldolaza (*fbp*), phosphofructokinaza (*pfkA*, *pfkB*), glucoza phosphat isomeraza (*pgi*), metyl xitrat syntaza (*prpC*), và v.v.. Tên của gen được mô tả trong ngoặc đơn sau tên enzym (điều tương tự cũng được áp dụng cho phần mô tả dưới đây).

Sự biểu hiện của các gen được đề cập trên đây có thể được làm tăng bằng

phương pháp được mô tả để làm tăng các hoạt tính của các enzym của quá trình NXA được mô tả trên đây.

Các ví dụ về các vi sinh vật được biến đổi sao cho sự biểu hiện của gen xitrat syntaza, gen pyruvat dehydrogenaza, và/hoặc gen glutamat dehydrogenaza trong số các gen của enzym được đề cập trên đây được làm tăng bao gồm các vi sinh vật được mô tả trong WO00/18935, Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 1010755, và v.v..

Hơn thế nữa, sự biến đổi để truyền khả năng sản xuất axit L-glutamic cũng có thể được tiến hành bằng cách làm giảm hoặc làm mất hoạt tính của enzym mà xúc tác phản ứng mà phân nhánh từ quá trình sinh tổng hợp axit L-glutamic và tạo ra hợp chất khác axit L-glutamic. Các ví dụ về các enzym mà xúc tác phản ứng mà phân nhánh từ quá trình sinh tổng hợp axit L-glutamic và tạo ra hợp chất khác axit L-glutamic bao gồm 2-oxoketoglutarat dehydrogenaza, succinat dehydrogenaza, isoxitrat lyaza, axit axetohydroxy syntaza, axetolactat syntaza, format axetyltransferaza, lactat dehydrogenaza, glutamat decarboxylaza, 1-pyrolin dehydrogenaza, axetyl-CoA hydraza (Công bố đơn sáng chế quốc tế WO2006/057450), và v.v..

Để làm giảm hoặc làm mất hoạt tính của enzym đích, sự đột biến có thể được đưa vào gen của enzym trên hệ gen bằng phương pháp gây đột biến thông thường hoặc kỹ thuật tái tổ hợp gen sao cho hoạt tính nội bào của enzym là thấp hoặc giảm. Việc đưa sự đột biến như vậy có thể đạt được bằng cách, ví dụ, sử dụng phương pháp tái tổ hợp di truyền để làm mất gen mã hóa enzym trên hệ gen hoặc để biến đổi trình tự kiểm soát sự biểu hiện như đoạn khởi đầu hoặc trình tự Shine-Dalgarno (SD). Cũng có thể đạt được bằng cách đưa sự đột biến để thay thế axit amin (đột biến sai nghĩa), codon kết thúc (đột biến vô nghĩa), hoặc sự đột biến dịch khung để bổ sung hoặc làm khuyết một hoặc hai nucleotit vào các vùng mã hóa enzym trên hệ gen, hoặc làm mất một phần hoặc hoàn toàn gen này (J. Biol. Chem., 272:8611-8617, 1997). Hoạt tính enzym cũng có thể được làm giảm hoặc làm mất bằng cách cấu trúc gen mã hóa enzym đột biến, mà vùng mã hóa của nó được làm khuyết hoàn toàn hoặc một phần, và thay thế nó cho gen thông thường trên hệ gen bằng quá trình tái tổ hợp đồng nhất hoặc phương pháp tương tự, hoặc bằng cách đưa đoạn di động hoặc yếu tố IS vào gen này.

Ví dụ, để đưa sự đột biến mà làm giảm hoặc làm mất các hoạt tính của enzym

được đề cập trên đây bằng phương pháp tái tổ hợp di truyền, các phương pháp sau được sử dụng. Gen đột biến được tạo ra bằng cách biến đổi một phần trình tự của gen đích sao cho nó không mã hóa enzym mà có thể thực hiện chức năng thông thường, và sau đó vi khuẩn thuộc họ *Enterobacteriaceae* có thể được biến nạp với ADN chứa gen đột biến để gây ra quá trình tái tổ hợp của gen tương ứng trên hệ gen với gen đột biến để thay thế gen đột biến cho gen đích trên hệ gen. Các ví dụ về sự thay thế gen như vậy sử dụng quá trình tái tổ hợp đồng nhất bao gồm phương pháp sử dụng ADN mạch thẳng như phương pháp được gọi là phương pháp hòa nhập được điều khiển bởi Red (Datsenko, K.A, và Wanner, B.L., 2000, Proc. Natl. Acad. Sci. USA, 97:6640-6645), và phương pháp sử dụng phương pháp hòa nhập được điều khiển bởi Red kết hợp với hệ cắt thu được từ thể thực khuẩn λ (Cho, E.H., Gumpert, R.I., Gardner, J.F., 2002, J. Bacteriol., 184:5200-5203, đề cập đến WO2005/010175, Đơn yêu cầu cấp bằng sáng chế Nga số 2006134574), phương pháp sử dụng plasmit chứa thể gốc sao mã nhạy nhiệt độ (patent Mỹ số 6,303,383, Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 05-007491), và v.v.. Hơn thế nữa, quá trình gây đột biến đặc hiệu điểm như vậy dựa trên sự thay thế gen sử dụng quá trình tái tổ hợp đồng nhất cũng có thể được tiến hành bằng cách sử dụng plasmit mà không thể sao chép trong vật chủ.

Hơn thế nữa, khả năng sản xuất axit L-glutamic ở vi khuẩn coryneform cũng có thể đạt được bằng phương pháp làm khuếch đại gen *yggB* (NCgl 1221;NP_600492. Reports small-conductance. [gi:19552490], WO2006/070944), và phương pháp nhập gen *yggB* đột biến mà trong đó sự đột biến được đưa vào vùng mã hóa.

Các ví dụ về phương pháp làm tăng khả năng sản xuất axit L-glutamic bao gồm bước đưa các gen mã hóa D-xyluloza-5-phosphat phosphoketolaza và/hoặc fructoza-6-phosphat phosphoketolaza (chúng được gọi chung là phosphoketolaza). Các ví dụ về các vi sinh vật mà được làm tăng hoạt tính của phosphoketolaza bao gồm các vi sinh vật sau (WO2006/016705):

Brevibacterium lactofermentum ATCC 13869 Δ sucA (pVK9-xfp)

Brevibacterium lactofermentum ATCC 13869 Δ sucA (pVK9-PS2_xpkA)

Khả năng sản xuất axit L-glutamic cũng có thể được truyền bằng cách làm tăng hoạt tính 6-phosphogluconat dehydrataza, hoạt tính 2-keto-3-deoxy-6-phosphogluconat aldolaza, hoặc cả hai hoạt tính đó. Các ví dụ về vi sinh vật mà hoạt

tính 6-phosphogluconat dehydrataza của nó và hoạt tính 2-keto-3-deoxy-6-phosphogluconat aldolaza được làm tăng bao gồm vi sinh vật được bộc lộ trong Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2003-274988. Hơn thế nữa, khả năng sản xuất axit L-glutamic cũng có thể được truyền bằng cách khuếch đại các gen *yhfK* và *ybjL*, mà là các gen tiết axit L-glutamic (WO2005/085419, WO2008/133161).

Dùng làm vi sinh vật sản xuất axit L-glutamic được sử dụng trong sáng chế, vi sinh vật có khả năng tích lũy axit L-glutamic trong môi trường lỏng với lượng lớn hơn nồng độ bão hòa của axit L-glutamic khi được nuôi cấy dưới điều kiện axit (dưới đây cũng được đề cập dưới dạng khả năng tích lũy axit L-glutamic dưới điều kiện axit) có thể được sử dụng. Ví dụ, bằng cách thu chủng mà kháng axit L-glutamic trong môi trường có độ pH thấp được cải thiện bằng phương pháp được mô tả trong Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 1078989, khả năng tích lũy axit L-glutamic với lượng lớn hơn nồng độ bão hòa có thể được truyền.

Dùng làm phương pháp truyền hoặc làm tăng khả năng sản xuất axit L-glutamic khác, cũng có thể đề cập đến các phương pháp truyền khả năng kháng chất tương tự axit hữu cơ, chất ức chế hô hấp hoặc các chất tương tự, và phương pháp truyền tính nhạy cho chất ức chế tổng hợp thành tế bào. Các ví dụ bao gồm, ví dụ, phương pháp truyền khả năng kháng axit monofloaxetic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 50-113209), phương pháp truyền khả năng kháng adenin hoặc khả năng kháng thymine (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-065198), phương pháp làm giảm ureaza (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 52-038088), phương pháp truyền khả năng kháng axit malonic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 52-038088), phương pháp truyền khả năng kháng benzopyron hoặc naphthoquinon (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-1889), phương pháp truyền khả năng kháng HOQNO (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-140895), phương pháp truyền khả năng kháng axit α -ketomalonic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-2689), phương pháp truyền khả năng kháng guanidin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-35981), phương pháp truyền tính nhạy penicillin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 4-88994), và v.v..

Các ví dụ cụ thể về vi khuẩn kháng như vậy bao gồm các chủng sau.

Brevibacterium flavum AJ3949 (FERM BP-2632, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 50-113209)

Corynebacterium glutamicum AJ11628 (FERM P-5736, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-065198)

Brevibacterium flavum AJ11355 (FERM P-5007, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-1889)

Corynebacterium glutamicum AJ11368 (FERM P-5020, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-1889)

Brevibacterium flavum AJ11217 (FERM P-4318, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-2869)

Corynebacterium glutamicum AJ11218 (FERM P-4319, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-2869)

Brevibacterium flavum AJ11564 (FERM BP-5472, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-140895)

Brevibacterium flavum AJ11439 (FERM BP-5136, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-35981)

Corynebacterium glutamicum H7684 (FERM BP-3004, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 04-88994)

Brevibacterium lactofermentum AJ11426 (FERM P-5123, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-048890)

Corynebacterium glutamicum AJ11440 (FERM P-5137, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-048890)

Brevibacterium lactofermentum AJ11796 (FERM P-6402, đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-158192)

Các ví dụ được ưu tiên về các vi sinh vật có khả năng sản xuất L-glutamin là vi khuẩn mà hoạt tính glutamat dehydrogenaza của nó được làm tăng, vi khuẩn mà hoạt tính glutamin synthetaza (*glnA*) của nó được làm tăng, và vi khuẩn mà gen glutaminaza của nó được bị gián đoạn (Đơn yêu cầu cấp bằng sáng chế châu Âu chưa

thẩm định số 1229121 và 1424398). Việc làm tăng hoạt tính glutamin synthetaza cũng có thể đạt được bằng cách làm gián đoạn glutamin adenylyltransferaza (*glnE*) hoặc làm gián đoạn protein đối chứng PII (*glnB*). Hơn thế nữa, chủng thuộc họ *Escherichia* và có glutamin synthetaza đột biến mà trong đó gốc tyrosin ở vị trí 397 được thay thế bằng gốc axit amin khác cũng có thể được minh họa dưới dạng vi khuẩn ưu tiên sản xuất L-glutamin (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2003/0148474).

Phương pháp truyền hoặc làm tăng khả năng sản xuất axit L-glutamic khác là phương pháp truyền khả năng kháng 6-diazo-5-oxo-norleuxin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 3-232497), phương pháp truyền khả năng kháng chất tương tự purin và khả năng kháng methionin sulfoxit (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 61-202694), phương pháp truyền khả năng kháng axit α -ketomalonic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-151495), và v.v.. Các ví dụ cụ thể về vi khuẩn coryneform có khả năng sản xuất axit L-glutamic bao gồm các chủng sau.

Brevibacterium flavum AJ11573 (FERM P-5492, Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-161495)

Brevibacterium flavum AJ11576 (FERM BP-10381, Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-151495)

Brevibacterium flavum AJ12212 (FERM P-8123, Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 61-202694)

Các ví dụ về các vi sinh vật có khả năng sản xuất L-prolin bao gồm, ví dụ, vi khuẩn có γ -glutamyl kinaza mà được khử nhạy quá trình ức chế ngược bởi L-prolin, và vi khuẩn mà hệ phân hủy L-prolin của nó được làm giảm. Phương pháp biến đổi vi khuẩn bằng cách sử dụng ADN mã hóa γ -glutamyl kinaza giảm khả năng ức chế ngược bởi L-prolin được bộc lộ trong tài liệu: Dandekar, A.M., Uratsu S.L., J. Bacteriol., 170, 12:5943-5 (1988). Hơn thế nữa, các ví dụ về phương pháp thu được vi khuẩn mà hệ phân hủy L-prolin của nó được làm giảm bao gồm, ví dụ, phương pháp đưa sự đột biến vào trong gen prolin dehydrogenaza để làm giảm hoạt tính enzym. Ví dụ về vi khuẩn có khả năng sản xuất L-prolin bao gồm chủng *Escherichia coli* NRRL B-12403 và chủng NRRL B-12404 (Patent Anh số 2075056), chủng *Escherichia coli* VKPM B-8012 (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2002/0058315), và

các chủng có plasmit đột biến được bộc lộ trong Patent Đức số 3127361 hoặc plasmit đột biến được bộc lộ trong tài liệu của Bloom F.R. và các đồng tác giả (The 15th Miami Winter Symposium, 1983, p,34).

Hơn thế nữa, các vi sinh vật có khả năng sản xuất L-prolin được ưu tiên cũng bao gồm chủng *Escherichia coli* 702 (VKPMB-8011), mà là chủng kháng 3,4-dehydroxyprolin và azetidine-2-carboxylat, chủng 702ilvA (chủng VKPMB-8012), mà là chủng thiếu hụt ilvA của chủng 702, chủng *E. coli* mà hoạt tính của protein của nó được mã hóa bởi gen b2682, b2683, b1242 hoặc b3434 được làm tăng (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2002-300874), và v.v..

Các ví dụ về các chủng sản xuất L-prolin của vi khuẩn coryneform bao gồm chủng kháng DL-3,4-dehydroprolin (FERM BP-1219, patent Mỹ số 4,224,409), các chủng mà hoạt tính xitrat synthetaza tăng 1,4 lần hoặc lớn hơn so với các chủng gốc của chúng (FERM P-5332, FERM P-5333, FERM P-5342, FERMP-5343, Patent Nhật Bản số 1426823), và chủng mà đặc tính khuyết dưỡng axit axetic được truyền (FERM P-5931).

Các ví dụ về vi sinh vật có khả năng sản xuất L-arginin bao gồm các chủng đột biến *Escherichia coli* có khả năng kháng α -metylmethionin, p-flophenylalanin, D-arginin, arginin hydroxamat, AEC (S-(2-aminoethyl)-xystein), α -metylserin, β -2-thienylalanin, hoặc sulfaguanidin (đề cập đến Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 56-106598). Chủng *Escherichia coli* 237, mà là chủng sản xuất L-arginin chứa N-axetylglutamat syntaza hoạt tính mạnh có sự đột biến về khả năng kháng sự ức chế ngược bởi L-arginin (Đơn yêu cầu cấp bằng sáng chế Nga số 2000117677), cũng là vi khuẩn sản xuất L-arginin được ưu tiên. Chủng 237 được lưu giữ tại Viện vi sinh vật công nghiệp quốc gia Nga (VKPM) (GNII Genetika) ngày 10 tháng 4 năm 2000 với số đăng ký VKPM B-7925, và việc lưu giữ ban đầu được chuyển thành lưu giữ quốc tế dựa trên Hiệp ước Budapest ngày 18 tháng 5 năm 2001. Chủng *Escherichia coli* 382, mà là dẫn xuất của chủng 237 và là chủng sản xuất L-arginin có khả năng sản xuất được cải thiện để đồng hóa axit axetic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2002-017342), cũng có thể được sử dụng. Chủng *Escherichia coli* 382 được lưu giữ tại Viện vi sinh vật công nghiệp quốc gia Nga (VKPM) ngày 10 tháng 4 năm 2000 với số đăng ký VKPM B-7926.

Dùng làm vi sinh vật có khả năng sản xuất L-arginin, các vi sinh vật trong đó lượng biểu hiện của một hoặc nhiều gen mã hóa enzym sinh tổng hợp L-arginin gia tăng cũng có thể được sử dụng. Các ví dụ về enzym sinh tổng hợp L-arginin bao gồm một hoặc nhiều enzym được chọn từ N-axetylglutaminat synthetaza (*argA*), N-axetylglutamyl phosphat reductaza (*argC*), ornithin axetyl transferaza (*argJ*), N-axetylglutamat kinaza (*argB*), axetylornithin transaminaza (*argD*), axetylornithin deaxetylaza (*argE*), ornithin carbamoyl transferaza (*argF*), axit argininosuccinic synthetaza (*argG*), axit argininosuccinic lyaza (*argH*), và carbamoyl phosphat syntaza (*carAB*). Tốt hơn nữa là sử dụng gen N-axetylglutamat syntaza đột biến (*argA*) mã hóa enzym mà trong đó trình tự axit amin tương ứng với các vị trí từ 15 đến 19 của enzym kiểu đại là được thay thế và sự ức chế ngược bởi L-arginin nhờ đó được bỏ qua (Đơn yêu cầu cấp bằng sáng chế châu Âu chưa thẩm định số 1170361).

Mặc dù vi khuẩn coryneform sản xuất L-arginin là không bị giới hạn cụ thể miễn sao vi khuẩn coryneform này có khả năng sản xuất L-arginin là được chọn, các ví dụ bao gồm các chủng kiểu đại của vi khuẩn coryneform; vi khuẩn coryneform kháng các tác nhân nhất định bao gồm các dược chất sulfa, 2-thiazolalanin, axit α -amino- β -hydroxyvaleric, và v.v., vi khuẩn coryneform thể hiện tính khuyết dưỡng đối với L-histidin, L-prolin, L-threonin, L-isoleuxin, L-methionin hoặc L-tryptophan bên cạnh khả năng kháng 2-thiazolalanin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 54-44096); vi khuẩn coryneform kháng axit ketomalonic, axit flomalonic hoặc axit monofloaxetic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 57-18989); vi khuẩn coryneform kháng argininol (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 62-24075); vi khuẩn coryneform kháng X-guanidin (X là dẫn xuất của axit béo hoặc mạch béo, Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2-186995), và v.v..

Vi khuẩn coryneform có khả năng sản xuất L-arginin có thể được lựa chọn làm chủng đột biến kháng 5-azauraxil, 6-azauraxil, 2-thiouraxil, 5-flouraxil, 5-bromouraxil, 5-azaxytosin, 6-azaxytosin và v.v.; chủng đột biến kháng arginin hydroxamat và 2-thiouraxil; chủng đột biến kháng arginin hydroxamat và 6-azauraxil (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 49-126819); chủng kháng chất tương tự histidin hoặc chất tương tự tryptophan (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 52-114092); chủng đột biến khuyết dưỡng ít nhất một trong số methionin,

histidin, threonin, prolin, isoleuxin, lysin, adenin, guanin và uraxil (hoặc tiền chất uraxil) (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 52-99289); chủng đột biến kháng arginin hydroxamat (công bố patent Nhật Bản số 51-6754); chủng đột biến khuyết dưỡng đối với axit succinic hoặc kháng chất tương tự trên cơ sở axit nucleic (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 58-9692); chủng đột biến thiếu khả năng phân hủy arginin, kháng chất đối kháng arginin và canavanin và khuyết dưỡng đối với lysin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 52-8729); chủng đột biến kháng arginin, arginin hydroxamat, homoarginin, D-arginin và canavanin, hoặc kháng arginin hydroxamat và 6-azauraxil (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 53-143288); chủng đột biến kháng canavanin (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 53-3586), hoặc các chủng tương tự.

Các ví dụ cụ thể về vi khuẩn coryneform có khả năng sản xuất L-arginin bao gồm các chủng sau.

Brevibacterium flavum AJ11169 (FERM P-4161)

Brevibacterium lactofermentum AJ12092 (FERM P-7273)

Brevibacterium flavum AJ11336 (FERM P-4939)

Brevibacterium flavum AJ11345 (FERM P-4948)

Brevibacterium lactofermentum AJ12430 (FERM BP-2228)

Hơn thế nữa, chủng thiếu hụt ArgR, mà là tiền chất arginin (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2002/0045223), và chủng mà trong đó hoạt tính glutamin synthetaza được làm tăng (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2005/0014236) cũng có thể được sử dụng.

L-xitruilin và L-ornithin có chung quá trình sinh tổng hợp thông thường với L-arginin, và khả năng sản xuất L-xitruilin và L-ornithin có thể được truyền bằng cách làm tăng các hoạt tính enzym của N-axetylglutamat syntaza (*argA*), N-axetylglutamylphosphat reductaza (*argC*), ornithin axetyltransferaza (*argJ*), N-axetylglutamat kinaza (*argB*), axetylornithin transaminaza (*argD*), và axetylornithin deaxetylaza (*argE*) (WO2006/35831).

Dùng làm axit vi khuẩn sản xuất γ -aminobutyric (GABA, chủng mà hoạt tính

của glutamat decarboxylaza của nó được làm tăng (Microb. Cell Fact., 2010, Nov. 12;9:85; Các axit amin, 2010 Nov., 39(5):1107-16; Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2010/0324258) có thể được sử dụng.

Dùng làm vi khuẩn sản xuất putresxin, chủng mà 4-hydroxybutyrat reductaza, succinyl-CoA reductaza (tạo aldehyt), và 4-hydroxybutyrat dehydrogenaza của nó được làm tăng (WO2011/047101), và chủng mà γ -aminobutyraldehyt dehydrogenaza của nó được làm tăng (FEBS Lett., 2005 Aug. 1, 579 (19):4107-12) có thể được sử dụng.

<2> Phương pháp sản xuất các hợp chất đích

Bằng cách nuôi cấy vi khuẩn như được mô tả trên đây trong môi trường nuôi cấy chứa xyloza dùng làm nguồn cacbon để tạo ra và tích lũy hợp chất đích trong môi trường này, và thu gom hợp chất đích từ môi trường này, hợp chất đích có thể được tạo ra.

Dùng làm môi trường được sử dụng cho quá trình nuôi cấy, môi trường thông thường chứa nguồn cacbon, nguồn nitơ và các muối khoáng cũng như các chất dinh dưỡng hữu cơ dạng vết như các axit amin và vitamin, nếu cần, có thể được sử dụng. Hoặc là môi trường tổng hợp hoặc môi trường tự nhiên có thể được sử dụng.

Dùng làm nguồn cacbon, miễn sao xyloza được chứa, các nguồn cacbon khác, ví dụ, đường như glucoza, glyxerol, fructoza, sucroza, maltoza, mannoza, galactoza, arabinoza, dịch thủy phân tinh bột và mật đường có thể được sử dụng. Ngoài ra, các axit hữu cơ như axit axetic và axit xitric, và các rượu như etanol cũng có thể được sử dụng riêng rẽ hoặc kết hợp với các nguồn cacbon khác.

Mặc dù tỷ lệ trộn của xyloza và các nguồn cacbon khác là không bị giới hạn một cách cụ thể, tỷ lệ của xyloza: nguồn cacbon khác (tỷ lệ trọng lượng) tốt hơn là 1:0,1 đến 100, tốt hơn nữa là 1:0,1 đến 10, tốt hơn nữa là 1:0,1 đến 5, vẫn tốt hơn nữa là 1:1 đến 5, tốt hơn nữa là 1:1 đến 3.

Nồng độ của nguồn cacbon trong môi trường là không bị giới hạn một cách cụ thể miễn sao nồng độ thích hợp để sản xuất hợp chất đích là được chọn. Tuy nhiên, nồng độ của nguồn cacbon trong môi trường tốt hơn là khoảng 0,1 đến 50% trọng lượng/thể tích, tốt hơn nữa là khoảng 0,5 đến 40% trọng lượng/thể tích, đặc biệt tốt

hơn nữa là khoảng 1 đến 30%.

Xyloza hoặc hỗn hợp của xyloza và hexoza như glucoza có thể thu được từ nguồn cấp sinh khối khác nhau mà không được sử dụng hoàn toàn. Các pentoza và hexoza như vậy có thể được giải phóng ra khỏi sinh khối bằng cách thủy phân bằng hơi nước và/hoặc axit đậm đặc, thủy phân bằng axit loãng, thủy phân bằng enzym như xenlulaza, hoặc xử lý bằng kiềm. Khi hợp chất là hợp chất loại xenluloza, xenluloza được thủy phân thành sacarit một cách đồng thời hoặc theo trình tự, và sacarit có thể được sử dụng để sản xuất hợp chất đích. Do hemixenluloza thường dễ dàng thủy phân hơn thành sacarit so với xenluloza, tốt hơn là thủy phân hợp chất loại xenluloza trước, tách các pentoza, và sau đó thủy phân xenluloza bằng cách xử lý với hơi nước, axit, kiềm, xenlulaza, hoặc sự kết hợp của các phương pháp này để tạo ra hexoza.

Xyloza được chứa trong môi trường được sử dụng trong sáng chế cũng có thể được cấp bằng cách chuyển hóa mỗi hexoza thành xyloza (D-xyloza) sử dụng vi sinh vật có quá trình chuyển hóa glucoza, galactoza hoặc arabinoza thành xyloza.

Dùng làm nguồn nitơ, amoniac, ure, các muối amoni như amoni sulfat, amoni cacbonat, amoni clorua, amoni phosphat và amoni axetat, các muối axit nitric và v.v. có thể được sử dụng. Dùng làm các chất dinh dưỡng hữu cơ dạng vết, các axit amin, vitamin, các axit béo, các axit nucleic, các chất chứa các chất nêu trên như pepton, axit casamino, dịch chiết nấm men, sản phẩm phân hủy protein nấm men và v.v. có thể được sử dụng. Khi chúng đột biến khuyết dưỡng mà cần axit amin hoặc các chất tương tự cho sự sinh trưởng của nó được sử dụng, tốt hơn là bổ sung chất dinh dưỡng cần thiết. Dùng làm các muối khoáng, các muối của axit phosphoric, các muối magie, các muối canxi, các muối sắt, các muối mangan và v.v. có thể được sử dụng.

Việc nuôi cấy tốt hơn là được tiến hành trong các điều kiện ưa khí, trong khi đó nhiệt độ lên men được kiểm soát ở 20 đến 45°C, và độ pH từ 3 đến 9. Để điều chỉnh độ pH, các chất kiềm hoặc axit hữu cơ hoặc vô cơ, khí amoniac, và v.v. có thể được sử dụng. Lượng cần thiết của hợp chất đích được tích lũy trong môi trường nuôi cấy hoặc các tế bào tốt hơn là sau 10 đến 120 giờ nuôi cấy dưới các điều kiện như được mô tả trên đây.

Hơn thế nữa, khi hợp chất đích là axit L-glutamic, việc nuôi cấy có thể được tiến hành để tạo ra và tích lũy axit L-glutamic với việc kết tủa axit L-glutamic trong

môi trường nuôi cấy bằng cách sử dụng, dùng làm môi trường, môi trường lỏng được điều chỉnh để thỏa mãn điều kiện mà dưới điều kiện đó axit L-glutamic được kết tủa. Các ví dụ về điều kiện mà dưới điều kiện này axit L-glutamic được kết tủa bao gồm, ví dụ, độ pH từ 5,0 đến 4,0, tốt hơn là từ 4,5 đến 4,0, tốt hơn nữa là từ 4,3 đến 4,0, đặc biệt tốt hơn nữa là từ 4,0. Để thu được một cách đồng thời cả sự cải thiện sự sinh trưởng dưới điều kiện axit và sự kết tủa hiệu quả axit L-glutamic, pH tốt hơn là 5,0 đến 4,0, tốt hơn nữa là 4,5 đến 4,0, vẫn tốt hơn nữa là 4,3 đến 4,0. Việc nuôi cấy có thể được tiến hành ở độ pH được đề cập trên đây trong toàn bộ hoặc một phần quá trình nuôi cấy.

Hợp chất đích được thu gom có thể chứa các tế bào vi sinh vật, các thành phần của môi trường, độ ẩm, và các sản phẩm trao đổi chất của vi sinh vật bên cạnh hợp chất đích. Độ tinh khiết của hợp chất đích thu gom được là 50% hoặc cao hơn, tốt hơn là 85% hoặc cao hơn, đặc biệt tốt hơn nữa là 95% hoặc cao hơn (patent Nhật Bản số 1214636, patent Mỹ số 5,431,933, 4,956,471, 4,777,051, 4,946,654, 5,840,358, 6,238,714, Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2005/0025878).

Hợp chất đích thu được theo sáng chế có thể được thu gom từ môi trường nuôi cấy sau khi hoàn thành việc nuôi cấy bằng cách kết hợp các phương pháp thông thường đã biết như phương pháp nhựa trao đổi ion (Nagai, H. et al., Separation Science và Technology, 39(16), 3691-3710), phương pháp tách màng (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 9-164323 và 9-173792), phương pháp kết tinh (WO2008/078448, WO2008/078646), và các phương pháp khác.

Hơn thế nữa, khi hợp chất đích lắng đọng trong môi trường này, nó có thể được thu gom bằng cách ly tâm, lọc hoặc phương pháp tương tự. Hợp chất đích được lắng đọng trong môi trường và hợp chất đích hòa tan trong môi trường có thể được tách đồng thời sau khi hợp chất đích hòa tan trong môi trường được kết tinh.

Ví dụ thực hiện sáng chế

Dưới đây, sáng chế sẽ được giải thích chi tiết hơn với việc tham khảo các ví dụ.

Thành phần của môi trường được sử dụng trong các ví dụ sau đây được thể hiện dưới đây.

[Môi trường LB]

Bacto trypton	10 g/l
---------------	--------

Dịch chiết nấm men	5 g/l
--------------------	-------

NaCl	5 g/l
------	-------

pH 7,0

[LBGM9]

Các thành phần tương tự như các thành phần của môi trường LB cộng với các thành phần của môi trường tối thiểu (5 g/l glucoza, 2 mM magie sulfat, 3 g/l monokali phosphat, 0,5 g/l natri clorua, 1 g/l amoni clorua, 6 g/l dinatri phosphat)

[Môi trường MSII-Glucoza]

Nhóm A

Glucoza	40 g/l
---------	--------

MgSO ₄ ·7H ₂ O	0,5 g/l
--------------------------------------	---------

Nhóm B

(NH ₄) ₂ SO ₄	20 g/l
---	--------

KH ₂ PO ₄	2 g/l
---------------------------------	-------

NaCl	0,5 g/l
------	---------

Dịch chiết nấm men	2 g/l
--------------------	-------

CaCl ₂ ·7H ₂ O	0,25 g/l
--------------------------------------	----------

FeSO ₄ ·7H ₂ O	20 mg/l
--------------------------------------	---------

MnSO ₄ ·nH ₂ O	20 mg/l
--------------------------------------	---------

Các thành phần ở dạng vết*	4 ml/l
----------------------------	--------

L-Lys	200 mg/l
-------	----------

DL-Met	200 mg/l
--------	----------

DAP	200 mg/l
-----	----------

Các thành phần của các Nhóm A và B được hấp riêng biệt ở 120°C trong 20 phút, và sau đó được trộn.

*Các thành phần ở dạng vết

$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	0,66 g/l
$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$	0,18 g/l
$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$	0,16 g/l
$\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$	0,15 g/l
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	0,18 g/l
H_3BO_3	0,10 g/l
Na_2MoO_4	0,30 g/l

[Môi trường MSII-Xyloza]

Các thành phần tương tự như các thành phần của môi trường MSII-Glucoza ngoại trừ rằng glucoza (40 g/l) được thay thế bằng xyloza (40 g/l)

[Môi trường MSII-GX]

Các thành phần tương tự như các thành phần của môi trường MSII-Glucoza ngoại trừ rằng glucoza (40 g/l) được thay thế bằng hỗn hợp của glucoza (20 g/l) và xyloza (20 g/l)

[Môi trường MSII-SX]

Các thành phần tương tự như các thành phần của môi trường MSII-Glucoza ngoại trừ rằng glucoza (40 g/l) được thay thế bằng hỗn hợp gồm sucroza (20 g/l) và xyloza (20 g/l).

[Môi trường tổng hợp E1]

Nhóm A

NH_4Cl	20 mM
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	2 mM
Na_2HPO_4	40 mM
KH_2PO_4	30 mM
CaCl_2	0,01 mM
$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$	0,01 mM

MnSO₄·4 đến 5H₂O 0,01 mM

Xitrat 5 mM

pH tự do

Lọc thanh trùng

Nhóm B-1

Nguồn cacbon 50 (hoặc 100) mM

Lọc thanh trùng

Nhóm B-2

Thiamin HCl 1 mM

Thành phần này được bổ sung vào thành phần của nhóm B-1 sau khi lọc thanh trùng (0,22 µm).

Nhóm C

MES-NaOH (pH 6,8) 50 mM

Lọc thanh trùng (0,22 µm)

Các dung dịch chứa các thành phần của các nhóm A đến C ở các nồng độ cao gấp 5 lần được điều chế làm dung dịch gốc.

[Môi trường CM-Dex]

Polypepton 10 g/l

Dịch chiết nấm men 10 g/l

Glucosa 5 g/l

KH₂PO₄ 1 g/l

Urea 3 g/l

MgSO₄·7H₂O 0,4 g/l

FeSO₄·7H₂O 0,01 g/l

MnSO₄·5H₂O 0,01 g/l

Dịch lọc đậu nành 1,2 g/l (T-N)

(Dịch thủy phân đậu nành)

pH 7,5 được điều chỉnh bằng KOH

[Môi trường Glc]

Glucosa	80 g/l
$(\text{NH}_4)_2\text{SO}_4$	30 g/l
KH_2PO_4	1 g/l
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	0,4 g/l
$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$	0,01 g/l
$\text{MnSO}_4 \cdot 5\text{H}_2\text{O}$	0,01 g/l
Vitamin B1	200 $\mu\text{g/l}$
Biotin	60 $\mu\text{g/l}$
Dịch lọc đậu nành	0,48 g/l (T-N)

(Dịch thủy phân đậu nành)

pH 8,0 được điều chỉnh bằng KOH

[Môi trường Xyl (biotin được hạn chế)]

Các thành phần tương tự như các thành phần của môi trường Glc ngoại trừ rằng glucosa (80 g/l) được thay thế bằng xyloza (80 g/l), và biotin không được bao gồm.

[Môi trường MS]

Nhóm A

Glucosa hoặc xyloza	40 g/l
Glucosa và xyloza (1:1)	40 g/l
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	1 g/l

Nhóm B

$(\text{NH}_4)_2\text{SO}_4$	20 g/l
KH_2PO_4	1 g/l
Dịch chiết nấm men	2 g/l

$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 10 mg/l

$\text{MnSO}_4 \cdot n\text{H}_2\text{O}$ 10 mg/l

Các thành phần của các Nhóm A và B được hấp riêng biệt ở 120°C trong 20 phút, và sau đó được trộn, và 50 g/l canxi cacbonat theo Dược điển Nhật Bản được bổ sung.

Ví dụ 1: Đưa quá trình NXA vào *Pantoea ananatis*

(1) Cấu trúc plasmit pTWV228Ptac_ccrNXA để đưa vào quá trình NXA

Đối với vi khuẩn đã biết mà quá trình NXA đã được thông báo, *C. crescentus* là đã biết (Stephens, C. et al., J. Bacteriol., 189(5):181-2185, 2007). Để thu được các gen mã hóa các enzym của quá trình NXA của *C. crescentu*, phương pháp sau được sử dụng.

Hệ gen của *C. crescentus* có độ dài khoảng 4 Mb, trong đó năm gen tạo ra cấu trúc operon (Journal of Bacteriology, 189:2181-2185, 2007). Hệ gen được tách ra từ chủng có hệ gen đã biết của *C. crescentus* (CB-15 (ATCC 19089), có sẵn từ ATCC), và sự tách dòng các gen này và việc cấu trúc vector biểu hiện đối với các gen này bị suy giảm.

Vector biểu hiện được cấu trúc bằng cách sử dụng kit tách dòng Clontech In-Fusion.

Bốn loại đoạn ADN sau đây được làm khuếch đại bằng PCR sử dụng ADN nhiễm sắc thể của *C. crescentus* CB-15 (ATCC 19089) vì i), ii) và iii) sau, và pMW119 vì iv) sau làm mẫu. Các đoạn mỗi được sử dụng để PCR được chỉ ra trong ngoặc đơn.

i) trình tự đoạn khởi đầu tac (giờ đây được đề cập đến dưới dạng "Ptac", PtwvPtacf: SEQ ID NO: 1, 0823Ptacr: SEQ ID NO: 2)

ii) Đoạn chứa *xylX*, *ccrxylA*, *ccrxylB* và *xylC* (Ptac0823f: SEQ ID NO: 3, 0819r: SEQ ID NO: 4)

iii) *xylD* và vùng phía sau nó khoảng 120 bp (0819f: SEQ ID NO: 5, 219cc0819r: SEQ ID NO: 6)

iv) pMW119/SmaI (219f: SEQ ID NO: 7, 219r: SEQ ID NO: 8)

Sau đó, bằng PCR sử dụng các sản phẩm PCR tinh khiết của i) và ii) làm mẫu,

cũng như PtwvPtacf và 0819r dùng làm các đoạn môi, đoạn Ptac_xylXccrAccrBC bao gồm các sản phẩm PCR nêu trên được liên kết đồng thời được khuếch đại. Phản ứng dung hợp được tiến hành với ba Ptac_xylXccrAccrBC thu được và sản phẩm PCR của iii) và iv) sử dụng kit tách dòng Clontech In-Fusion, chủng *E. coli* JM109 được biến nạp với sản phẩm phản ứng, và plasmit pMW119 Ptac_ccrNXA đích thu được từ thể biến nạp đó.

Sau đó, bằng cách sử dụng pMW119 Ptac_ccrNXA làm mẫu, cũng như PtwvPtacf và 219CC0819r làm các đoạn môi, operon ccrNXA chứa Ptac được khuếch đại. Phản ứng dung hợp được tiến hành với sản phẩm được khuếch đại thu được và pTWV228 được phân cắt bởi *Sma*I, chủng *E. coli* JM109 được biến nạp với sản phẩm phản ứng, và plasmit đích pTWV228Ptac_ccrNXA thu được từ thể biến nạp đó.

(2) Cấu trúc plasmit pUT-MuKm chứa pUT399 mang Mini-Mu kháng kanamycin

pUT399 là plasmit có gốc sao chép R6K và vùng mob cần thiết để vận chuyển liên kết, và không thể sao ở chủng không có gen *pir* (có sẵn từ Biomedal, đề cập đến tài liệu R. Simon., et al., BIO/TECHNOLOGY NOVEMBER 1983, 784-791; patent Mỹ số 7,090,998).

pCE1134 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2-109985) là plasmit chứa MudII1734, và mang gen kháng Km và gen *lacXYZ* trong đơn vị Mini-Mu. Bằng phương pháp được mô tả dưới đây, đoạn ADN không có vùng *lacXYZ* được điều chế từ đơn vị Mini-Mu của pCE1134, và được tách dòng thành pUT399.

Bằng PCR sử dụng pCE1134 làm mẫu, cũng như các đoạn môi attL-F (SEQ ID NO: 9) và nptII-R (SEQ ID NO: 10), đoạn chứa yếu tố ức chế MuCts của gen *MuAB* mã hóa đầu bên trái và transposaza, và gen kháng Km là thu được. Hơn thế nữa, bằng cách sử dụng pCE1134 làm mẫu, cũng như các đoạn môi attR-F (SEQ ID NO: 11) và attR-R (SEQ ID NO: 12), đoạn chứa đầu bên phải là thu được theo cách tương tự. PCR trao đổi chéo được tiến hành bằng cách sử dụng hai đoạn này làm mẫu, cũng như các đoạn môi attL-F và attR-R, và đoạn thu được khoảng 2,3 kb được đưa vào pUT399 ở vị trí *Sma*I. Theo cách này, plasmit pUT-MuKm là thu được.

Do đơn vị Mini-Mu được cấu trúc như được mô tả trên đây có gen kháng Km và vị trí *Not*I ghi nhận 8 bazơ làm vị trí tách dòng trong đơn vị vận động, các gen khác

nhau có thể được tách dòng thành nó.

(3) Sự thay thế của gen kháng thuốc pTWV228Ptac_ccrNXA

Gen kháng ampicilin pTWV228Ptac_ccrNXA được thay thế bằng gen kháng kanamycin bằng phương pháp λ Red.

Bằng cách sử dụng pUT_MuKm làm mẫu, cũng như các đoạn môi Ap-Km-fw (SEQ ID NO: 13) và Ap-Km-rv (SEQ ID NO: 14), trình tự chứa gen kháng kanamycin (đoạn *ntpII*) được khuếch đại.

PCR này được tiến hành bằng cách sử dụng PrimeSTAR HS Polymeraza (Takara Bio) theo phương pháp được hướng dẫn đối với enzym này.

Plasmid hỗ trợ RSF_Red_TER (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2009/0286290A1, WO2008/075483) được đưa vào *E. coli* JM109 có pTWV228Ptac_ccrNXA, và các tế bào được nuôi cấy trong 50 ml môi trường LB (containing 1 mM IPTG, 100 mg/l ampicilin, và 25 mg/l cloramphenicol) ở 37°C đến khi giá trị OD₆₆₀ bằng 0,4.

RSF_Red_TER được đề cập trên đây là plasmid trợ giúp để gây ra sự hòa nhập phụ thuộc λ (phương pháp hòa nhập được điều khiển bởi Red, phương pháp λ Red), và nó có thể gây cảm ứng sự biểu hiện của các gen của λ *gam*, *bet* và *exo* với gen *lacI*. Plasmid này cũng chứa gen levansucraza (*sacB*), và có thể loại plasmid ra khỏi tế bào có gen này trong môi trường nuôi cấy chứa sucroza. Hơn thế nữa, plasmid này cũng chứa gen kháng cloramphenicol.

Các tế bào được nuôi cấy như được mô tả trên đây được thu gom, được rửa hai lần bằng dung dịch 10% glyxerol bằng cách ly tâm, và tạo huyền phù 1 ml dung dịch 10% glyxerol. Sau đó, các tế bào được biến nạp với đoạn *ntpII* thu được trên đây bằng cách mở lỗ bằng điện, và các thể biến nạp được lựa chọn trong môi trường LB agar chứa 40 mg/l kanamycin. Các thể biến nạp thu được được ủ trong môi trường LB agar (chứa 1 mM IPTG, 10% sucroza, và 40 mg/l kanamycin), và được nuôi cấy qua đêm ở 37°C để thu được dòng đơn. Đã khẳng định được rằng thể biến nạp thu được không thể sinh trưởng trong môi trường LB agar chứa 100 mg/l ampicilin, và nhờ đó đã khẳng định được rằng gen kháng ampicilin pTWV228Ptac_ccrNXA được thay thế bằng gen kháng kanamycin. Plasmid thu được được chỉ định là

pTWVPtac_ccrNXA_Km.

(4) Cấu trúc plasmit chứa *xyID*

Việc cấu trúc được tiến hành bằng cách sử dụng kit tách dòng Clontech In-Fusion.

Đầu tiên, bằng PCR sử dụng plasmit chứa pUC18 trong đó mỗi gen được tách dòng dùng làm mẫu, cũng như *xyID_IFS_5742-10-5* (SEQ ID NO: 15) và *xyID_IFS_5742-10-6* (SEQ ID NO: 16) dùng làm các đoạn môi, đoạn ADN chứa *xyID* được khuếch đại. Cụ thể, nó được tách dòng thành pUC18 mà vị trí *SfiI* của nó được loại bỏ bằng phương pháp được mô tả dưới đây.

Bằng PCR sử dụng ADN hệ gen của chủng *C. crescentus* CB-15 làm mẫu, CC0819-01F_4691-88-7 (SEQ ID NO: 17) và CC0819-01R_5659-9-1 (SEQ ID NO: 18), cũng như CC0819-02F_5659-9-2 (SEQ ID NO: 19) và CC0819-02R_4691-88-10 (SEQ ID NO: 20) dùng làm các đoạn môi, các đoạn 1130 bp và 653 bp lần lượt được khuếch đại. Sau đó, pUC18 được phân cắt bởi *SmaI* và hai đoạn được khuếch đại nêu trên được lắp ghép bằng phương pháp lắp ghép *in vitro* (Nature Methods, 6(5), 343-345, 2009) để thu được pUC18-*xyID* mà trong đó vị trí *SfiI* được loại bỏ, và gen *xyID* được xen đoạn.

Theo cách khác, pSTV28-Ptac-Ttrp được phân cắt bởi *SmaI* bằng phương pháp thông thường. Phản ứng dung hợp được tiến hành với đoạn ADN của gen *xyID* và đoạn vector ADN, chủng *E. coli* JM109 được biến nạp với sản phẩm phản ứng, và plasmit đích pSTVPtac_*xyID*_Ttrp thu được từ thể biến nạp này.

pSTV28-Ptac-Ttrp được cấu trúc như sau.

Đoạn ADN (PtacTtrp) có đoạn khởi đầu tac (có trình tự SEQ ID NO: 32) và trình tự của đoạn kết thúc *trp* được tổng hợp, và được liên kết giữa các vị trí *KpnI*-*BamHI* của vector pMW219 để thu được pMW219-Ptac-Ttrp. Những lượng tương tự của pSTV28 và pMW219-Ptac-Ttrp, cả hai được phân cắt bởi *KpnI* và *BamHI* được trộn, và được liên kết, JM109 được biến nạp với sản phẩm liên kết này, và plasmit được tách ra từ khuẩn lạc mà thể hiện khả năng kháng Cm. Đã khẳng định được rằng plasmit thu được thể hiện các dải khoảng 400 bp và 3 kbp (chính xác, 389 bp và 2994 bp), mà được mong đợi là kết quả của sự phân cắt kép bởi *KpnI* và *BamHI*, và do đó

thu được pSTV28-Ptac-Ttrp.

(4) Sản xuất axit L-glutamic bằng *Pantoea ananatis* được nhập quá trình NXA

Chủng *P. ananatis* NA1 được biến nạp với pTWVPtac_ccrNXA_Km bằng cách mở lỗ bằng điện (đề cập đến patent Mỹ số 6,682,912). Đối với chủng được đưa pTWVPtac_ccrNXA_Km, môi trường trong đĩa chứa LBGM9 được bổ sung kanamycin ở nồng độ cuối bằng 40 mg/l được sử dụng.

Các tế bào của chủng *P. ananatis* NA1 và chủng biến nạp được nuôi cấy qua đêm ở 34°C trên đĩa LBGM9, mỗi đĩa được cạo lấy một lượng tương ứng với 1/6 đĩa, được ủ trong 5 ml môi trường MSII-Xyloza hoặc MSII-GX được chứa trong ống thử nghiệm lớn, và được nuôi cấy ở 34°C và 120 vòng/phút trong 48 giờ, và sacarit dư, lượng axit L-glutamic (Glu), và axit xylonic tích tụ được đo. Các kết quả được thể hiện trong các bảng 2 và 3.

Bảng 2 (Sản xuất Glu trong môi trường MSII-GX)

Chủng	OD660 (x51)	Glc (g/l) được tiêu thụ	Xyl (g/l) được tiêu thụ	Glu (g/l)	Hiệu suất(%)	Axit xylonic (g/l)
NA1	0,129 ± 0,001	21,8	22,9	11,5 ± 0,1	25,7 ± 0,3	18,4 ± 0,2
NA1/pTWV228Pt ac_ccrNXA_Km	0,134 ± 0,006	21,8	22,9	31,3 ± 0,1	69,8 ± 0,1	0,0

Bảng 3 (Sản xuất Glu trong môi trường MSII-Xyloza)

Chủng	OD660 (x51)	Xyl (g/l) tiêu thụ	Glu (g/l)	Hiệu suất(%)	Axit xylonic (g/l)
NA1	0,019 ± 0,006	0,0	0,0	0,0	0,0
NA1/pTWV228Ptac_cc rNXA_Km	0,069 ± 0,008	41,0	32,7 ± 0,9	80,5 ± 2,2	10,1 ± 1,9

Khi chủng *P. ananatis* NA1 được nuôi cấy với nguồn cacbon hỗn hợp của glucoza và xyloza (môi trường MSII-GX), hiệu suất axit glutamic là 25,7% (bảng 2). Trong trường hợp này, sự tích lũy axit xylonic được quan sát, và do đó đã cho thấy rằng hầu hết xyloza được chuyển hóa thành axit xylonic. Ước tính rằng sự tích lũy axit xylonic ở chủng *P. ananatis* NA1 được tạo ra bởi hoạt tính của glucoza dehydrogenaza của *P. ananatis*.

Mặt khác, khi chủng *P. ananatis* NA1 được đưa vào pTWVPtac_ccrNXA_Km được nuôi cấy với nguồn cacbon hỗn hợp của glucoza và xyloza (môi trường MSII-GX), đã cho thấy hiệu suất axit glutamic rất cao so với chủng gốc (hiệu suất: 69,9%). Nếu coi như rằng chủng gốc hầu như không sản xuất axit glutamic từ xyloza, và thấy rằng hiệu suất axit glutamic từ glucoza của chủng được gắn pTWVPtac_ccrNXA_Km là tương đương với chủng gốc, hiệu suất của axit glutamic tạo ra được từ xyloza nhờ quá trình NXA có thể được ước tính khoảng 86%. Thực tế là, khi việc nuôi cấy được tiến hành với xyloza dùng làm nguồn cacbon duy nhất (MSII-Xyloza), chủng được gắn pTWVPtac_ccrNXA_Km tạo ra Glu ở hiệu suất là 80% (bảng 3).

Ví dụ 2: Đưa quá trình NXA vào *Escherichia coli*

(1) Sự biểu hiện của quá trình NXA trong *E. coli*

Bằng cách sử dụng chủng thiếu hụt isoxitrat dehydrogenaza (Δicd), mà là enzym của chu trình TCA và tạo ra α KG từ axit isoxitric, sự biểu hiện của quá trình NXA đạt được nhờ sự sinh trưởng bổ trợ trong môi trường tối thiểu chứa xyloza dùng làm nguồn cacbon duy nhất. Vì chủng thiếu hụt gen *icd* không thể tạo ra α KG, nó không thể sinh trưởng trong môi trường tối thiểu chứa xyloza dùng làm nguồn cacbon duy nhất, nhưng nếu khả năng sản xuất α KG từ xyloza có thể được truyền bằng cách

đưa quá trình NXA, chủng này cần có khả năng sinh trưởng trong môi trường như vậy.

Cụ thể, chủng JW1122, mà là chủng thiếu hụt gen *icd* của Keio Collection (<http://cgsc.biology.yale.edu/Person.php?ID99553>, có sẵn từ Trung tâm nguồn gen *E. coli* ở Yale CGSC, Trung tâm vật liệu di truyền gốc Coli), được sử dụng làm chủng vi khuẩn vật chủ, và bằng cách đưa và biểu hiện quá trình NXA ở chủng đó sử dụng plasmit, được đánh giá xem liệu quá trình NXA cũng có thể thực hiện chức năng trong *E. coli* hay không.

Trong bảng 4, các plasmit được cấu trúc và các kết quả của sự sinh trưởng bổ trợ ở chủng thiếu hụt gen *icd* đã được chỉ ra. Đã khẳng định được rằng chủng này được gắn plasmit pMW119 Ptac_ccrNXA (điều chế được trong ví dụ 1) chứa operon của quá trình NXA (*xylX*, *ccrxylA*, *ccrxylB*, *xylC*, *xylD*) và đoạn khởi đầu tác ở dạng kết hợp có thể sinh trưởng trong môi trường tối thiểu M9 (đĩa) (Sambrook, J. et al., Molecular Cloning, Cold Spring Harbor Laboratory Press (1989)) chứa xyloza dùng làm nguồn cacbon duy nhất.

Nghiên cứu tương tự cũng đã được tiến hành bằng cách nuôi cấy trong môi trường lỏng. Sự sinh trưởng (O.D.) môi trường tối thiểu M9 và môi trường tổng hợp E1 chứa xyloza hoặc α KG dùng làm nguồn cacbon duy nhất được đo theo thời gian bằng cách sử dụng dụng cụ nuôi cấy cho 36 mẫu. Các kết quả được thể hiện trong Fig. 1. Tương tự việc nuôi cấy trên đĩa, chủng được đưa vào quá trình NXA tốt hơn là được sinh trưởng trong môi trường M9 hoặc E1 chứa xyloza dùng làm nguồn cacbon duy nhất, trong khi đó chủng có vectơ đối chứng không sinh trưởng trong môi trường như vậy. Xem như là các kết quả là thu được bởi chủng này sinh trưởng bằng cách đồng hóa xyloza thông qua quá trình NXA, tức là, quá trình NXA thu được từ *C. crescentus* cũng thực hiện được chức năng trong *E. coli*.

Bảng 4

Vector	Đoạn khởi đầu	Gen	Sự sinh trưởng trong môi trường M9-xyloza
pTWV229	Nguyên thể	<i>xylX</i> <i>ccrxyIA</i> <i>ccrxyIB</i> <i>xylC</i> <i>xylD</i>	X
pMW119	Ptac	<i>xylX</i> <i>ccrxyIA</i> <i>ccrxyIB</i> <i>xylC</i> <i>xylD</i>	○
pTWV228	Ptac	<i>xylX</i> <i>ccrxyIA</i> <i>ccrxyIB</i> <i>xylC</i> <i>xylD</i>	○

(2) Sự biểu hiện của quá trình NXA ở chủng sản xuất axit L-glutamic E. coli

Dùng làm chủng sản xuất axit L-glutamic E. coli, MG1655 Δ sucA (Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2005/0106688), mà là chủng thiếu hụt α KGDH, được sử dụng. pMW119 Ptac_ccrNXA hoặc pMW119 dùng làm đối chứng được đưa vào chủng nêu trên để thu được MG1655 Δ sucA/pMW119Ptac_ccrNXA và MG1655 Δ sucA/pMW119. Mỗi chủng này được nuôi cấy trong bình cầu chứa môi trường nuôi cấy MS chứa glucoza (40 g/l), xyloza (40 g/l), hoặc glucoza và xyloza (20 g/l) dùng làm nguồn cacbon. Việc nuôi cấy được tiến hành trong 24 giờ trong môi trường chỉ chứa glucoza dùng làm nguồn cacbon, và trong 48 giờ trong môi trường còn lại. Các kết quả được thể hiện trong Fig. 2. Các số 325, 425 và 513 gắn với tên chủng được thể hiện trong hình vẽ là các số dòng.

Khi hệ hỗn hợp của glucoza và xyloza được sử dụng dùng làm nguồn cacbon, trong khi chủng đối chứng (MG1655 Δ sucA/pMW119) tích lũy 15 đến 16 g/l axit L-glutamic, chủng biểu hiện operon ccrNXA (MG1655 Δ sucA/pMW119Ptac_ccrNXA) cho thấy sự tích lũy axit L-glutamic khoảng 12 g/l, và do đó có xu hướng thể hiện sự

tích lũy và hiệu suất của axit L-glutamic giảm. Ngoài ra, khi xyloza được sử dụng dùng làm nguồn cacbon duy nhất, kết quả tương tự là thu được. Đối với các sản phẩm phụ, các axit hữu cơ và axit xylonic được phân tích. Kết quả là, đã biết được rằng axit axetic và axit xylonic chủ yếu được tích lũy.

(3) Phân tích điểm giới hạn tốc độ của quá trình NXA trong *E. coli*

Vì axit xylonic, mà là chất trung gian của quá trình NXA, được xác định trong dịch nổi nuôi cấy vi khuẩn *E. coli* sản xuất axit L-glutamic biểu hiện operon *ccrNXA* như được mô tả trên đây, được xem như rằng có khả năng cao rằng một phần xyloza được kết hợp được đồng hóa thông qua quá trình NXA. Hơn thế nữa, các vấn đề được dự định là như sau.

i) trong khi xyloza được kết hợp vào các tế bào có thể được đồng hóa bởi cả hệ đồng hóa xyloza đặc trưng cho *E. coli* và quá trình NXA, lượng nhất định của xyloza có thể được sử dụng bởi hệ đặc trưng cho *E. coli* này do sự khác nhau về hoạt tính hoặc đặc trưng về mặt cơ chất của enzym thứ nhất của hệ đặc trưng cho *E. coli*, xyloza isomeraza (XylA), và enzym thứ nhất của quá trình NXA, xyloza dehydrogenaza (XDH), và do đó tốc độ dòng của dòng trao đổi chất đi qua quá trình NXA có thể trở nên nhỏ hơn.

ii) có thể có điểm giới hạn tốc độ trong quá trình NXA, hoặc quá trình phân nhánh đã biết, và do đó α KG có thể không được tạo ra.

Được xem như là vấn đề của i) có thể được làm giảm bằng cách làm tăng lượng của các enzym của quá trình NXA bằng cách thay đổi vector biểu hiện operon NXA từ vector loại số bản sao thấp (pMW119) cho vector loại số bản sao trung bình (pTWV228), và nhờ đó làm tăng lượng hấp thu của cơ chất vào quá trình *ccrNXA*.

Cũng được xem như là vấn đề của ii) có thể được khắc phục bằng cách cải thiện chúng bằng cách nhân giống dựa trên phân tích điểm giới hạn tốc độ và các kết quả của chúng.

Dựa trên việc xem xét nêu trên, các quá trình sau đây đã được tiến hành:

a) cấu trúc chủng từ chủng thiếu hụt quá trình đồng hóa xyloza đặc hiệu ở *E. coli* (Δ sucA Δ xylA) làm vật chủ, trong đó operon *ccrNXA* được biểu hiện và do đó dòng cacbon được ép đi qua quá trình NXA, và việc đánh giá nó bằng cách nuôi cấy,

b) cấu trúc vector biểu hiện operon *ccrNXA* sử dụng vector số bản sao trung bình, cấu trúc chủng sử dụng vector biểu hiện như vậy, và việc đánh giá nó bằng cách nuôi cấy, và

c) phân tích điểm giới hạn tốc độ.

Bằng cách làm khuyết gen đồng hóa xyloza đặc hiệu *E. coli xylA* từ MG1655 Δ sucA theo phương pháp λ -Red sử dụng các đoạn mồi *xylA*-H1P1-5742-5-1 (SEQ ID NO: 21) và *xylA*-H2P2-5742-5-2 (SEQ ID NO: 22), chủng MG1655 Δ sucA Δ *xylA* là thu được. pMW119Ptac_*ccrNXA* được đưa vào chủng này để thu được chủng biểu hiện operon *ccrNXA* thiếu hụt *xylA*.

Các kết quả của việc nuôi cấy sản xuất axit L-glutamic được tiến hành bằng cách nuôi cấy chủng biểu hiện operon *ccrNXA* thiếu hụt *xylA* theo cách tương tự như được mô tả trong phần (3) nêu trên được thể hiện trong Fig. 3. Trong Fig. 3, "*ccrNXA*" là pMW119Ptac_*ccrNXA*, và các số phía sau là số dòng.

Trong khi chủng đối chứng vector của chủng Δ sucA Δ *xylA* không thể đồng hóa xyloza, và có thể tạo ra tế bào và tạo ra axit L-glutamic chỉ từ glucoza, chủng biểu hiện operon *ccrNXA* thể hiện sự tiêu thụ xyloza và việc sản xuất axit L-glutamic, mà được xem là thu được từ xyloza. Tuy nhiên, đã khám phá ra rằng lượng tích lũy axit L-glutamic của nó là nhỏ hơn lượng thu được bằng chủng mẫu (chủng Δ sucA), và nó tích lũy axit xylonic, mà là chất trung gian trao đổi chất của quá trình *ccrNXA*. Từ các kết quả này, đã cho thấy rằng dòng trao đổi chất của toàn bộ quá trình *ccrNXA* có thể không đầy đủ. Hơn thế nữa, do việc sản xuất sản phẩm phụ α KG là không được thấy, nên nó được xem là cung cấp NADPH cần thiết cho sự biểu hiện của hoạt tính của GDH không có vấn đề bất kỳ ở gia đoạn này.

Sau đó, vector biểu hiện operon *ccrNXA* được cấu trúc sử dụng vector số bản sao trung bình. Vùng tac chứa đoạn khởi đầu operon *ccrNXA* được khuếch đại bằng cách sử dụng pMW119 Ptac_*ccrNXA* làm mẫu, cũng như PtwvPtacf (SEQ ID NO: 1) và 0819r (SEQ ID NO: 4) dùng làm các đoạn mồi. pTWV228 được phân cắt bởi *Sma*I, và được sử dụng cùng với đoạn PCR của vùng tac chứa đoạn khởi đầu operon *ccrNXA* để tiến hành phản ứng dung hợp, chủng *E. coli* JM109 được biến nạp với sản phẩm phản ứng, và plasmit đích pTWVPtac_*ccrNXA* thu được từ thể biến nạp đó. Plasmit này được đưa vào chủng MG1655 Δ sucA, và chủng thu được được nuôi cấy theo cách

tương tự như được mô tả trong phần (3) nêu trên. Các kết quả được thể hiện trong Fig. 4. Trong hình này, " Δ sucA" là chủng MG1655 Δ sucA, và "/pTWV" và "/v" có nghĩa là chủng chứa pTWV228. Hơn thế nữa, từ pTWV110 đến pTWV119 chỉ ra số dòng của pTWV228Ptac_ccrNXA.

Khi chỉ glucoza được sử dụng dùng làm nguồn cacbon, chủng biểu hiện operon ccrNXA số bản sao trung bình tích lũy axit L-glutamic với lượng về cơ bản tương đương với lượng thu được bằng chủng đối chứng. Tuy nhiên, khi hệ nuôi cấy hỗn hợp của glucoza và xyloza được sử dụng, sự tích lũy axit L-glutamic có xu hướng suy giảm. Hơn thế nữa, các kết quả của các chủng anh em cũng dao động. Một trong các lý do chấp nhận được là sự suy giảm của vector biểu hiện số bản sao trung bình. Hơn thế nữa, giống như các chủng được mô tả trên đây, sự tích lũy axit xylonic là thu được.

Để khẳng định xem liệu các hoạt tính của các enzym của quá trình ccrNXA được làm tăng bởi sự gia tăng về số bản sao của operon NXA, hoạt tính của XDH, mà là enzym thứ nhất của quá trình NXA, là được đo. Các kết quả được thể hiện trong Bảng 5. "7513" và "1110" trong các tên chủng được đề cập trong bảng 5 là các số dòng.

Bảng 5: Các kết quả đo hoạt tính XDH (xyloza dehydrogenaza)

Chủng	Hoạt tính đặc hiệu (μ mol/min/mg-protein)	Hoạt tính tương đối
MG1655 Δ sucA/pMW119	ND	
MG1655 Δ sucA/pMWccrNXA7513	21,5	1,0
MG1655 Δ sucA/pMW228	ND	
MG1655 Δ sucA/pMWccrNXA1110	140,4	6,5

ND: không xác định được

Lưu ý: hoạt tính tương đối được chỉ ra dưới dạng hoạt tính tương đối dựa trên hoạt tính đặc hiệu của ccrNXA7513 được lấy là 1.

Chủng được gắn vector biểu hiện bản sao trung bình thể hiện hoạt tính XDH cao hơn khoảng 7 lần so với chủng được gắn vector biểu hiện số bản sao thấp. Đã không khẳng định được làm sao xyloza được phân bố thực sự vào trong tế bào ở điểm

phân nhánh của xyloza isomeraza (XylA) đặc trưng đối với *E. coli* và XDH, và do đó cũng vẫn được xem như là hoạt tính của XDH, mà là enzym thứ nhất của quá trình NXA, có thể là không đầy đủ. Tuy nhiên, do sự tăng hoạt tính XDH không tạo ra hiệu quả cải thiện sự tích lũy axit L-glutamic, và dựa trên sự tích lũy axit xylonic, và v.v., được xem như là hoặc là một hoặc nhiều enzym của quá trình NXA có thể làm giới hạn tốc độ.

Do đó, điểm giới hạn tốc độ của quá trình NXA được phân tích. Do axit xylonic được tích lũy dưới dạng chất trung gian trao đổi chất, được xem là ít nhất một điểm giới hạn tốc độ có thể tồn tại trong quá trình từ axit xylonic đến α KG mà không phải là quá trình trao đổi chất từ xyloza đến axit xylonic. Hơn thế nữa, trong cấu trúc của operon *ccrNXA*, trong khi các enzym của quá trình từ xyloza đến axit xylonic được mã hóa bởi các gen ở các vị trí thứ ba và thứ tư, các enzym của quá trình từ axit xylonic đến α KG được mã hóa bởi các gen định vị ở các vị trí đầu tiên, thứ hai và thứ năm. Hoạt tính của XDH, mà gen của nó ở vị trí thứ ba trong operon, đã được xác định *in vitro*, và do đó được xem như là nêu hoạt tính của enzym được mã hóa bởi gen ở vị trí thứ năm trong operon (*XylD*) cũng có thể được xác định hơn nữa, có thể dùng làm bằng chứng chi tiết về quá trình sao mã và dịch mã của toàn bộ operon NXA. Do đó, ước tính rằng một trong ba phản ứng từ axit xylonic đến α KG tạo ra điểm giới hạn tốc độ, và các thử nghiệm dưới đây được tiến hành.

Các plasmit *pSTVPtac_xylD_Ttrp*, *pSTVPtac_xylX_Ttrp*, và *pSTVPtac_ccrxylA_Ttrp* mà lần lượt biểu hiện các gen *xylD*, *xylX*, và *ccrxylA*, được điều chế như sau.

pSTV28-Ptac-xylX-Ttrp được tạo ra bằng cách tạo ra đoạn *xylX* bởi PCR sử dụng *pUC18-xylX*, mà là plasmit tạo ra được bằng cách tách dòng *xylX* từ vị trí mà *SfiI* đã được tách loại thành *pUC18*, dùng làm mẫu, cũng như *xylX*-IFS-5742-10-1 (SEQ ID NO: 38) và *xylX*-IFA-5742-10-2 (SEQ ID NO: 39) dùng làm các đoạn nối, và tách dòng plasmit thu được thành *pSTV28-Ptac-Ttrp* được phân cắt bởi *SmaI* bằng phương pháp tách dòng nội dung hợp.

pSTV28-Ptac-ccrxylA-Ttrp được tạo ra bằng cách điều chế đoạn bằng phương pháp PCR sử dụng *pUC18-ccrxylA*, mà là plasmit điều chế được bằng cách tách dòng *ccrxylA* mà từ đó vị trí *SfiI* đã được tách loại thành *pUC18*, dùng làm mẫu, cũng như

xylA_IFS_5742-10-3 (SEQ ID NO: 40) và *xylA_IFA_5742-10-4* (SEQ ID NO: 41) dùng làm các đoạn mồi, và tách dòng plasmid thu được thành pSTV28-Ptac-Ttrp được phân cắt bởi *SmaI* bằng phương pháp nội dung hợp.

pSTV28-Ptac-*xylD*-Ttrp được tạo ra bằng cách điều chế đoạn *xylD* bằng phương pháp PCR sử dụng pUC18-*xylD*, mà là plasmid điều chế được bằng cách tách dòng *xylD* mà từ đó vị trí *SfiI* đã được tách loại thành pUC18, dùng làm mẫu, cũng như *xylD_IFS_5742-10-5* (SEQ ID NO: 42) và *xylD_IFA_5742-10-6* (SEQ ID NO: 43) dùng làm các đoạn mồi, và tách dòng plasmid thu được thành pSTV28-Ptac-Ttrp được phân cắt bởi *SmaI* bằng phương pháp nội dung hợp.

Các plasmid pUC18-*xylX*, pUC18-*ccrxylA*, và pUC18-*xylD* được đề cập trên đây lần lượt được điều chế như được mô tả trên đây.

Bằng phương pháp PCR sử dụng hệ gen ADN của chủng *C. crescentus* CB-15 dùng làm mẫu, CC0823-01F_4691-87-1 (SEQ ID NO: 44) và CC0823-01R_4691-87-2 (SEQ ID NO: 45), cũng như CC0823-02F_4691-87-3 (SEQ ID NO: 46) và CC0823-02R_4691-87-4 (SEQ ID NO: 47) làm các đoạn mồi, các đoạn 900 bp và 280 bp lần lượt được khuếch đại. Sau đó, pUC18 được phân cắt bởi *SmaI*, và hai đoạn được khuếch đại được đề cập trên đây được lắp ghép bằng phương pháp lắp ghép in vitro (Nature Methods, 6(5), 343-345, 2009) để thu được pUC18-*xylX* mà trong đó vị trí *SfiI* được loại bỏ, và gen *xylX* được chèn vào.

Bằng phương pháp PCR sử dụng hệ gen ADN của chủng *C. crescentus* CB-15 làm mẫu, CC0822-01F_4691-87-5 (SEQ ID NO: 48) và CC0822-01R_5659-8-7 (SEQ ID NO: 49), CC0822-02F_5659-8-8 (SEQ ID NO: 50) và CC0822-02R_5659-8-9 (SEQ ID NO: 51), CC0822-03F_5659-8-10 (SEQ ID NO: 52) và CC0822-03R_5659-8-11 (SEQ ID NO: 53), CC0822-04F_5659-8-12 (SEQ ID NO: 54) và CC0822-04R_5659-8-13 (SEQ ID NO: 55), cũng như CC0822-05F_5659-8-14 (SEQ ID NO: 56) và CC0822-05R_4691-87-14 (SEQ ID NO: 57) dùng làm các đoạn mồi, năm đoạn, 11v02 (175 bp), 12v02 (325 bp), 13v02 (260 bp), 14v02 (193 bp), và 15v02 (544 bp), lần lượt được khuếch đại. Sau đó, hai đoạn 11v02 và 12v02 được liên kết bằng PCR trao đổi chéo sử dụng hai đoạn này làm mẫu, cũng như CC0822-01F_4691-87-5 và CC0822-02R_5659-8-9 dùng làm các đoạn mồi. Tương tự, hai đoạn 13v02 và 14v02 được liên kết bằng PCR trao đổi chéo sử dụng hai đoạn này làm mẫu, cũng như

CC0822-03F_5659-8-12 và CC0822-04R_5659-8-13 dùng làm các đoạn mồi. Hai đoạn này và đoạn 15v02 được đề cập trên đây được liên kết bằng phương pháp lắp ghép in vitro (Nature Methods, 6(5), 343-345, 2009). Đoạn được liên kết thu được được khuếch đại bằng cách PCR sử dụng nó làm mẫu, cũng như CC0822-01F_4691-87-5 và CC0822-05R_4691-87-14 dùng làm các đoạn mồi. Sau đó, pUC18 được phân cắt bởi *Sma*I, và đoạn được liên kết được đề cập trên đây được lắp ghép bằng phương pháp lắp ghép in vitro (Nature Methods, 6(5), 343-345, 2009) để thu được pUC18-*ccrxylA* mà trong đó vị trí *Sfi*I được loại bỏ, và gen *ccrxylA* được chèn.

Bằng phương pháp PCR sử dụng hệ gen ADN của chủng *C. crescentus* CB-15 làm mẫu, CC0819-01F_4691-88-7 (SEQ ID NO: 17) và CC0819-01R_5659-9-1 (SEQ ID NO: 18), cũng như CC0819-02F_5659-9-2 (SEQ ID NO: 19) và CC0819-02R_4691-88-10 (SEQ ID NO: 20) dùng làm các đoạn mồi, các đoạn 1130 bp và 653 bp lần lượt được khuếch đại. Sau đó, pUC18 được phân cắt bởi *Sma*I, và hai đoạn được khuếch đại được đề cập trên đây được lắp ghép bằng phương pháp lắp ghép in vitro (Nature Methods, 6(5), 343-345, 2009) để thu được pUC18-*xylD* mà trong đó vị trí *Sfi*I được loại bỏ, và gen *xylD* được chèn.

Phân chiết enzym thô được điều chế từ chủng biểu hiện operon *ccrNXA* (MG1655 Δ sucA/pTWV228Ptac-*ccrNXA*), và các chủng chứa mỗi plasmid biểu hiện một trong các gen *xylD*, *xylX*, và *ccrxylA* được đề cập trên đây, (MG1655 Δ sucA/pSTVPtac-*xylD*_Ttrp, MG1655 Δ sucA/pSTVPtac-*xylX*_Ttrp, và MG1655 Δ sucA/pSTVPtac-*ccrxylA*_Ttrp), và sau đó mỗi phân chiết enzym thô của chủng chỉ chứa vector, hoặc các chủng biểu hiện một trong các gen *xylD*, *xylX*, và *ccrxylA* được bổ sung vào phân chiết enzym thô của chủng biểu hiện operon *ccrNXA* đó, và hoạt tính để sản xuất α KG từ axit xylonic của mỗi hỗn hợp là được đo. Các kết quả được thể hiện trong bảng 6. Trong Bảng 6, "1110" gắn với tên chủng là số dòng.

Bảng 6: Đo các kết quả hoạt tính sản xuất α KG từ axit xylonic

CFE(①+②)	①	②			
	Δ <i>lacA</i> /pTWVP <i>lac ccrNXA 1110</i>				
	JM109/pSTV <i>P_{lac} T_{trp}</i>				
	Không	<i>xylD</i>	<i>xylX</i>	<i>ccrxylA</i>	
Hoạt tính đặc hiệu ($\mu\text{mol/min/mg-protein}$)	1,35	3,48	1,34	1,91	
	1,0	2,6	1,0	1,4	

Lưu ý: Hoạt tính tương đối được chỉ ra dưới dạng hoạt tính tương đối dựa trên hoạt tính đặc hiệu của hệ hỗn hợp ccrNXA1110 và pSTV28-Ptac-Ttrp lấy là 1.

Với hệ mà phần chiết enzyme thô của chủng chỉ biểu hiện gen *xylD* được bổ sung, sự gia tăng hoạt tính sản xuất α KG được quan sát. Từ kết quả này, đã cho thấy rằng xylonat dehydrataza (XylD) được mã hóa bởi gen *xylD* tạo ra điểm giới hạn tốc độ của quá trình NXA được cấu trúc trong *E. coli* bằng cách biểu hiện dị thể.

Do dòng trao đổi chất của toàn bộ quá trình này có thể được cải thiện bằng cách làm tăng hơn nữa sự biểu hiện của gen *xylD* trong chủng biểu hiện operon ccrNXA như được đề xuất bằng cách đo hoạt tính enzym, chủng biểu hiện tăng gen *xylD* được cấu trúc bằng cách đưa vector biểu hiện gen *xylD* vào chủng biểu hiện operon ccrNXA, và được đánh giá bằng cách nuôi cấy để sản xuất axit L-glutamic sử dụng glucoza và xyloza dùng làm nguồn cacbon. Dùng làm chủng biểu hiện operon ccrNXA, MG1655 Δ sucA/pMW119Ptac_ccrNXA và MG1655 Δ sucA/pTWV228Ptac_ccrNXA được sử dụng.

Việc nuôi cấy được tiến hành theo cách tương tự như được mô tả trong phần (3) nêu trên.

Các kết quả được thể hiện trong bảng 7. Chủng làm tăng sự biểu hiện gen *xylD* thể hiện sự tích lũy và hiệu suất axit L-glutamic được làm tăng đáng kể. Sự tích lũy axit L-glutamic là 23 đến 25 g/l trái ngược với 15 đến 16 g/l chủng đối chứng, và hiệu suất dựa trên sacarit tiêu thụ đạt từ 57 đến 60% trái ngược với từ 37 đến 40% của chủng đối chứng. axit xylonic, mà là chất trung gian trao đổi chất, là không thấy được với chủng được làm tăng sự biểu hiện gen *xylD*. Mặt khác, hiệu quả làm tăng sự biểu hiện gen *xylD* chỉ được thấy ở chủng biểu hiện mang vector biểu hiện operon ccrNXA loại số bản sao trung bình, và hiệu quả này không thấy được ở chủng biểu hiện mang vector loại số bản sao thấp. Từ các kết quả này, được xem như là việc loại bỏ điểm giới hạn tốc độ của quy trình này bằng cách làm tăng hoạt tính của toàn bộ quá trình NXA nhờ việc làm tăng số bản sao của các vector và làm tăng hơn nữa sự biểu hiện của gen *xylD* tạo ra sự cải thiện lượng sản xuất axit L-glutamic. Ngoài ra, khi hoạt tính sản xuất α KG từ axit xylonic của chủng được làm tăng sự biểu hiện của gen *xylD* được đo, nó tăng khoảng 10 lần so với giá trị trước khi làm tăng (Bảng 8). Các số "1110", "17" và "19" gắn với các tên chủng được đề cập trong các bảng 7 và 8 là các số dòng.

Bảng 7: Kết quả đánh giá chủng biểu hiện operon *ccrNXA* + *xylD* bằng quá trình nuôi cấy để sản xuất L-Glu

Chủng	L-Glu(g/l)	Hiệu suất (%)	O.D,600
MG1655 Δ sucA/pTWV228	15,7	37,6	23,9
MG1655 Δ sucA/pTWVccrNXA1110	16,3	39,2	16,6
MG1655 Δ sucA/pTWVccrNXA/pSTVPtacTtrp	16,3	39,2	22,9
MG1655 Δ sucA/pTWVccrNXA/ <i>xylD</i> 17	25,4	61,1	14,3
MG1655 Δ sucA/pTWVccrNXA/ <i>xylD</i> 19	24,2	58,1	16,0

Bảng 8: Các kết quả đo hoạt tính sản xuất α KG từ axit xylonic

Chủng	Hoạt tính đặc hiệu (μ mol/min/mg-protein)	Hoạt tính tương đối
MG1655 Δ sucA/pTWV228/pSTVPtacTtrp	ND	
MG1655 Δ sucA/pTWVccrNXA1110	1,9	1,0
MG1655 Δ sucA/pTWVccrNXA/pSTV <i>xylD</i> 17	20,2	10,8
MG1655 Δ sucA/pTWVccrNXA/pSTV <i>xylD</i> 19	17,1	9,2

Lưu ý: Các kết quả đối với MG1655 Δ sucA/pTWVccrNXA1110 là các giá trị thu được trong các thử nghiệm sử dụng các mẻ khác nhau.

ND: không xác định được

Lưu ý: Hoạt tính tương đối được chỉ ra dưới dạng hoạt tính tương đối dựa trên hoạt tính đặc hiệu của *ccrNXA*1110 lấy làm 1.

Ví dụ 3: Đưa quá trình NXA vào *Corynebacterium glutamicum*

(1) Cấu trúc plasmit pVK9Pefu_*ccrNXA* để đưa vào đó quá trình NXA

Plasmit có trình tự chứa trình tự đoạn khởi đầu của gen Tu yếu tố kéo dài (EF-Tu), tuf (WO2008/114721, SEQ ID NO: 33, giờ đây được đề cập đến dưới dạng "Pefu") và *xylD* được liên kết ở phía sau từ trình tự đoạn khởi đầu được cấu trúc bằng

sử dụng kit tách dòng Clontech In-Fusion Cloning HD (Clontech). Đầu tiên, PCR được tiến hành bằng cách sử dụng ADN nhiễm sắc thể của chủng *C. glutamicum* ATCC 13869 làm mẫu, cũng như các đoạn mồi Peftu(Pst) (SEQ ID NO: 58) và Peftu_Rv (SEQ ID NO: 59) để thu được đoạn chứa trình tự Peftu. PCR này được tiến hành bằng cách sử dụng PrimeSTAR HS Polymeraza theo phương pháp hướng dẫn kèm theo enzym này.

Hơn thế nữa, PCR được tiến hành bằng cách sử dụng pTWV228Ptac_ccrNXA làm mẫu, cũng như các đoạn mồi Peftu_xylXABCD_fw (SEQ ID NO: 60) và Peftu_xylXABCD_rv (SEQ ID NO: 61) để thu được đoạn chứa trình tự *xylXABCD* của *C. crescentus*. PCR này được tiến hành bằng cách sử dụng PrimeSTAR GXL Polymeraza theo theo phương pháp hướng dẫn kèm theo enzym này.

Sau đó, đoạn Peftu và đoạn chứa *xylXABCD* thu được trên đây được trộn với pVK9 được xử lý bằng *Pst*I và *Bam*HI, và được sử dụng để tiến hành phản ứng dung hợp theo phương pháp hướng dẫn sử dụng kit tách dòng Clontech In-fusion HD. pVK9 là vector con thoi thu được bởi pHSG299 đầu bằng (Takara Bio) ở vị trí *Ava*II, và chèn vùng có thể tự sao chép trong vi khuẩn coryneform được chứa trong pHK4 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 05-007491), mà được cắt bằng *Bam*HI và *Kpn*I, và được làm bằng đầu (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2007-97573, Công bố Đơn yêu cầu cấp bằng sáng chế Mỹ số 2005/0196846). *E. coli* JM109 được biến nạp với hỗn hợp phản ứng dung hợp. Các thể biến nạp được lựa chọn trong môi trường agar chứa môi trường LB được bổ sung kanamycin ở nồng độ cuối bằng 50 mg/l. Plasmit đích pVK9Peftu_ccrNXA thu được từ thể biến nạp thu được.

(2) Sản xuất axit L-glutamic bằng *Corynebacterium glutamicum* được đưa vào đó quá trình NXA

Chủng *C. glutamicum* ATCC 13869 được biến nạp với pVK9Peftu_ccrNXA được đề cập trên đây bằng phương pháp xung điện (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2-207791). Chủng được gắn pVK9Peftu_ccrNXA được lựa chọn trong môi trường agar chứa môi trường CM-Dex được bổ sung kanamycin ở nồng độ cuối bằng 25 mg/l. Hơn thế nữa, chủng *C. glutamicum* ATCC13869 được biến nạp với pVK9 cũng được lựa chọn theo cách tương tự dùng làm chủng đối chứng.

Khả năng đồng hóa xyloza và khả năng sản xuất axit L-glutamic của các thể biến nạp thu được được đánh giá bằng cách tiến hành nuôi cấy sử dụng bình cầu Sakaguchi. Các tế bào của mỗi chủng biến nạp được nuôi cấy ở 31,5°C trong 24 giờ trong môi trường CM-Dex agar được bổ sung kanamycin ở nồng độ cuối bằng 25 mg/l được cạo lấy lượng tương ứng với 1/6 đĩa, và được ủ trong 20 ml môi trường Glc được chứa trong bình cầu Sakaguchi, 1 g canxi cacbonat được thanh trùng trước bằng không khí nóng được bổ sung, và lắc môi trường nuôi cấy ở 31,5°C và 120 vòng/phút trong 24 giờ. Môi trường nuôi cấy thu được với thể tích 1 ml được ủ trong 20 ml môi trường Xyl (biotin được hạn chế) được chứa trong bình cầu Sakaguchi, 1 g canxi cacbonat được thanh trùng trước bằng không khí nóng được bổ sung, và việc lắc môi trường nuôi cấy được tiến hành ở 31,5°C và 120 vòng/phút trong 73 giờ. Các kết quả được thể hiện trong bảng 9.

Chủng *C. glutamicum* ATCC 13869 được nhập pVK9 hầu như không sinh trưởng trong môi trường Xyl, và cũng không thể hiện khả năng sản xuất axit L-glutamic. Mặt khác, chủng *C. glutamicum* ATCC 13869 được nhập pVK9Peftu_ccrNXA sinh trưởng trong môi trường Xyl, và tích lũy axit L-glutamic. Từ các kết quả này, đã cho thấy rằng việc đưa quá trình NXA vào vi khuẩn coryneform được cải thiện khả năng đồng hóa xyloza, và chủng này sản xuất axit L-glutamic từ xyloza.

Bảng 9

Chủng	OD620 (x101)	Xyl tiêu thụ (g/l)	Glu (g/l)	Hiệu suất Axit xylonic (%)	Axit xylonic (g/l)
ATCC13869/pVK9	0,048 ± 0,013	0	0	0	0
ATCC13869/ pVK9Peftu_ccrNXA	0,404 ± 0,010	57,1 ± 0,22	2,5 ± 0,3	4,3 ± 0,5	44,5 ± 1,5

(3) Cấu trúc chủng được làm tăng hơn nữa sự biểu hiện gen *xylD*

Do chủng *C. glutamicum* ATCC 13869 chứa pVK9Peftu_ccrNXA tích lũy axit xylonic, nên được cho rằng hoạt tính của sản phẩm gen *xylD* là không đầy đủ. Do đó, đã nỗ lực để cấu trúc plasmit cho sự biểu hiện của quá trình NXA trong đó gen *xylD* được làm tăng hơn nữa bằng cách đưa một hoặc nhiều bản sao của gen *xylD* vào

plasmid pVK9 *Peftu_ccrNXA*.

(4) Cấu trúc plasmid pVS7PmsrA_xylD cho sự biểu hiện của gen *xylD*

Plasmid có trình tự bao gồm trình tự đoạn khởi đầu của gen *msrA* (peptid methionin sulfoxit reductaza A) (giờ đây được đề cập đến dưới dạng "PmsrA") và gen *xylD* của *C. crescentus* được liên kết ở phía sau từ trình tự đoạn khởi đầu được cấu trúc bằng sử dụng kit Clontech In-Fusion Cloning HD.

Đầu tiên, PCR được tiến hành bằng cách sử dụng ADN nhiễm sắc thể của chủng *C. glutamicum* ATCC 13869 làm mẫu, cũng như các đoạn mồi PmsrA(Pst) (SEQ ID NO: 62) và PmsrAR (SEQ ID NO: 63) để thu được đoạn chứa trình tự PmsrA. Hơn thế nữa, PCR được tiến hành bằng cách sử dụng pTWV228Ptac_ccrNXA làm mẫu, cũng như các đoạn mồi PmsrA_xylD_fw (SEQ ID NO: 64) và *Peftu_xylXABCD_rv* để thu được đoạn chứa trình tự gen *xylD* của *C. crescentus*. Các PCR được tiến hành bằng cách sử dụng PrimeSTAR HS Polymeraza theo phương pháp hướng dẫn kèm theo enzym này.

Sau đó, các đoạn chứa đoạn PmsrA và gen *xylD* thu được trên đây được trộn với vectơ con thoi pVS7 (phương pháp cấu trúc được thể hiện dưới đây) được xử lý bằng *Pst*I và *Bam*HI, và được sử dụng để tiến hành phản ứng dung hợp theo phương pháp hướng dẫn sử dụng kit tách dòng Clontech In-fusion HD, và sau đó *E. coli* JM109 được biến nạp với hỗn hợp phản ứng này. Các thể biến nạp này được lựa chọn trong môi trường agar chứa môi trường LB được bổ sung spectinomycin ở nồng độ cuối bằng 25 mg/l. Plasmid đích pVS7PmsrA_xylD thu được từ thể biến nạp thu được.

pVS7 là plasmid thu được bằng cách thay thế gen kháng cloramphenicol của pVC7 (Đơn yêu cầu cấp bằng sáng chế Nhật Bản chưa thẩm định số 2000-201692, patent châu Âu số 1004671) bằng gen kháng spectinomycin. Gen kháng spectinomycin này có thể thu được bằng cách điều chế plasmid pDG1726 từ chủng *Escherichia coli* ECE101E được bán bởi Bacillus Genetic Stock Center (BGSC), và thu lấy gen kháng này từ plasmid dùng làm caset. Bằng PCR sử dụng pDG1726 làm mẫu, cũng như các đoạn mồi SpcR-F (SEQ ID NO: 65) và SpcR-R (SEQ ID NO: 66), gen kháng spectinomycin được khuếch đại. Đoạn gen thu được được trộn với pVC7 được xử lý bằng *Sma*I, phản ứng liên kết được tiến hành theo phương pháp của Ligation Mix <Mighty Mix> của Takara Bio, và *E. coli* JM109 được biến nạp với hỗn

hợp phản ứng này. Các thể biến nạp được lựa chọn trong môi trường agar chứa môi trường LB được bổ sung spectinomycin ở nồng độ cuối bằng 25 mg/l. pVC7-spc chứa pVC7 được chèn bằng gen kháng spectinomycin thu được từ thể biến nạp thu được. Hơn thế nữa, PCR được tiến hành bằng cách sử dụng pVC7-spc làm mẫu, cũng như các đoạn môi spc(GTG khởi đầu)-F (SEQ ID NO: 67) và spc(kết thúc)-R (SEQ ID NO: 68) để làm khuếch đại gen kháng spectinomycin.

Theo cách riêng biệt, PCR được tiến hành bằng cách sử dụng pVC7 làm mẫu, cũng như các đoạn môi Spc-pVC7-Cm-F (SEQ ID NO: 69) và Spc-pVC7-Cm-R (SEQ ID NO: 70) để thu được đoạn ADN bao gồm pVC7 trong đó gen kháng chloramphenicol được loại bỏ. Đoạn ADN này và đoạn ADN của gen kháng spectinomycin thu được trên đây từ pVC7-spc được trộn, và được sử dụng để tiến hành phản ứng dung hợp theo phương pháp hướng dẫn kèm theo kit tách dòng Clontech In-fusion HD, và *E. coli* JM109 được biến nạp với hỗn hợp phản ứng thu được. Các thể biến nạp được trải qua quá trình chọn lọc trong môi trường agar chứa môi trường LB được bổ sung spectinomycin ở nồng độ cuối bằng 25 mg/l. pVS7 thu được từ thể biến nạp thu được.

(5) Cấu trúc plasmit pVK9Peftu_ccrNXA^{+D} được làm tăng hơn nữa gen *xylD*

Đoạn ADN chứa gen kháng spectinomycin (Spc) và PmsrA_*xylD* được khuếch đại bằng cách sử dụng pVS7PmsrA_*xylD* làm mẫu, cũng như các đoạn môi ME_Spc_fw (SEQ ID NO: 71) và ME_Peftu_*xyl*XABCD_rv (SEQ ID NO: 72). Đoạn ADN thu được được xen vào pVK9Peftu_ccrNXA *in vitro* theo phương pháp hướng dẫn kèm theo kit Ez-Tn5TM Custom Transposome Construction Kit (Epicentre), và đoạn thu được được sử dụng để biến nạp *E. coli* DH5 α . Các thể biến nạp được trải qua quá trình chọn lọc trong môi trường agar chứa môi trường LB agar được bổ sung kanamycin và spectinomycin ở nồng độ cuối bằng 50 mg/l và 25 mg/l. Các plasmit được tách ra từ các thể biến nạp thu được, và plasmit mà đã khẳng định được rằng vị trí chèn của trình tự Spc-PmsrA_*xylD* không nằm trên trình tự Peftu_ccrNXA bằng phép phân tích trình tự nucleotit xung quanh trình tự Spc-PmsrA_*xylD* được chỉ định pVK9Peftu_ccrNXA^{+D}.

(6) Sản xuất axit L-glutamic bằng chủng được nhập quá trình NXA có gen *xylD* được làm tăng hơn nữa

pVK9Peftu_ccrNXA⁺^D được đưa vào chủng *C. glutamicum* ATCC 13869 bằng phương pháp xung điện, và các tế bào được sử dụng cho môi trường CMDex agar chứa 25 mg/l kanamycin. Khả năng sản xuất axit L-glutamic của chủng được sinh trưởng ở 31,5°C được đánh giá theo phương pháp tương tự như phương pháp của phần (2) nêu trên. Các kết quả được thể hiện trong bảng 10.

Chủng chứa pVK9Peftu_ccrNXA⁺^D tích lũy axit D-xylonic với lượng nhỏ, nhưng tích lũy axit L-glutamic với lượng lớn hơn so với chủng chứa pVK9Peftu_ccrNXA. Bảng kết quả này, đã chứng minh được rằng axit L-glutamic có thể được tạo ra hiệu quả hơn từ D-xyloza thông qua quá trình NXA bằng cách làm tăng hơn nữa gen *xylD*.

Bảng 10: Sản xuất axit L-glutamic của chủng được nhập quá trình NXA được làm tăng hơn nữa gen *xylD*.

Chủng	OD620 (x101)	Xyl tiêu thụ (g/l)	L-Glu (g/l)	Hiệu suất (%)	Axit xylonic (g/l)
ATCC13869/ pVK9Peftu_ccrNXA	0,297 ± 0,004	58,3 ± 0,11	4,2 ± 0,1	7,1 ± 0,1	50,2 ± 0,1
ATCC13869/ pVK9Peftu_ccrNXA ⁺ ^D	0,351 ± 0,002	71,9 ± 0,66	27,5 ± 0,1	38,2 ± 0,4	22,7 ± 0,4

Ví dụ 4

(1) Sự thay thế các gen của quá trình NXA

Trong các ví dụ được đề cập trên đây, bằng cách sử dụng các gen *xylX*, *ccrxylA*, *ccrxylB*, *xylC* và *xylD* của vi khuẩn đã biết *C. crescentus*, mà quá trình NXA đã được thông báo, đã chứng minh được rằng axit glutamic có thể được tạo ra từ xyloza thông qua quá trình NXA. Trong ví dụ này, các gen tương tự của *xylD*, *xylX*, và *ccrxylA* thu được từ các loài sinh vật khác *C. crescentus*, và được đánh giá xem liệu các gen này có thể thay thế cho các gen của *C. crescentus*. Dùng làm nguồn gen, các loài sinh vật được mô tả trong bảng 11 là được chọn. Ký hiệu gen của các gen (Ngân hàng gen) là cũng đã biết. Các số xác định trình tự của các trình tự nucleotit của các gen và các trình tự axit amin được mã hóa bởi chúng được sử dụng trong danh mục trình tự được

thể hiện trong bảng 12. Trong bảng 12 này, "gốc" có nghĩa là trình tự nucleotit gen có trong tự nhiên, và "được tối ưu" có nghĩa là trình tự nucleotit mà codon của nó được tối ưu hóa theo việc sử dụng codon trong *E. coli*.

Trong phần mô tả sau đây, các enzym được mã hóa bởi các gen đồng đẳng *xylD*, *xylX*, và *xylA* lần lượt có thể được đề cập ở dạng XylD, XylX, và XylA.

Bảng 11

Sinh vật	Phân loại	Viết tắt	Ký hiệu gen
<i>Agrobacterium tumefaciens 5A</i>	α -proteobacteria	xylD(Atu)	EHJ96830
<i>Herbaspirillum seropedicae</i>	β -proteobacteria	xylD(Hse)	Hsero_4498
<i>Escherichia coli</i>	γ -proteobacteria	yjhG	ECK4286
<i>Escherichia coli</i>	γ -proteobacteria	yagF	ECK0270
<i>Actinoplanes missouriensis</i>	Actinobacteria	xylD(Amis)	AMIS_27920
<i>Aspergillus oryzae</i>	Fungi	xylD(Aor)	AOR_1_412134
<i>Agrobacterium tumefaciens 5A</i>	α -proteobacteria	xylX(Atu)	EHJ96825
<i>Cupriavidus necator</i>	β -proteobacteria	xylX(Cne)	CNE_2c03420
<i>Pseudomonas elodea</i>	γ -proteobacteria	xylX(Selo)	ZP_09955741
<i>Zobellia galactanivorans</i>	Bacteroidetes	xylX(Zga)	zobellia_2318
<i>Thermobacillus composti</i>	Firmicutes	xylX(Tco)	ZP_08919992.1
<i>Arthrobacter globiformis</i>	Actinobacteria	xylX(Art)	ARGLB_037_02150
<i>Azospirillum brasilense</i>	α -proteobacteria	xylA(Abr)	BAE94276.1
<i>Halomonas boliviensis</i>	γ -proteobacteria	xylA(Hbo)	ZP_09188044.1
<i>Bacillus subtilis</i>	Firmicutes	ycbD	BSU02470

Bảng 12

Viết tắt		SEQ ID NO (Trình tự nucleotit)	SEQ ID NO (trình tự axit amin)
xylD(Atu)	Gốc	73	74
	Được tối ưu hóa	75	
xylD(Hse)	Gốc	76	77
	Được tối ưu hóa	78	
yjhG		34	35
yagF		36	37
xylD(Amis)	Gốc	79	80
	Được tối ưu hóa	81	
xylD(Aor)	Gốc	82	83
	Được tối ưu hóa	84	
xylX(Atu)	Gốc	85	86
	Được tối ưu hóa	87	
xylX(Cne)	Gốc	88	89
	Được tối ưu hóa	90	
xylX(Selo)	Gốc	91	92
	Được tối ưu hóa	93	
xylX(Zga)	Gốc	94	95
	Được tối ưu hóa	96	
xylX(Tco)	Gốc	97	98
	Được tối ưu hóa	99	
xylX(Art)	Gốc	100	101
	Được tối ưu hóa	102	
xylA(Abr)	Gốc	103	104
	Được tối ưu hóa	105	
xylA(Hbo)	Gốc	106	107
	Được tối ưu hóa	108	
ycbD		109	110

(2) Cấu trúc các plasmit để xác định hoạt tính của XylD, XylX, và XylA, pTWVPtac_ccrNXA_ΔxylD_Km, pTWVPtac_ccrNXA_ΔxylX_Km, và pTWVPtac_ccrNXA_ΔccrxylA_Km

Việc cấu trúc được tiến hành bằng cách sử dụng kit tách dòng Clontech In-Fusion.

Đầu tiên, bằng PCR sử dụng pTWVPtac_ccrNXA_Km làm mẫu, cũng như Ptac_xylXABC_F (SEQ ID NO: 111) và Ptac_xylXABC_R (SEQ ID NO: 112) dùng làm các đoạn môi, đoạn ADN ngoại trừ *xylD* được khuếch đại. Sản phẩm PCR được sử dụng để tiến hành phản ứng dung hợp theo phương pháp của kit tách dòng Clontech In-fusion HD, chủng *E. coli* JM109 được biến nạp với sản phẩm phản ứng, và plasmit đích pTWVPtac_ccrNXA_ΔxylID_Km thu được từ thể biến nạp đó.

Theo cách tương tự như được mô tả trên đây, pTWVPtac_ccrNXA_ΔxylX_Km được cấu trúc bằng sử dụng Ptac_xylABCD_F (SEQ ID NO: 113) và Ptac_xylABCD_R (SEQ ID NO: 114) dùng làm các đoạn môi, và pTWVPtac_ccrNXA_ΔccrxylA_Km được cấu trúc bằng sử dụng Ptac_xylXBCD_F (SEQ ID NO: 115) và Ptac_xylXBCD_R (SEQ ID NO: 116) dùng làm các đoạn môi.

(3) Cấu trúc *Pantoea ananatis* để xác định các hoạt tính XylD, XylX, và XylA

Chủng *P. ananatis* NA1 được biến nạp với pTWVPtac_ccrNXA_ΔxylID_Km được mô tả trên đây bằng phương pháp mở lỗ bằng điện. Chủng được cấu trúc này được đề cập đến *P. ananatis* NA2 ΔxylID. Để nuôi cấy *P. ananatis* NA2 ΔxylID, môi trường trong đĩa chứa LBGM9 mà kanamycin và tetracyclin được bổ sung lần lượt ở các nồng độ cuối 40 mg/l và 12,5 mg/l, được sử dụng.

Theo phương pháp tương tự, chủng *P. ananatis* NA1 được biến nạp với pTWVPtac_ccrNXA_ΔxylX_Km hoặc pTWVPtac_ccrNXA_ΔccrxylA_Km để cấu trúc chủng *P. ananatis* NA2 ΔxylX và chủng *P. ananatis* NA2 ΔccrxylA.

(4) Cấu trúc các plasmit biểu hiện đồng đẳng *xylD*, *xylX*, và *xylA*

Các plasmit để biểu hiện của *yjhG* hoặc *yagF*, pSTV28-Ptac-yjhG-Ttrp và pSTV28-Ptac-yagF-Ttrp, được điều chế như sau.

pSTV28-Ptac-yjhG-Ttrp được điều chế bằng cách khuếch đại đoạn *yjhG* bằng PCR sử dụng hệ gen ADN của chủng *E. coli* MG1655 làm mẫu, cũng như *yjhG_F* (SEQ ID NO: 117) và *yjhG_R* (SEQ ID NO: 118) dùng làm các đoạn môi, và tách dòng đoạn được khuếch đại này thành pSTV28-Ptac-Ttrp được phân cắt bởi *SmaI* theo phương pháp tách dòng nội dung hợp.

pSTV28-Ptac-yagF-Ttrp được điều chế theo cách tương tự như được mô tả trên đây bằng cách sử dụng yagF_F (SEQ ID NO: 119) và yagF_R (SEQ ID NO: 120) dùng làm các đoạn nối.

Plasmid để biểu hiện của *xylD*(Hse), pSTV28-Ptac-*xylD*(Hse)-Ttrp, được điều chế như sau. Đoạn ADN có các trình tự của đoạn khởi đầu tac, *xylD*(Hse), và yếu tố kết thúc *trp* (Ptac-*xylD*(Hse)-Ttrp) được tổng hợp, và được liên kết với vector pUC57 (được mua từ Thermo Fischer Scientific) được phân cắt bởi *EcoRV* để thu được pUC57-Ptac-*xylD*(Hse)-Ttrp. Khi đoạn ADN được tổng hợp, các codon được tối ưu hóa sao cho đoạn này là thích hợp để biểu hiện trong *E. coli*. Những lượng tương đương của pSTV28 và pUC57-Ptac-*xylD*(Hse)-Ttrp, cả hai được phân cắt bởi *EcoRI* và *KpnI*, được trộn, và phản ứng liên kết được tiến hành. Sau đó, JM109 được biến nạp với sản phẩm liên kết này, và plasmid được tách ra từ thể thực khuẩn thể hiện khả năng kháng Cm để thu được pSTV28-Ptac-*xylD*(Hse)-Ttrp.

Các plasmid đối với sự biểu hiện của *xylD*(Amis), *xylD*(Aor), *xylX*(Cne), *xylX*(Zga), *xylX*(Tco), *xylA*(Abr), và *ycbD*, pSTV28-Ptac-*xylD*(Amis)-Ttrp, pSTV28-Ptac-*xylD*(Aor)-Ttrp, pSTV28-Ptac-*xylX*(Cne)-Ttrp, pSTV28-Ptac-*xylX*(Zga)-Ttrp, pSTV28-Ptac-*xylX*(Tco)-Ttrp, pSTV28-Ptac-*xylA*(Abr)-Ttrp, và pSTV28-Ptac-*ycbD*-Ttrp, cũng được điều chế theo phương pháp tương tự. Việc tối ưu hóa codon không được tiến hành đối với *YcbD*.

Plasmid đối với sự biểu hiện của *xylD*(Atu), pSTV28-Ptac-*xylD*(Atu)-Ttrp, được điều chế như sau. Đoạn ADN có các trình tự của đoạn khởi đầu tac, *xylD*(Atu), và yếu tố kết thúc *trp* (Ptac-*xylD*(Atu)-Ttrp) được tổng hợp, và được liên kết với vector pJET1.2 (được mua từ Thermo Fischer Scientific) được phân cắt bởi *EcoRV* để thu được pJET1.2-Ptac-*xylD*(Atu)-Ttrp. Khi đoạn ADN được tổng hợp, các codon được tối ưu hóa sao cho đoạn này là thích hợp để biểu hiện trong *E. coli*. Những lượng tương đương của pSTV28 và pJET1.2-Ptac-*xylD*(Atu)-Ttrp, cả hai được phân cắt bởi *EcoRI* và *KpnI*, được trộn, và phản ứng liên kết được tiến hành. Sau đó, JM109 được biến nạp với sản phẩm liên kết này, và plasmid được tách ra từ thể thực khuẩn thể hiện khả năng kháng Cm để thu được pSTV28-Ptac-*xylD*(Atu)-Ttrp.

Các plasmid đối với sự biểu hiện của *xylX*(Atu) hoặc *xylA*(Hbo), pSTV28-Ptac-*xylX*(Atu)-Ttrp và pSTV28-Ptac-*xylA*(Hbo)-Ttrp, cũng được điều chế theo phương

pháp tương tự.

Plasmid đối với sự biểu hiện của *xylX*(Art), pSTV28-Ptac-*xylX*(Art)-Ttrp, được điều chế như sau. Đoạn ADN có các trình tự của đoạn khởi đầu tac, *xylX*(Art), và yếu tố khởi đầu trp (Ptac-*xylX*(Art)-Ttrp) được tổng hợp, và được liên kết với vector pCC1 (được mua từ Epicentre) được phân cắt bởi *EcoRV* để thu được pCC1-Ptac-*xylX*(Art)-Ttrp. Khi đoạn ADN được tổng hợp, các codon được tối ưu hóa sao cho đoạn này là thích hợp để biểu hiện trong *E. coli*. Những lượng tương đương của pSTV28 và pCC1-Ptac-*xylX*(Art)-Ttrp, cả hai được phân cắt bởi *EcoRI* và *KpnI*, được trộn, và phản ứng liên kết được tiến hành. Sau đó, JM109 được biến nạp với sản phẩm liên kết này, và plasmid được tách ra từ thể thực khuẩn thể hiện khả năng kháng Cm để thu được pSTV28-Ptac-*xylX*(Art)-Ttrp.

(5) Xác định hoạt tính của các đồng đẳng *xylD*

Chủng *P. ananatis* NA2 Δ *xylD* được biến nạp với pSTV28-Ptac-Ttrp, pSTV28-Ptac-*xylD*-Ttrp, pSTV28-Ptac-*xylD*(Atu)-Ttrp, pSTV28-Ptac-*xylD*(Hse)-Ttrp, pSTV28-Ptac-*yjhG*-Ttrp, pSTV28-Ptac-*yagF*-Ttrp, pSTV28-Ptac-*xylD*(Amis)-Ttrp, hoặc pSTV28-Ptac-*xylD*(Aor)-Ttrp bằng phương pháp mở lỗ bằng điện (đề cập đến patent Mỹ số 6,682,912). Đối với việc nuôi cấy các thể biến nạp, môi trường chứa trong đĩa chứa LBGM9 mà kanamycin, tetracycline và chloramphenicol được bổ sung ở các nồng độ cuối 40 mg/l, 12,5 mg/l và 25 mg/l, được sử dụng.

Các tế bào của mỗi thể biến nạp được nuôi cấy qua đêm ở 34°C trên đĩa LBGM9 mà các dược chất được bổ sung được cạo với lượng tương ứng với 1/6 tế bào trên đĩa, được ủ trong 5 ml môi trường MSII-SX được chứa trong ống thử nghiệm lớn, và được nuôi cấy ở 34°C và 120 vòng/phút trong 48 giờ, và lượng axit L-glutamic (Glu) được tích lũy được đo. Các kết quả được thể hiện trong Fig. 5.

Trong khi chủng *P. ananatis* NA2 Δ *xylD* được nhập pSTV28-Ptac-Ttrp tích lũy 6,9 g/l axit L-glutamic, các thể biến nạp khác tích lũy 11,1 đến 23,1 g/l axit L-glutamic. Ở chủng được nhập pSTV28-Ptac-Ttrp, axit L-glutamic hầu như được tạo ra từ xyloza, và do đó được xem như là axit L-glutamic được tạo ra từ xyloza thông qua quá trình NXA ở các thể biến nạp khác. Đó là, đã chứng minh được rằng đồng đẳng *xylD* thu được từ loài sinh vật bất kỳ khác *C. crescentus* có thể thay thế cho *xylD*.

(6) Xác định hoạt tính của các đồng đẳng *XylX*

Chủng *P. ananatis* NA2 $\Delta xylX$ được biến nạp với pSTV28-Ptac-Ttrp, pSTV28-Ptac-xylX-Ttrp, pSTV28-Ptac-xylX(Art)-Ttrp, pSTV28-Ptac-xylX(Atu)-Ttrp, pSTV28-Ptac-xylX(Cne)-Ttrp, pSTV28-Ptac-xylX(Zga)-Ttrp, pSTV28-Ptac-xylX(Tco)-Ttrp, hoặc pSTV28-Ptac-xylX(Selo)-Ttrp bằng phương pháp mở lỗ bằng điện. Đối với việc nuôi cấy các thể biến nạp, môi trường chứa trong đĩa chứa LBGM9 mà kanamycin, tetracyclin và chloramphenicol được bổ sung ở các nồng độ cuối 40 mg/l, 12,5 mg/l và 25 mg/l, được sử dụng.

Các tế bào của mỗi thể biến nạp được nuôi cấy qua đêm ở 34°C trên đĩa LBGM9 mà được chất được bổ sung được cạo ra với lượng tương ứng với 1/6 tế bào trên đĩa, được ủ trong 5 ml môi trường MSII-SX chứa trong ống thử nghiệm lớn, và được nuôi cấy ở 34°C và 120 vòng/phút trong 48 giờ, và lượng của axit L-glutamic (Glu) được tích lũy được đo. Các kết quả được thể hiện trong Fig. 6.

Trong khi chủng *P. ananatis* NA2 $\Delta xylX$ được nhập pSTV28-Ptac-Ttrp tích lũy 8,7 g/l axit L-glutamic, các thể biến nạp khác tích lũy 20,5 đến 27,8 g/l axit L-glutamic. Ở chủng được nhập pSTV28-Ptac-Ttrp, axit L-glutamic hầu như được tạo ra từ xyloza, và do đó được xem như là axit L-glutamic được tạo ra từ xyloza thông qua quá trình NXA ở các thể biến nạp khác. Đó là, đã chứng minh được rằng đồng phân *xylX* thu được từ loài sinh vật bất kỳ khác *C. crescentus* có thể thay thế cho *xylX*.

(7) Xác định hoạt tính của các đồng đẳng XylA

Chủng *P. ananatis* NA2 $\Delta ccrxylA$ được biến nạp với pSTV28-Ptac-Ttrp, pSTV28-Ptac-ccrxylA-Ttrp, pSTV28-Ptac-ycbD-Ttrp, pSTV28-Ptac-xylA(Hbo)-Ttrp, hoặc pSTV28-Ptac-xylA(Abr)-Ttrp bằng phương pháp mở lỗ bằng điện. Đối với việc nuôi cấy các thể biến nạp, môi trường chứa trong đĩa chứa LBGM9 mà kanamycin, tetracyclin và chloramphenicol được bổ sung ở các nồng độ cuối 40 mg/l, 12,5 mg/l và 25 mg/l, được sử dụng.

Các tế bào của mỗi thể biến nạp được nuôi cấy qua đêm ở 34°C trên đĩa LBGM9 mà các được chất được bổ sung được cạo ra với lượng tương ứng với 1/6 tế bào trên đĩa, được ủ trong 5 ml môi trường MSII-SX được chứa trong ống thử nghiệm lớn, và được nuôi cấy ở 34°C và 120 vòng/phút trong 48 giờ, và lượng của axit L-glutamic (Glu) được tích lũy được đo. Các kết quả được thể hiện trong Fig. 7.

Trong khi chủng *P. ananatis* NA2 $\Delta ccrxylA$ được nhập pSTV28-Ptac-Ttrp tích

lũy 1,0 g/l axit L-glutamic, các thể biến nạp khác tích lũy 18,1 đến 30,7 g/l axit L-glutamic. Ở chủng được nhập pSTV28-Ptac-Ttrp, axit L-glutamic hầu như được tạo ra từ xyloza, và do đó được xem như là axit L-glutamic được tạo ra từ xyloza thông qua quá trình NXA trong các thể biến nạp khác. Đó là, đã chứng minh được rằng đồng đẳng *ccrxylA* thu được từ loài sinh vật bất kỳ khác *C. crescentus* có thể thay thế cho *xylA*.

Khả năng ứng dụng trong công nghiệp

Theo sáng chế, hợp chất đích có thể được tạo ra một cách hiệu quả bằng cách lên men sử dụng nguyên liệu thô xyloza.

YÊU CẦU BẢO HỘ

1. Phương pháp sản xuất hợp chất đích bao gồm bước nuôi cấy vi khuẩn có khả năng tạo ra hợp chất đích trong môi trường nuôi cấy chứa xyloza làm nguồn cacbon để tạo ra và tích lũy hợp chất đích trong môi trường này, và thu gom hợp chất đích từ môi trường này, trong đó:

hợp chất đích là axit 2-ketoglutaric hoặc dẫn xuất của nó,

vi khuẩn có khả năng sản xuất axit xylonic từ xyloza, và hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza đã được truyền hoặc làm tăng trong vi khuẩn

vi khuẩn là vi khuẩn thuộc loài *Pantoea* hoặc loài *Corynebacterium*, và

dẫn xuất của axit 2-ketoglutaric là hợp chất được chọn từ nhóm bao gồm axit L-glutamic, L-glutamin, L-arginin, L-xitruin, L-ornithin, L-prolin, putresxin, và axit γ -aminobutyric.

2. Phương pháp theo điểm 1, trong đó các hoạt tính enzym của xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza được truyền hoặc được làm tăng trong vi khuẩn bằng cách đưa các gen mã hóa các enzym này, ở dạng có thể biểu hiện, vào vi khuẩn.

3. Phương pháp theo điểm 1 hoặc 2, trong đó các gen mã hóa xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza thu được từ vi sinh vật thuộc loài *Caulobacter*, *Escherichia*, *Agrobacterium*, *Herbaspirillum*, *Actinoplanes*, *Cupriavidus*, *Pseudomonas*, *Zobellia*, *Thermobacillus*, *Arthrobacter*, *Azospirillum*, *Halomonas*, *Bacillus*, hoặc *Aspergillus*.

4. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 3, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza có một trong các đặc điểm sau:

(A) hoạt tính xyloza dehydrogenaza, hoặc hoạt tính xyloza dehydrogenaza và hoạt tính xylonolactonaza đã được truyền hoặc làm tăng trong vi khuẩn, hoặc

(B) vi khuẩn có hoạt tính glucoza dehydrogenaza mà xúc tác phản ứng sản xuất axit xylonic từ xyloza.

5. Phương pháp theo điểm 4, trong đó glucoza dehydrogenaza sử dụng pyroloquinolin

quinon làm coenzym, và vi khuẩn thể hiện hoạt tính glucoza dehydrogenaza vì nó có khả năng sản xuất pyroloquinolin quinon, hoặc được nuôi cấy trong môi trường nuôi cấy chứa pyroloquinolin quinon.

6. Phương pháp theo điểm 4, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza vì nó đã được đưa gen mã hóa xyloza dehydrogenaza, hoặc các gen mã hóa xyloza dehydrogenaza và xylonolactonaza ở dạng có thể biểu hiện.

7. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 6, trong đó vi khuẩn được tiếp tục biến đổi sao cho hoạt tính của 2-ketoglutarat dehydrogenaza giảm.

8. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 7, trong đó vi khuẩn được tiếp tục biến đổi sao cho hoạt tính của succinat dehydrogenaza giảm.

9. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 8, trong đó vi khuẩn này là *Pantoea ananatis*.

10. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 8, trong đó vi khuẩn này là *Corynebacterium glutamicum*.

11. Vi khuẩn có khả năng tạo ra hợp chất đích trong môi trường nuôi cấy chứa xyloza làm nguồn cacbon, trong đó:

hợp chất đích là axit 2-ketoglutaric hoặc dẫn xuất của nó,

vi khuẩn có khả năng sản xuất axit xylonic từ xyloza, và hoạt tính xylonat dehydrataza, hoạt tính 2-keto-3-deoxyxylonat dehydrataza và hoạt tính 2-ketoglutaric semialdehyt dehydrogenaza đã được truyền hoặc làm tăng trong vi khuẩn,

trong đó vi khuẩn là vi khuẩn thuộc loài *Pantoea* hoặc loài *Corynebacterium*, và

trong đó dẫn xuất của axit 2-ketoglutaric là hợp chất được chọn từ nhóm bao gồm axit L-glutamic, L-glutamin, L-arginin, L-xitrulin, L-ornithin, L-prolin, putresxin, và axit γ -aminobutyric.

12. Vi khuẩn theo điểm 11, trong đó các hoạt tính enzym của xylonat dehydrataza, 2-keto-3-deoxyxylonat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza được truyền hoặc được làm tăng trong vi khuẩn bằng cách đưa các gen mã hóa các enzym

này, ở dạng có thể biểu hiện, vào vi khuẩn.

13. Vi khuẩn theo điểm 11 hoặc 12, trong đó các gen mã hóa xylozat dehydrataza, 2-keto-3-deoxyxylozat dehydrataza, và 2-ketoglutaric semialdehyt dehydrogenaza thu được từ vi sinh vật thuộc loài *Caulobacter*, *Escherichia*, *Agrobacterium*, *Herbaspirillum*, *Actinoplanes*, *Cupriavidus*, *Pseudomonas*, *Zobellia*, *Thermobacillus*, *Arthrobacter*, *Azospirillum*, *Halomonas*, *Bacillus*, hoặc *Aspergillus*.

14. Vi khuẩn theo điểm bất kỳ trong số các điểm từ 11 đến 13, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza có một trong các đặc điểm sau:

(A) hoạt tính xyloza dehydrogenaza, hoặc hoạt tính xyloza dehydrogenaza và hoạt tính xylonolactonaza đã được truyền hoặc làm tăng trong vi khuẩn, hoặc

(B) vi khuẩn có hoạt tính glucoza dehydrogenaza mà xúc tác phản ứng sản xuất axit xylonic từ xyloza.

15. Vi khuẩn theo điểm 14, trong đó glucoza dehydrogenaza sử dụng pyroloquinolin quinon làm coenzym, và vi khuẩn thể hiện hoạt tính glucoza dehydrogenaza vì nó có khả năng sản xuất pyroloquinolin quinon, hoặc được nuôi cấy trong môi trường nuôi cấy chứa pyroloquinolin quinon.

16. Vi khuẩn theo điểm 14, trong đó vi khuẩn có thể sản xuất axit xylonic từ xyloza vì nó đã được đưa gen mã hóa xyloza dehydrogenaza, hoặc các gen mã hóa xyloza dehydrogenaza và xylonolactonaza ở dạng có thể biểu hiện.

17. Vi khuẩn theo điểm bất kỳ trong số các điểm từ 11 đến 16, trong đó vi khuẩn được tiếp tục biến đổi sao cho hoạt tính của 2-ketoglutarat dehydrogenaza giảm.

18. Vi khuẩn theo điểm bất kỳ trong số các điểm từ 11 đến 17, trong đó vi khuẩn được tiếp tục biến đổi sao cho hoạt tính của succinat dehydrogenaza giảm.

19. Vi khuẩn theo điểm bất kỳ trong số các điểm từ 11 đến 18, trong đó vi khuẩn này là *Pantoea ananatis*.

20. Vi khuẩn theo điểm bất kỳ trong số các điểm từ 11 đến 18, trong đó vi khuẩn này là *Corynebacterium glutamicum*.

Fig.1

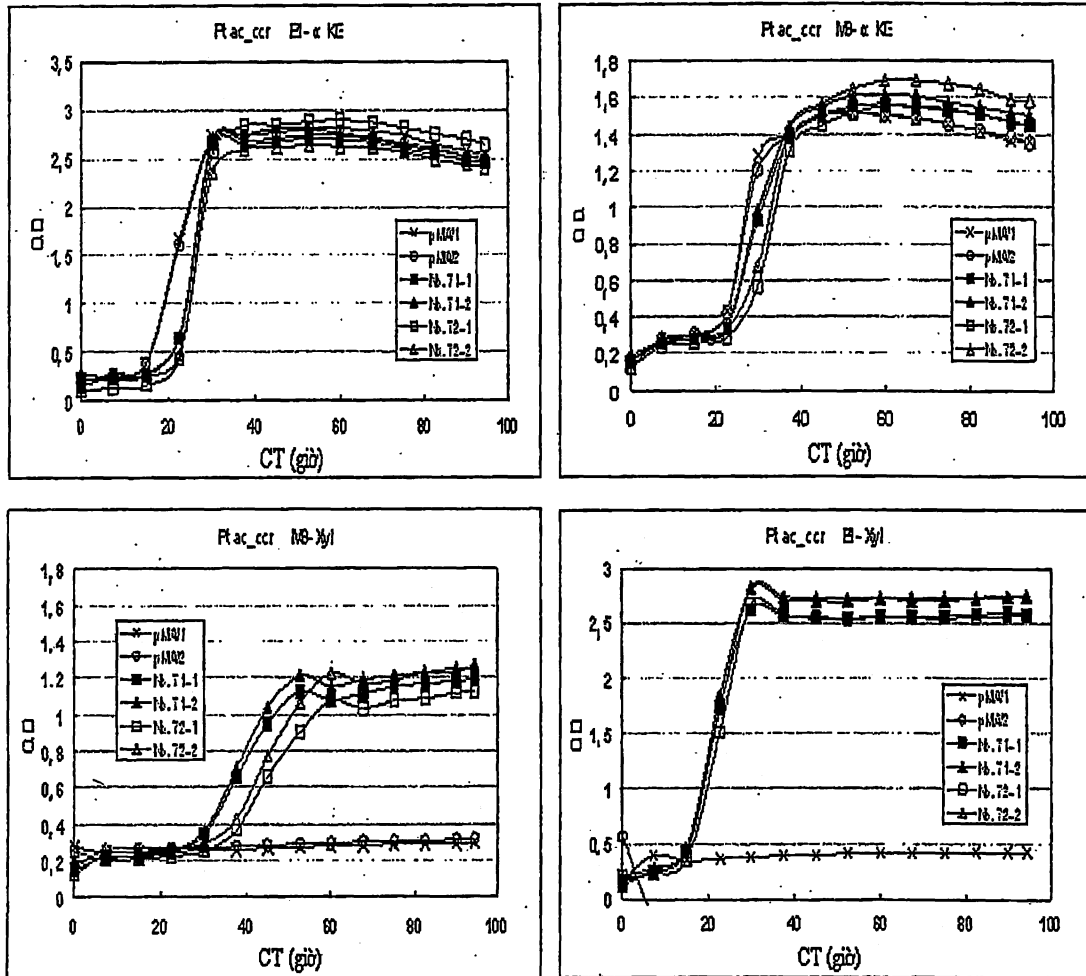


Fig. 2

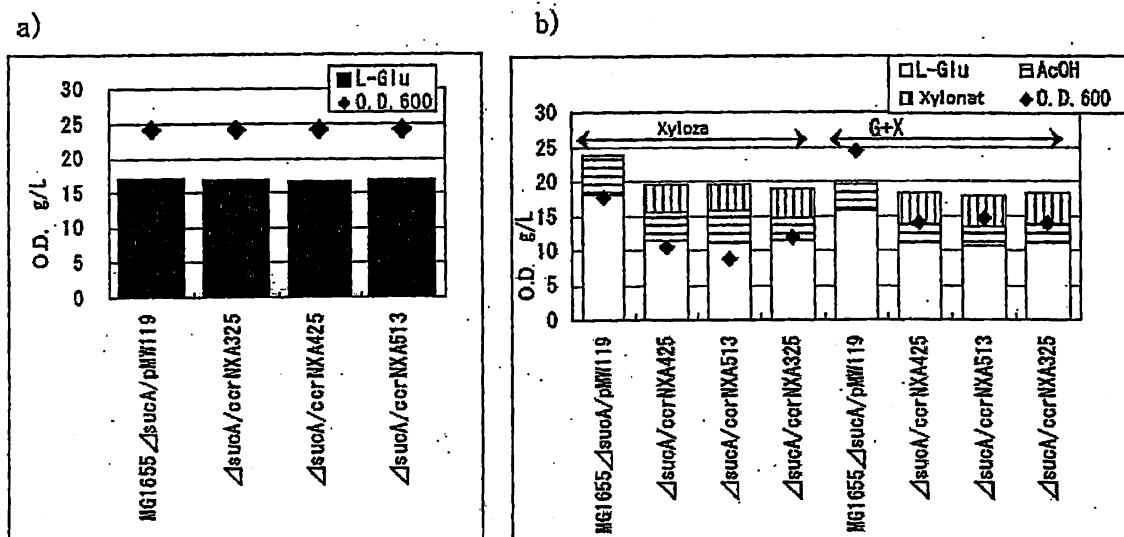


Fig. 3

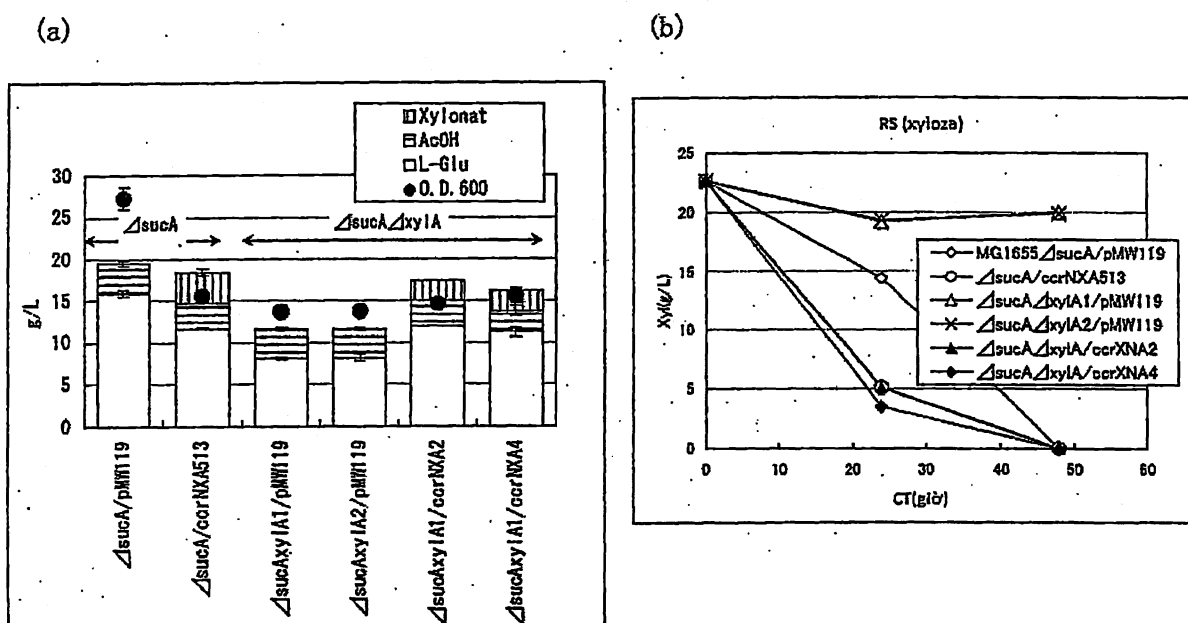


Fig. 4

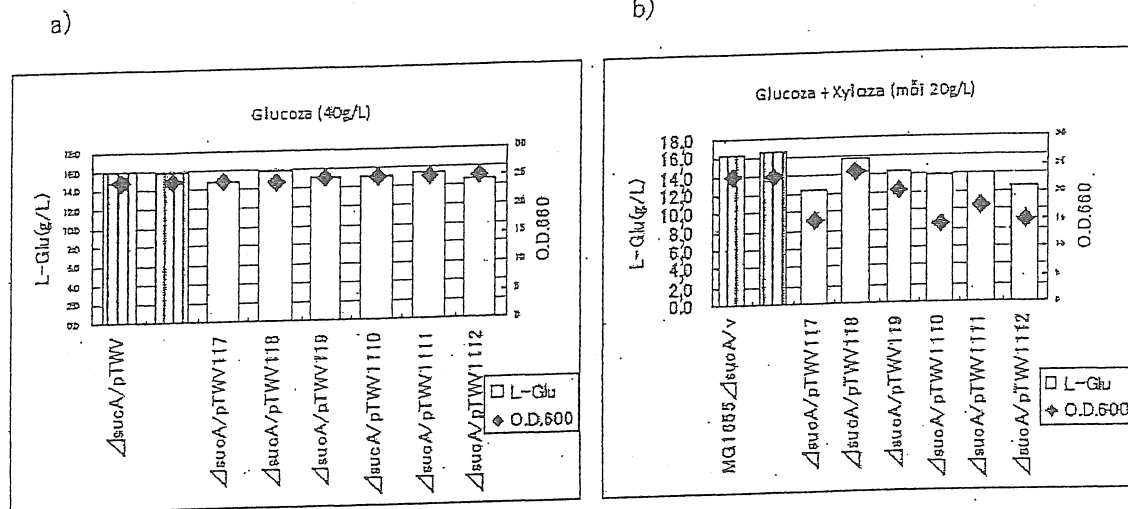


Fig. 5

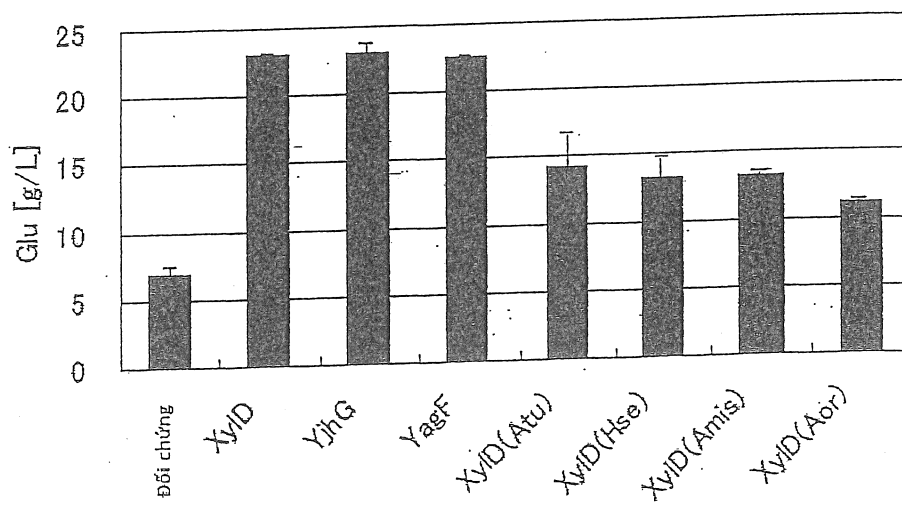


Fig. 6

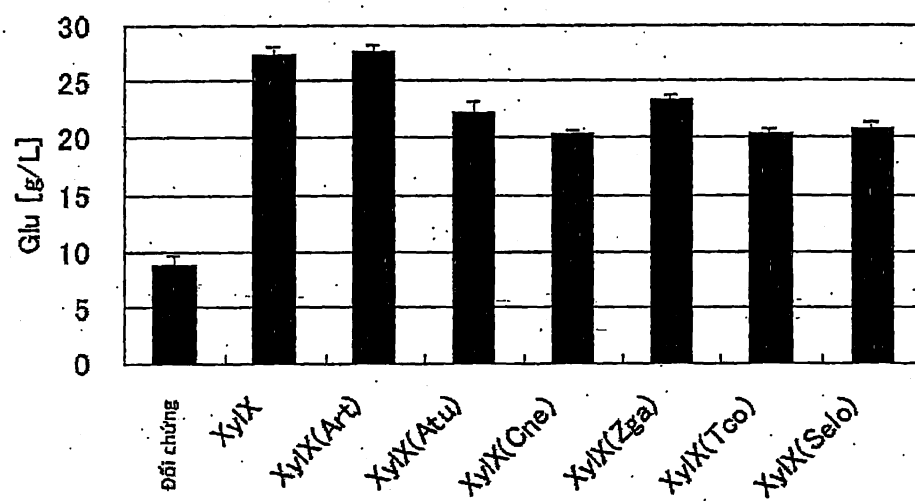
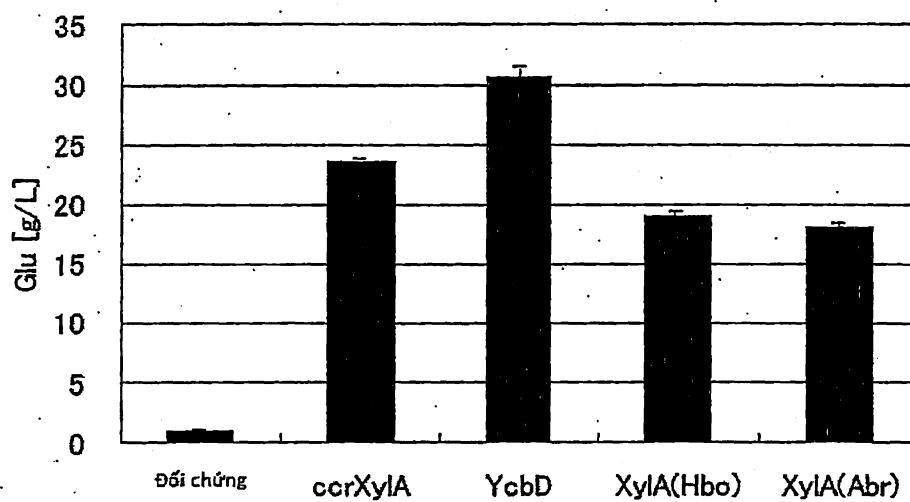


Fig. 7



DANH MỤC TRÌNH TỰ

<110> Ajinomoto Co., Inc.

<120> Phương pháp sản xuất hợp chất đích bằng cách lên men

<130> D313-11346

<150> JP-2011-247031

<151> 2011-11-11

<160> 120

<170> PatentIn version 3.5

<210> 1

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi PtwvPtacf

<400> 1

gaattcgagc tcggtaccca gatctccctg ttgacaatta

40

<210> 2

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi 0823Ptacr

<400> 2

aggaattcac tcacgccac cctcctgtgt gaaattgtta

40

<210> 3

<211> 39

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi Ptac0823f

<400> 3

taacaatttc acacaggagg gtgggcgtga gtgaattcc

39

<210> 4

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi 0819r

 <400> 4
 gcggggcgtg cggtagaca tggcggacct catgctgggg 40

 <210> 5
 <211> 40
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi 0819f

 <400> 5
 cccagcatg aggtccgcca tgtctaaccg caccgcccgc 40

 <210> 6
 <211> 36
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi 219cc0819r

 <400> 6
 ctctagagga tccccttcag cgtttggcga cggaga 36

 <210> 7
 <211> 21
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi 219f

 <400> 7
 ggggatcctc tagagtcgac c 21

 <210> 8
 <211> 21
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi 219r

 <400> 8
 gggtagcgag ctccaattca c 21

 <210> 9
 <211> 35

<212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi attL-F

 <400> 9
 tatattgatt cacttgaagt acgaaaaaaa ccggg 35

 <210> 10
 <211> 45
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi ntpII-R

 <400> 10
 cctgcaggcg gccgctcata gaaggcggcg gtggaatcga aatct 45

 <210> 11
 <211> 45
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi attR-F

 <400> 11
 gcggccgcct gcaggcccat gtaatgaata aaaagcagta attaa 45

 <210> 12
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi attR-R

 <400> 12
 tgaagcggcg caccgaaaaac gcgaaagcgt 30

 <210> 13
 <211> 80
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi Ap-Km-fw

 <400> 13
 ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc aagcttcacg 60
 ctgccgcaag cactcagggc 80

<210> 14
 <211> 80
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi Ap-Km-rv

<400> 14
 atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt cgctcataga 60
 aggcggcggt ggaatcgaaa 80

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

<220>
 <223> đoạn mỗi xyLD_IFS_5742-10-5

<400> 15
 acacaaggag actcccattgt ctaaccgcac gccccgcccgg 40

<210> 16
 <211> 40
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi xyLD_IFS_5742-10-6

<400> 16
 ggaactggcg gctccctcag tggttgtggc ggggcagctt 40

<210> 17
 <211> 45
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0819-01F_4691-88-7

<400> 17
 gtcgactcta gaggatcccc atgtctaacc gcacgccccg ccggt 45

<210> 18
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0819-01R_5659-9-1

<400> 18

aaccaggaac ccggccttct ctgc

24

<210> 19

<211> 24

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0819-02F_5659-9-2

<400> 19

tcccgtacca cgagccgctg gcag

24

<210> 20

<211> 44

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0819-02R_4691-88-10

<400> 20

cgaattcgag ctcggtaccc tcagtgggtg tggcggggca gctt

44

<210> 21

<211> 68

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi xylA-H1P1-5742-5-1

<400> 21

acgacatcat ccatcaccgc cggcattacc tgattatgga gttcaatatg tgaagcctgc
ttttttat

60

68

<210> 22

<211> 68

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi xylA-H2P2-5742-5-2

<400> 22

atcggggccaa cggactgcac agttagccgt tatttgtcga acagataatg cgctcaagtt
agtataaa

60

68

<210> 23

<211> 8296
 <212> ADN
 <213> *Caulobacter crescentus*

<220>
 <221> CDS
 <222> (1175)..(2329)

<220>
 <221> CDS
 <222> (2346)..(3782)

<220>
 <221> CDS
 <222> (3788)..(4534)

<220>
 <221> CDS
 <222> (4659)..(5528)

<400> 23
 caggacttca gaccgcgcgag gctgtagggg tcgtcaaagg ccttggccag gaccggcggc 60
 acctcgcgct tggacaggcc gacgtcgcg acgggtggg tgccgatgtc gacggccttc 120
 ttgcccgcct cgcgccagcg gccggaagggt tcctcggggt tggccggcgg cttgacctgg 180
 gcgtggggcg tggaggccag cagcacggcg gcggtccgg ccatgatcag cagcttgctc 240
 ggcatacaaaa ggcgcctccc gacaaaacac gccacgggaa cggcaagcct cgccgcgggt 300
 tcccgcccg aagtcctctg gcacgaactc taaggcctgc gcacgatcca tcgccaccac 360
 aaagcgtgtc atcccgcccg aagcgcagcg gagagccggg acccaggggc acaagcacgg 420
 cgcttcgtca ccccttgggt cccggatagc gctccgcgct tccgggatga cacgggttta 480
 agcacttggg tgagaagccc tagttcggcg cgcgcgccgc ttccaggctc aacgcggcgt 540
 cgaaggcgct gtcgccagga accgtccctt ccaggggcggc gtcgaacagg tcttcgagcg 600
 atgaccagat gaaggccatc tgctccatga tcgtcgggac gctggccgga tcgaagccgg 660
 ggtagaggct ttcaccagc cacaggcccg tgcggggtc gaaggcgaag cggcagccga 720
 tcaggcggaa ctgggcgcgg gccatggtct cgaacagggc cgtcaggctc cgggcggcgt 780
 ccaggctcgt gtggtccagc accaggcagc ggatctgcag ccagccgtgg tcgggcagca 840
 ggtagaaggc gccctcgtcc tgatcctcgc ccgaaacctc cagcccccg tcgatggctt 900
 cgacgacata gccggccgcg cggcaggtgt cggtagcgc ggccagcagg gcggcttcct 960
 ggtcaggggt caggtcggtc atgggcaaga ggtccaggctc gtggtttgtc ggcggcttct 1020
 agcatggacc gccgcgcgcc gtgaggccga ggatttcgcg ctggtcagac aacctacttg 1080
 ccgtccccac atgttagcgc taccaagtgc cgacgaacgc gcgccgccga cgggtgtcggc 1140
 gottcagacg ctcgagtttt ggggagaaga cgcc gtg ggc gtg agt gaa ttc ctg 1195
 Val Gly Val Ser Glu Phe Leu
 1 5
 ccg gaa gat tgg aaa gcc gcg acc ctg ctg ggg cgc atc gac ttt ggc 1243
 Pro Glu Asp Trp Lys Ala Ala Thr Leu Leu Gly Arg Ile Asp Phe Gly
 10 15 20
 gaa ggc ccg acg ccg gtg ctg gtg cgc ggc ggc cgc gtc gag gac gtc 1291
 Glu Gly Pro Thr Pro Val Leu Val Arg Gly Gly Arg Val Glu Asp Val
 25 30 35
 tcg aag atc gcc ccc acc gtc gct gac ctg atg aac gcc ttc cag ccc 1339
 Ser Lys Ile Ala Pro Thr Val Ala Asp Leu Met Asn Ala Phe Gln Pro
 40 45 50 55
 ggc gcg gtg atc ccg cgc ggc gag gac aag ggt ccg ctg gaa gcc ctc 1387
 Gly Ala Val Ile Pro Arg Gly Glu Asp Lys Gly Pro Leu Glu Ala Leu
 60 65 70
 gac atc cgc cca gtc tgg gaa gac ccg gac ggc gcc gcg ccg gtc aag 1435
 Asp Ile Arg Pro Val Trp Glu Asp Pro Asp Gly Ala Ala Pro Val Lys

75						80				85						
ctg	ttg	gcc	ccc	gtc	gac	ctg	caa	tgc	ctg	aag	gcc	gcc	ggc	gtg	acc	1483
Leu	Leu	Ala	Pro	Val	Asp	Leu	Gln	Cys	Leu	Lys	Ala	Ala	Gly	Val	Thr	
90						95				100						
ttc	gcg	gtc	tcg	acc	ctt	gag	cgg	gtc	atc	gag	gag	cgc	gcg	cgc	ggc	1531
Phe	Ala	Val	Ser	Thr	Leu	Glu	Arg	Val	Ile	Glu	Glu	Arg	Ala	Arg	Gly	
105						110				115						
gac	gcc	ggc	gag	gcg	ctg	aag	atc	cgc	acc	ctg	ctg	gcc	gaa	cgc	atg	1579
Asp	Ala	Gly	Glu	Ala	Leu	Lys	Ile	Arg	Thr	Leu	Leu	Ala	Glu	Arg	Met	
120						125				130						
ggc	ggc	gac	ctc	aag	agc	gtc	gag	ccg	ggc	tcg	cag	ggc	gcc	cag	cgc	1627
Gly	Gly	Asp	Leu	Lys	Ser	Val	Glu	Pro	Gly	Ser	Gln	Gly	Ala	Gln	Arg	
140						145				150						
ctg	aag	gac	gcc	ctg	atc	gcc	gac	ggc	ctg	tgg	tcg	cag	tat	ctg	gaa	1675
Leu	Lys	Asp	Ala	Leu	Ile	Ala	Asp	Gly	Leu	Trp	Ser	Gln	Tyr	Leu	Glu	
155						160				165						
gtg	gcg	atc	ggc	ccg	gac	gcc	gag	atc	ttc	acc	aag	ggc	ccg	acc	ctg	1723
Val	Ala	Ile	Gly	Pro	Asp	Ala	Glu	Ile	Phe	Thr	Lys	Gly	Pro	Thr	Leu	
170						175				180						
tcc	tcg	atg	ggc	tgg	ggc	gac	cag	gtc	ggc	gtc	cgc	tat	gac	agc	cac	1771
Ser	Ser	Met	Gly	Trp	Gly	Asp	Gln	Val	Gly	Val	Arg	Tyr	Asp	Ser	His	
185						190				195						
tgg	aac	aat	ccc	gag	ccg	gaa	gtc	gtg	ctg	ctg	tgc	gac	ggt	tcg	ggc	1819
Trp	Asn	Asn	Pro	Glu	Pro	Glu	Val	Val	Leu	Leu	Cys	Asp	Gly	Ser	Gly	
200						205				210						
ctg	atc	cgc	ggc	gcg	gcg	ctg	ggc	aac	gac	gtc	aat	ctg	cgc	gac	ttc	1867
Leu	Ile	Arg	Gly	Ala	Ala	Leu	Gly	Asn	Asp	Val	Asn	Leu	Arg	Asp	Phe	
220						225				230						
gaa	ggt	cgt	tcg	gcc	ctg	ctg	ctc	agc	aag	gcc	aag	gac	aac	aac	gcc	1915
Glu	Gly	Arg	Ser	Ala	Leu	Leu	Leu	Ser	Lys	Ala	Lys	Asp	Asn	Asn	Ala	
235						240				245						
agc	tgc	gcc	atc	ggt	ccg	ttc	ttc	cgc	ctg	ttc	gac	gag	acc	ttc	ggc	1963
Ser	Cys	Ala	Ile	Gly	Pro	Phe	Phe	Arg	Leu	Phe	Asp	Glu	Thr	Phe	Gly	
250						255				260						
ctg	gac	gac	gtc	cgt	tcg	gcc	gag	gtc	gag	ctg	aag	atc	acc	ggc	cgc	2011
Leu	Asp	Asp	Val	Arg	Ser	Ala	Glu	Val	Glu	Leu	Lys	Ile	Thr	Gly	Arg	
265						270				275						
gac	aac	ttc	gtg	ctc	gac	ggc	aag	tcg	aac	atg	agc	ctg	atc	agc	cgc	2059
Asp	Asn	Phe	Val	Leu	Asp	Gly	Lys	Ser	Asn	Met	Ser	Leu	Ile	Ser	Arg	
280						285				290						
gac	ccg	gcc	gtg	ctg	gcc	gga	cag	gcc	tat	ggc	aag	cag	cac	cag	tat	2107
Asp	Pro	Ala	Val	Leu	Ala	Gly	Gln	Ala	Tyr	Gly	Lys	Gln	His	Gln	Tyr	
300						305				310						
ccg	gac	ggc	ttt	gct	ttg	ttc	ctg	ggc	acc	atg	ttc	gcc	ccg	atc	cag	2155
Pro	Asp	Gly	Phe	Ala	Leu	Phe	Leu	Gly	Thr	Met	Phe	Ala	Pro	Ile	Gln	
315						320				325						
gac	cgc	gac	acc	ccc	ggc	cag	ggt	ttc	acc	cac	aag	gtc	ggc	gac	cgc	2203
Asp	Arg	Asp	Thr	Pro	Gly	Gln	Gly	Phe	Thr	His	Lys	Val	Gly	Asp	Arg	
330						335				340						
gtg	cgt	gtc	tcg	acg	ccg	aag	ctg	ggc	gtg	ctc	gag	aac	gaa	gtc	acc	2251
Val	Arg	Val	Ser	Thr	Pro	Lys	Leu	Gly	Val	Leu	Glu	Asn	Glu	Val	Thr	
345						350				355						
acc	tgc	gac	aag	gcc	aag	ccg	tgg	acg	ttc	ggc	atc	tcg	gcc	ctg	atc	2299
Thr	Cys	Asp	Lys	Ala	Lys	Pro	Trp	Thr	Phe	Gly	Ile	Ser	Ala	Leu	Ile	
360						365				370						
cgc	aac	ctg	gcc	ggc	cgc	ggc	ctc	ctc	taa	tcgaaagc	ttt	cccgcg	atg	acc	2351	
Arg	Asn	Leu	Ala	Gly	Arg	Gly	Leu	Leu					Met	Thr		

				380													385	
gac	acc	ctg	cgc	cat	tac	atc	ggc	ggc	gaa	cgc	gtc	gcg	gcc	gac	gcc			2399
Asp	Thr	Leu	Arg	His	Tyr	Ile	Gly	Gly	Glu	Arg	Val	Ala	Ala	Asp	Ala			
			390					395					400					
ccg	gcc	gag	agc	ctg	aac	ccg	tcg	aac	acc	aac	gac	gtc	gtc	gcc	aag			2447
Pro	Ala	Glu	Ser	Leu	Asn	Pro	Ser	Asn	Thr	Asn	Asp	Val	Val	Ala	Lys			
		405					410					415						
gtc	ccg	atg	ggc	ggc	cag	gcc	gag	gtc	gac	gcc	gcc	gtc	gac	gcc	gcg			2495
Val	Pro	Met	Gly	Gly	Gln	Ala	Glu	Val	Asp	Ala	Ala	Val	Asp	Ala	Ala			
		420				425					430							
cgc	aag	gcc	ttc	ccg	gcc	tgg	gcc	gac	gcc	tcg	ccg	gag	gtc	cgc	tcg			2543
Arg	Lys	Ala	Phe	Pro	Ala	Trp	Ala	Asp	Ala	Ser	Pro	Glu	Val	Arg	Ser			
	435				440				445					450				
gac	ctc	ttg	gac	aag	gtc	ggg	tcg	acc	atc	atc	gcc	cgc	agc	gcc	gac			2591
Asp	Leu	Leu	Asp	Lys	Val	Gly	Ser	Thr	Ile	Ile	Ala	Arg	Ser	Ala	Asp			
			455					460					465					
atc	ggc	cgc	ctg	ctg	gcc	cgc	gaa	gag	ggc	aag	acc	ctg	gcg	gaa	ggc			2639
Ile	Gly	Arg	Leu	Leu	Ala	Arg	Glu	Glu	Gly	Lys	Thr	Leu	Ala	Glu	Gly			
			470				475						480					
atc	ggc	gag	acc	gtc	cgc	gcc	ggc	cgc	atc	ttc	aag	tac	ttc	gcc	ggg			2687
Ile	Gly	Glu	Thr	Val	Arg	Ala	Gly	Arg	Ile	Phe	Lys	Tyr	Phe	Ala	Gly			
		485					490					495						
gaa	gcc	ctg	cgc	cgt	cac	ggc	cag	aac	ctg	gaa	agc	acc	cgt	ccg	ggc			2735
Glu	Ala	Leu	Arg	Arg	His	Gly	Gln	Asn	Leu	Glu	Ser	Thr	Arg	Pro	Gly			
		500				505					510							
gtc	gag	atc	cag	acc	tat	cgt	cag	gcc	gtg	ggc	gtc	tat	ggc	ctg	atc			2783
Val	Glu	Ile	Gln	Thr	Tyr	Arg	Gln	Ala	Val	Gly	Val	Tyr	Gly	Leu	Ile			
	515				520				525					530				
acg	ccc	tgg	aac	ttc	ccg	atc	gcc	atc	ccg	gcc	tgg	aag	gcc	gcc	ccg			2831
Thr	Pro	Trp	Asn	Phe	Pro	Ile	Ala	Ile	Pro	Ala	Trp	Lys	Ala	Ala	Pro			
			535						540				545					
gcg	ctc	gcc	ttc	ggc	aac	acc	gtg	gtg	atc	aag	ccg	gcc	ggc	ccg	acg			2879
Ala	Leu	Ala	Phe	Gly	Asn	Thr	Val	Val	Ile	Lys	Pro	Ala	Gly	Pro	Thr			
			550				555						560					
ccc	gcc	acc	gcc	aac	gtg	ctg	gcc	gac	atc	atg	gcc	gag	tgc	ggc	gcc			2927
Pro	Ala	Thr	Ala	Asn	Val	Leu	Ala	Asp	Ile	Met	Ala	Glu	Cys	Gly	Ala			
		565					570					575						
ccg	gcc	ggc	gtg	ttc	aac	atg	ctg	ttt	ggg	cgc	ggc	tcg	atg	ggc	gat			2975
Pro	Ala	Gly	Val	Phe	Asn	Met	Leu	Phe	Gly	Arg	Gly	Ser	Met	Gly	Asp			
		580				585					590							
gcg	ctg	atc	aag	cac	aag	gac	gtg	gac	ggc	gtc	tcg	ttc	acc	ggc	tcg			3023
Ala	Leu	Ile	Lys	His	Lys	Asp	Val	Asp	Gly	Val	Ser	Phe	Thr	Gly	Ser			
		595			600				605				610					
cag	ggc	gtg	ggc	gcg	cag	gtc	gcc	gcc	gcc	gcc	gtg	gcc	cgt	cag	gcc			3071
Gln	Gly	Val	Gly	Ala	Gln	Val	Ala	Ala	Ala	Ala	Val	Ala	Arg	Gln	Ala			
			615				620						625					
cgc	gtg	cag	ctg	gag	atg	ggc	ggc	aag	aac	ccg	ctg	atc	gtg	ctg	gac			3119
Arg	Val	Gln	Leu	Glu	Met	Gly	Gly	Lys	Asn	Pro	Leu	Ile	Val	Leu	Asp			
		630					635						640					
gac	gcc	gac	ctg	gag	cgc	gcg	gtc	gcc	atc	gcc	ctg	gac	ggc	tcg	ttc			3167
Asp	Ala	Asp	Leu	Glu	Arg	Ala	Val	Ala	Ile	Ala	Leu	Asp	Gly	Ser	Phe			
		645					650					655						
ttc	gcc	acc	ggc	cag	cgc	tgc	acc	gcc	agc	tcg	cgc	ctg	atc	gtc	cag			3215
Phe	Ala	Thr	Gly	Gln	Arg	Cys	Thr	Ala	Ser	Ser	Arg	Leu	Ile	Val	Gln			
		660				665					670							
gac	ggg	att	cac	gac	aag	ttc	gtg	gcc	ctg	ctg	gcc	gag	aag	gtc	gcc			3263
Asp	Gly	Ile	His	Asp	Lys	Phe	Val	Ala	Leu	Leu	Ala	Glu	Lys	Val	Ala			

675	680	685	690	
gcc ctg cgc gtg ggc gac gct ctg gac ccc aac acc cag atc ggc ccg				3311
Ala Leu Arg Val Gly Asp Ala Leu Asp Pro Asn Thr Gln Ile Gly Pro				
695	700	705		
gcc gtc tcc gaa gac cag atg gag act tcg tac cgc tat atc gac atc				3359
Ala Val Ser Glu Asp Gln Met Glu Thr Ser Tyr Arg Tyr Ile Asp Ile				
710	715	720		
gct gcc tcc gaa ggc ggc cgc gtg gtc acc ggc ggc gac cgc atc aag				3407
Ala Ala Ser Glu Gly Gly Arg Val Val Thr Gly Gly Asp Arg Ile Lys				
725	730	735		
ctc gac aat ccg ggc tgg tac gtg cgt ccg acc ctg atc gcc gac acg				3455
Leu Asp Asn Pro Gly Trp Tyr Val Arg Pro Thr Leu Ile Ala Asp Thr				
740	745	750		
caa gcc ggc atg cgg atc aac gag gag gtc ttc ggc ccc gtc gcc				3503
Gln Ala Gly Met Arg Ile Asn Asn Glu Glu Val Phe Gly Pro Val Ala				
755	760	765	770	
tcg acc atc cgc gtc aag agc tac gaa gag gcg ctg gag atc gcc aac				3551
Ser Thr Ile Arg Val Lys Ser Tyr Glu Glu Ala Leu Glu Ile Ala Asn				
775	780	785		
ggc gtc gag ttc ggc ctt tcg gcc ggc atc gcc acc acc tcg ctc aag				3599
Gly Val Glu Phe Gly Leu Ser Ala Gly Ile Ala Thr Thr Ser Leu Lys				
790	795	800		
cac gcc cgc cac ttc cag cgc tat gcc cgc gcg ggc atg acc atg gtc				3647
His Ala Arg His Phe Gln Arg Tyr Ala Arg Ala Gly Met Thr Met Val				
805	810	815		
aat ctg gcc acg gcc ggc gtc gac tat cac gtg ccg ttc ggc ggc acg				3695
Asn Leu Ala Thr Ala Gly Val Asp Tyr His Val Pro Phe Gly Gly Thr				
820	825	830		
aag agc agc tcg tac ggc gcc cgc gag cag ggc ttc gcg gcg gtc gag				3743
Lys Ser Ser Ser Tyr Gly Ala Arg Glu Gln Gly Phe Ala Ala Val Glu				
835	840	845	850	
ttc ttc acc cag acc aaa acc tcc tac tcg tgg tcg taa gcagc atg tcc				3793
Phe Phe Thr Gln Thr Lys Thr Ser Tyr Ser Trp Ser Met Ser				
855	860			
tca gcc atc tat ccc agc ctg aag ggc aag cgc gtc gtc atc acc ggc				3841
Ser Ala Ile Tyr Pro Ser Leu Lys Gly Lys Arg Val Val Ile Thr Gly				
865	870	875	880	
ggc ggc tcg ggc atc ggg gcc ggc ctc acc gcc ggc ttc gcc cgt cag				3889
Gly Gly Ser Gly Ile Gly Ala Gly Leu Thr Ala Gly Phe Ala Arg Gln				
885	890	895		
ggc gcg gag gtg atc ttc ctc gac atc gcc gac gag gac tcc agg gct				3937
Gly Ala Glu Val Ile Phe Leu Asp Ile Ala Asp Glu Asp Ser Arg Ala				
900	905	910		
ctt gag gcc gag ctg gcc ggc tcg ccg atc ccg ccg gtc tac aag cgc				3985
Leu Glu Ala Glu Leu Ala Gly Ser Pro Ile Pro Pro Val Tyr Lys Arg				
915	920	925		
tgc gac ctg atg aac ctc gag gcg atc aag gcg gtc ttc gcc gag atc				4033
Cys Asp Leu Met Asn Leu Glu Ala Ile Lys Ala Val Phe Ala Glu Ile				
930	935	940		
ggc gac gtc gac gtg ctg gtc aac aac gcc ggc aat gac gac cgc cac				4081
Gly Asp Val Asp Val Leu Val Asn Asn Ala Gly Asn Asp Asp Arg His				
945	950	955	960	
aag ctg gcc gac gtg acc ggc gcc tat tgg gac gag ccg atc aac gtc				4129
Lys Leu Ala Asp Val Thr Gly Ala Tyr Trp Asp Glu Arg Ile Asn Val				
965	970	975		
aac ctg cgc cac atg ctg ttc tgc acc cag gcc gtc gcg ccg ggc atg				4177
Asn Leu Arg His Met Leu Phe Cys Thr Gln Ala Val Ala Pro Gly Met				

980	985	990	
aag aag cgt ggc ggc ggg gcg gtg atc aac ttc ggt tcg atc agc tgg			4225
Lys Lys Arg Gly Gly Gly Ala Val Ile Asn Phe Gly Ser Ile Ser Trp			
995	1000	1005	
cac ctg ggg ctt gag gac ctc gtc ctc tac gaa acc gcc aag gcc			4270
His Leu Gly Leu Glu Asp Leu Val Leu Tyr Glu Thr Ala Lys Ala			
1010	1015	1020	
ggc atc gaa ggc atg acc cgc gcg ctg gcc cgg gag ctg ggt ccc			4315
Gly Ile Glu Gly Met Thr Arg Ala Leu Ala Arg Glu Leu Gly Pro			
1025	1030	1035	
gac gac atc cgc gtc acc tgc gtg gtg ccg ggc aac gtc aag acc			4360
Asp Asp Ile Arg Val Thr Cys Val Val Pro Gly Asn Val Lys Thr			
1040	1045	1050	
aag cgc cag gag aag tgg tac acg ccc gaa ggc gag gcc cag atc			4405
Lys Arg Gln Glu Lys Trp Tyr Thr Pro Glu Gly Glu Ala Gln Ile			
1055	1060	1065	
gtg gcg gcc caa tgc ctg aag ggc cgc atc gtc ccg gag aac gtc			4450
Val Ala Ala Gln Cys Leu Lys Gly Arg Ile Val Pro Glu Asn Val			
1070	1075	1080	
gcc gcg ctg gtg ctg ttc ctg gcc tcg gat gac gcg tcg ctc tgc			4495
Ala Ala Leu Val Leu Phe Leu Ala Ser Asp Asp Ala Ser Leu Cys			
1085	1090	1095	
acc ggc cac gaa tac tgg atc gac gcc ggc tgg cgt tga cctaagaaaa			4544
Thr Gly His Glu Tyr Trp Ile Asp Ala Gly Trp Arg			
1100	1105	1110	
ctgtcatccc ggcccagcgt gaagcgcgcc gagccgggac cacggcaagc gccacgcgtc			4604
cggagggtccc ggctctccgc tgtgctacgg ccgggatgac agaggaatga ttgt atg			4661
		Met	
acc gct caa gtc act tgc gta tgg gat ctg aag gcc acg ttg ggc			4706
Thr Ala Gln Val Thr Cys Val Trp Asp Leu Lys Ala Thr Leu Gly			
1115	1120	1125	
gaa ggc ccg atc tgg cat ggc gac acc ctg tgg ttc gtc gac atc			4751
Glu Gly Pro Ile Trp His Gly Asp Thr Leu Trp Phe Val Asp Ile			
1130	1135	1140	
aag cag cgt aaa atc cac aac tac cac ccc gcc acc ggc gag cgc			4796
Lys Gln Arg Lys Ile His Asn Tyr His Pro Ala Thr Gly Glu Arg			
1145	1150	1155	
ttc agc ttc gac gcg ccg gat cag gtg aac ttc ctc gcg ccg atc			4841
Phe Ser Phe Asp Ala Pro Asp Gln Val Thr Phe Leu Ala Pro Ile			
1160	1165	1170	
gtc ggc gcg acc ggc ttt gtc gtc ggt ctg aag acc ggg att cac			4886
Val Gly Ala Thr Gly Phe Val Val Gly Leu Lys Thr Gly Ile His			
1175	1180	1185	
cgc ttc cac ccg gcc acg ggc ttc agc ctg ctg ctc gag gtc gag			4931
Arg Phe His Pro Ala Thr Gly Phe Ser Leu Leu Leu Glu Val Glu			
1190	1195	1200	
gac gcg gcg ctg aac aac cgc ccc aac gac gcc acg gtc gac gcg			4976
Asp Ala Ala Leu Asn Asn Arg Pro Asn Asp Ala Thr Val Asp Ala			
1205	1210	1215	
caa ggc cgt ctg tgg ttc ggc acc atg cac gac ggg gaa gag aac			5021
Gln Gly Arg Leu Trp Phe Gly Thr Met His Asp Gly Glu Glu Asn			
1220	1225	1230	
aat agc ggc tcg ctc tat cgg atg gac ctc acc ggc gtc gcc cgg			5066
Asn Ser Gly Ser Leu Tyr Arg Met Asp Leu Thr Gly Val Ala Arg			
1235	1240	1245	
atg gac cgc gac atc tgc atc acc aac ggc ccg tgc gtc tcg ccc			5111

Met	Asp	Arg	Asp	Ile	Cys	Ile	Thr	Asn	Gly	Pro	Cys	Val	Ser	Pro	
			1250					1255					1260		
gac	ggc	aag	acc	ttc	tac	cac	acc	gac	acc	ctg	gaa	aag	acg	atc	5156
Asp	Gly	Lys	Thr	Phe	Tyr	His	Thr	Asp	Thr	Leu	Glu	Lys	Thr	Ile	
			1265					1270					1275		
tac	gcc	ttc	gac	ctg	gcc	gag	gac	ggc	ctg	ctg	tcg	aac	aag	cgc	5201
Tyr	Ala	Phe	Asp	Leu	Ala	Glu	Asp	Gly	Leu	Leu	Ser	Asn	Lys	Arg	
			1280					1285					1290		
gtc	ttc	gtg	cag	ttc	gcc	ctg	ggc	gac	gat	gtc	tat	ccg	gac	ggt	5246
Val	Phe	Val	Gln	Phe	Ala	Leu	Gly	Asp	Asp	Val	Tyr	Pro	Asp	Gly	
			1295					1300					1305		
tcg	gtc	gtc	gat	tcc	gaa	ggc	tat	ctg	tgg	acc	gcc	ctg	tgg	ggc	5291
Ser	Val	Val	Asp	Ser	Glu	Gly	Tyr	Leu	Trp	Thr	Ala	Leu	Trp	Gly	
			1310					1315					1320		
ggt	ttc	ggc	gcg	gtc	cgc	ttc	tcg	ccg	caa	ggc	gac	gcc	gtg	acg	5336
Gly	Phe	Gly	Ala	Val	Arg	Phe	Ser	Pro	Gln	Gly	Asp	Ala	Val	Thr	
			1325					1330					1335		
cgc	atc	gaa	ctg	ccc	gcc	ccc	aac	gtc	acc	aag	ccc	tgc	ttc	ggc	5381
Arg	Ile	Glu	Leu	Pro	Ala	Pro	Asn	Val	Thr	Lys	Pro	Cys	Phe	Gly	
			1340					1345					1350		
ggg	cct	gac	ctg	aag	acc	ctc	tat	ttc	acc	acc	gcc	cgc	aag	ggc	5426
Gly	Pro	Asp	Leu	Lys	Thr	Leu	Tyr	Phe	Thr	Thr	Ala	Arg	Lys	Gly	
			1355					1360					1365		
ctg	agc	gac	gag	acc	ctg	gcc	cag	tac	ccg	ctg	gcc	ggc	ggt	gtg	5471
Leu	Ser	Asp	Glu	Thr	Leu	Ala	Gln	Tyr	Pro	Leu	Ala	Gly	Gly	Val	
			1370					1375					1380		
ttc	gcc	gtt	ccg	gtc	gat	gtg	gcc	ggc	caa	ccc	cag	cat	gag	gtc	5516
Phe	Ala	Val	Pro	Val	Asp	Val	Ala	Gly	Gln	Pro	Gln	His	Glu	Val	
			1385					1390					1395		
cgc	ctt	gtc	taa	ccgcacgccc	cgccgggttcc	ggccccgcga	ttggttcgat								5568
Arg	Leu	Val													
aacccccgacc	atatcgacat	gaccgcgctc	tatctggagc	gcttcatgaa	ctacggggatc										5628
acgccggagg	agctgcgcag	cggcaagccg	atcatcgga	tcgcccagac	cggcagcgac										5688
atctcgccct	gcaaccgcat	ccacctggac	ctgggtccagc	gggtgcggga	cgggatccgc										5748
gacgcggggg	gcatccccat	ggagttcccg	gtccatccga	tcttcgagaa	ctgccgtcgc										5808
ccgacggcgg	cgctggaccg	gaacctctcg	tacctgggtc	tcgtcgagac	cctgcacggc										5868
tatccgatcg	acgccgtggt	tctgaccacc	ggctgcgaca	agaccacccc	ggccggggatc										5928
atggccgcca	ccacgggtcaa	tatcccggcc	atcgtgctgt	cgggcggccc	gatgctggac										5988
ggctggcacg	agaacgagct	cgtgggctcg	ggcaccgtga	tctggcgctc	gcgcgcgaag										6048
ctggcgggcc	gcgagatcac	cgaggaagag	ttcatcgacc	gcgcgcgcag	ctcggcgccg										6108
tcggcgggcc	actgcaacac	catgggcacg	gcctcgacca	tgaacgcccgt	ggccgagggc										6168
ctgggcctgt	cgctgaccgg	ctgcgcggcc	atccccgccc	cctaccgcga	gcgcggccag										6228
atggcctaca	agaccggcca	gcgcacgtc	gatctggcct	atgacgacgt	caaaccgctc										6288
gacatcctga	ccaagcaagc	cttcgagaac	gccatcgccc	tgggtggcggc	ggccggcgcc										6348
tcgaccaacg	cccagccgca	catcgtggcc	atggcccgtc	acgccggcgt	cgagatcacc										6408
gccgacgact	ggcgcgcgcc	ctatgacatc	ccgctgatcg	tcaacatgca	gccggccggc										6468
aagtatctgg	gcgagcgctt	ccaccgagcc	ggcgcgcgcc	cggcggtgct	gtgggagctg										6528
ttgcagcaag	gcccgcctgca	cggcgacgtg	ctgaccgtca	ccggcaagac	gatgagcgag										6588
aacctgcaag	gcccgcgaac	cagcgaccgc	gaggtgatct	tcccgtaacca	cgagccgctg										6648
gccgagaagg	ccgggttccct	ggtttctcaag	ggcaacctct	tcgacttcgc	gatcatgaag										6708
tccagcgtga	tcggcgagga	gttcgcgaag	cgctacctgt	cgcagcccgg	ccaggaaggc										6768
gtgttcgaag	ccgcgcgccat	cgtgttcgac	ggctcggaag	actatcacia	gcggatcaac										6828
gatccggccc	tggagatcga	cgagcgctgc	atcctgggtga	tccgcggcgc	gggtccgatc										6888
ggctggcccg	gctcggccga	ggtcgtcaac	atgcagccgc	cggatcaoct	tctgaagaag										6948
gggatcatga	gcctgcccac	cctgggcgat	ggccgtcagt	cgggcaccgc	cgacagcccc										7008
tcgatcctga	acgcctcgcc	cgaagcgcg	atcggcgggc	gcctgtcgtg	gctgcgcacc										7068

```

ggcgacacca tccgcatoga cctcaacacc ggccgctgcg acgcccctggt cgaacgaggcg 7128
acgatcgccg cgcgcaagca ggacggcatc ccggcggttc ccgccaccat gacgccctgg 7188
caggaaatct accgcgcccc cgcagtcag ctgcacaccg gcggcggtgt ggagttcgcg 7248
gtcaagtacc aggacctggc ggccaagctg ccccgccaca accactgatg cgaagggcct 7308
tcggagcgat ccggaggccc ttttctttgc gccccccgga tcgcccggcc ccgcttcaca 7368
acaccagacc gaatcgcat tttccgctg ccaaacgctg aaatgcgaag gggctggggg 7428
ggcgccatgg aactcaagga atactttctg ggctgttg ccacatcac cggcctggcg 7488
atcaccgaca tgatcctgag cgtgcacggc ctgctacggc gcgtcagcgc cgtacgatgg 7548
gattggctgc cgtcacggc cgcgccttg gttttcgtgc tcatcgtccg cagctgggtg 7608
atcgctggg agcccgactg gatcaacacg ccgatgtggc agttcctgtt gatcctgatc 7668
cagctgacct gctgttcct ggccgccaag tctgtgttg cggacgacag cgacgccgaa 7728
gaggtcgacc tgatggctca ctactggctg cagaaccggg tcgtctgggc gtccctgatc 7788
ggcatgatca cggcggttcgc cgcgcctcg gtcttcctgc gcatcaacga tccggcgagc 7848
ctgtcgcat ggttctggac gttcggttg gagctgccg tctactgctt gccgctggg 7908
gtcctgatcc tcctgcagc accgatcgt catcgcggc tggttccgc gctcctgctc 7968
gcttggtgg tcctgaacgg gacggacgt atgcactaca gctagggctc ggcggctctg 8028
gtaacaacc caaccccgat caacgaagg ttagaaagct ttaagagaaa cgttgccaat 8088
tgtaaaattc ttctctgggt tgaaacccat ttcggtgta tcgatgaggt gtcgggccaat 8148
ttattgagcg cggacaggat gttgtcagtg agccgagccc ttgatcaacg accggagtcg 8208
cccagcctgg tgatcgccac cagatcctc cagcttggcc tcgttggtgt ggtttggtg 8268
gtggcgctga tcccgcgcgc gctggcgg 8296

```

```

<210> 24
<211> 384
<212> PRT
<213> Caulobacter crescentus

```

```

<400> 24
Val Gly Val Ser Glu Phe Leu Pro Glu Asp Trp Lys Ala Ala Thr Leu
1          5          10          15
Leu Gly Arg Ile Asp Phe Gly Glu Gly Pro Thr Pro Val Leu Val Arg
20        25        30
Gly Gly Arg Val Glu Asp Val Ser Lys Ile Ala Pro Thr Val Ala Asp
35        40        45
Leu Met Asn Ala Phe Gln Pro Gly Ala Val Ile Pro Arg Gly Glu Asp
50        55        60
Lys Gly Pro Leu Glu Ala Leu Asp Ile Arg Pro Val Trp Glu Asp Pro
65        70        75        80
Asp Gly Ala Ala Pro Val Lys Leu Leu Ala Pro Val Asp Leu Gln Cys
85        90        95
Leu Lys Ala Ala Gly Val Thr Phe Ala Val Ser Thr Leu Glu Arg Val
100       105       110
Ile Glu Glu Arg Ala Arg Gly Asp Ala Gly Glu Ala Leu Lys Ile Arg
115       120       125
Thr Leu Leu Ala Glu Arg Met Gly Gly Asp Leu Lys Ser Val Glu Pro
130       135       140
Gly Ser Gln Gly Ala Gln Arg Leu Lys Asp Ala Leu Ile Ala Asp Gly
145       150       155       160
Leu Trp Ser Gln Tyr Leu Glu Val Ala Ile Gly Pro Asp Ala Glu Ile
165       170       175
Phe Thr Lys Gly Pro Thr Leu Ser Ser Met Gly Trp Gly Asp Gln Val
180       185       190
Gly Val Arg Tyr Asp Ser His Trp Asn Asn Pro Glu Pro Glu Val Val
195       200       205
Leu Leu Cys Asp Gly Ser Gly Leu Ile Arg Gly Ala Ala Leu Gly Asn
210       215       220

```

```

Asp Val Asn Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu Leu Ser
225      230      235      240
Lys Ala Lys Asp Asn Asn Ala Ser Cys Ala Ile Gly Pro Phe Phe Arg
      245      250      255
Leu Phe Asp Glu Thr Phe Gly Leu Asp Asp Val Arg Ser Ala Glu Val
      260      265      270
Glu Leu Lys Ile Thr Gly Arg Asp Asn Phe Val Leu Asp Gly Lys Ser
      275      280      285
Asn Met Ser Leu Ile Ser Arg Asp Pro Ala Val Leu Ala Gly Gln Ala
      290      295      300
Tyr Gly Lys Gln His Gln Tyr Pro Asp Gly Phe Ala Leu Phe Leu Gly
305      310      315      320
Thr Met Phe Ala Pro Ile Gln Asp Arg Asp Thr Pro Gly Gln Gly Phe
      325      330      335
Thr His Lys Val Gly Asp Arg Val Arg Val Ser Thr Pro Lys Leu Gly
      340      345      350
Val Leu Glu Asn Glu Val Thr Thr Cys Asp Lys Ala Lys Pro Trp Thr
      355      360      365
Phe Gly Ile Ser Ala Leu Ile Arg Asn Leu Ala Gly Arg Gly Leu Leu
      370      375      380

```

```

<210> 25
<211> 478
<212> PRT
<213> Caulobacter crescentus

```

```

<400> 25
Met Thr Asp Thr Leu Arg His Tyr Ile Gly Gly Glu Arg Val Ala Ala
1      5      10      15
Asp Ala Pro Ala Glu Ser Leu Asn Pro Ser Asn Thr Asn Asp Val Val
      20      25      30
Ala Lys Val Pro Met Gly Gly Gln Ala Glu Val Asp Ala Ala Val Asp
      35      40      45
Ala Ala Arg Lys Ala Phe Pro Ala Trp Ala Asp Ala Ser Pro Glu Val
      50      55      60
Arg Ser Asp Leu Leu Asp Lys Val Gly Ser Thr Ile Ile Ala Arg Ser
65      70      75      80
Ala Asp Ile Gly Arg Leu Leu Ala Arg Glu Glu Gly Lys Thr Leu Ala
      85      90      95
Glu Gly Ile Gly Glu Thr Val Arg Ala Gly Arg Ile Phe Lys Tyr Phe
      100      105      110
Ala Gly Glu Ala Leu Arg Arg His Gly Gln Asn Leu Glu Ser Thr Arg
      115      120      125
Pro Gly Val Glu Ile Gln Thr Tyr Arg Gln Ala Val Gly Val Tyr Gly
      130      135      140
Leu Ile Thr Pro Trp Asn Phe Pro Ile Ala Ile Pro Ala Trp Lys Ala
145      150      155      160
Ala Pro Ala Leu Ala Phe Gly Asn Thr Val Val Ile Lys Pro Ala Gly
      165      170      175
Pro Thr Pro Ala Thr Ala Asn Val Leu Ala Asp Ile Met Ala Glu Cys
      180      185      190
Gly Ala Pro Ala Gly Val Phe Asn Met Leu Phe Gly Arg Gly Ser Met
      195      200      205
Gly Asp Ala Leu Ile Lys His Lys Asp Val Asp Gly Val Ser Phe Thr
      210      215      220
Gly Ser Gln Gly Val Gly Ala Gln Val Ala Ala Ala Val Ala Arg

```

225 230 235 240
 Gln Ala Arg Val Gln Leu Glu Met Gly Gly Lys Asn Pro Leu Ile Val
 245 250 255
 Leu Asp Asp Ala Asp Leu Glu Arg Ala Val Ala Ile Ala Leu Asp Gly
 260 265 270
 Ser Phe Phe Ala Thr Gly Gln Arg Cys Thr Ala Ser Ser Arg Leu Ile
 275 280 285
 Val Gln Asp Gly Ile His Asp Lys Phe Val Ala Leu Leu Ala Glu Lys
 290 295 300
 Val Ala Ala Leu Arg Val Gly Asp Ala Leu Asp Pro Asn Thr Gln Ile
 305 310 315 320
 Gly Pro Ala Val Ser Glu Asp Gln Met Glu Thr Ser Tyr Arg Tyr Ile
 325 330 335
 Asp Ile Ala Ala Ser Glu Gly Gly Arg Val Val Thr Gly Gly Asp Arg
 340 345 350
 Ile Lys Leu Asp Asn Pro Gly Trp Tyr Val Arg Pro Thr Leu Ile Ala
 355 360 365
 Asp Thr Gln Ala Gly Met Arg Ile Asn Asn Glu Glu Val Phe Gly Pro
 370 375 380
 Val Ala Ser Thr Ile Arg Val Lys Ser Tyr Glu Glu Ala Leu Glu Ile
 385 390 395 400
 Ala Asn Gly Val Glu Phe Gly Leu Ser Ala Gly Ile Ala Thr Thr Ser
 405 410 415
 Leu Lys His Ala Arg His Phe Gln Arg Tyr Ala Arg Ala Gly Met Thr
 420 425 430
 Met Val Asn Leu Ala Thr Ala Gly Val Asp Tyr His Val Pro Phe Gly
 435 440 445
 Gly Thr Lys Ser Ser Ser Tyr Gly Ala Arg Glu Gln Gly Phe Ala Ala
 450 455 460
 Val Glu Phe Phe Thr Gln Thr Lys Thr Ser Tyr Ser Trp Ser
 465 470 475

<210> 26
 <211> 248
 <212> PRT
 <213> *Caulobacter crescentus*

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

Ser Trp His Leu Gly Leu Glu Asp Leu Val Leu Tyr Glu Thr Ala Lys
 145 150 155 160
 Ala Gly Ile Glu Gly Met Thr Arg Ala Leu Ala Arg Glu Leu Gly Pro
 165 170 175
 Asp Asp Ile Arg Val Thr Cys Val Val Pro Gly Asn Val Lys Thr Lys
 180 185 190
 Arg Gln Glu Lys Trp Tyr Thr Pro Glu Gly Glu Ala Gln Ile Val Ala
 195 200 205
 Ala Gln Cys Leu Lys Gly Arg Ile Val Pro Glu Asn Val Ala Ala Leu
 210 215 220
 Val Leu Phe Leu Ala Ser Asp Asp Ala Ser Leu Cys Thr Gly His Glu
 225 230 235 240
 Tyr Trp Ile Asp Ala Gly Trp Arg
 245

<210> 27
 <211> 289
 <212> PRT
 <213> *Caulobacter crescentus*

<400> 27
 Met Thr Ala Gln Val Thr Cys Val Trp Asp Leu Lys Ala Thr Leu Gly
 1 5 10 15
 Glu Gly Pro Ile Trp His Gly Asp Thr Leu Trp Phe Val Asp Ile Lys
 20 25 30
 Gln Arg Lys Ile His Asn Tyr His Pro Ala Thr Gly Glu Arg Phe Ser
 35 40 45
 Phe Asp Ala Pro Asp Gln Val Thr Phe Leu Ala Pro Ile Val Gly Ala
 50 55 60
 Thr Gly Phe Val Val Gly Leu Lys Thr Gly Ile His Arg Phe His Pro
 65 70 75 80
 Ala Thr Gly Phe Ser Leu Leu Leu Glu Val Glu Asp Ala Ala Leu Asn
 85 90 95
 Asn Arg Pro Asn Asp Ala Thr Val Asp Ala Gln Gly Arg Leu Trp Phe
 100 105 110
 Gly Thr Met His Asp Gly Glu Glu Asn Asn Ser Gly Ser Leu Tyr Arg
 115 120 125
 Met Asp Leu Thr Gly Val Ala Arg Met Asp Arg Asp Ile Cys Ile Thr
 130 135 140
 Asn Gly Pro Cys Val Ser Pro Asp Gly Lys Thr Phe Tyr His Thr Asp
 145 150 155 160
 Thr Leu Glu Lys Thr Ile Tyr Ala Phe Asp Leu Ala Glu Asp Gly Leu
 165 170 175
 Leu Ser Asn Lys Arg Val Phe Val Gln Phe Ala Leu Gly Asp Asp Val
 180 185 190
 Tyr Pro Asp Gly Ser Val Val Asp Ser Glu Gly Tyr Leu Trp Thr Ala
 195 200 205
 Leu Trp Gly Gly Phe Gly Ala Val Arg Phe Ser Pro Gln Gly Asp Ala
 210 215 220
 Val Thr Arg Ile Glu Leu Pro Ala Pro Asn Val Thr Lys Pro Cys Phe
 225 230 235 240
 Gly Gly Pro Asp Leu Lys Thr Leu Tyr Phe Thr Thr Ala Arg Lys Gly
 245 250 255
 Leu Ser Asp Glu Thr Leu Ala Gln Tyr Pro Leu Ala Gly Gly Val Phe
 260 265 270
 Ala Val Pro Val Asp Val Ala Gly Gln Pro Gln His Glu Val Arg Leu

275 280 285
 Val

 <210> 28
 <211> 1788
 <212> ADN
 <213> *Caulobacter crescentus*

 <220>
 <221> CDS
 <222> (1)..(1788)

 <400> 28
 atg agg tcc gcc ttg tct aac cgc acg ccc cgc cgg ttc cgg tcc cgc 48
 Met Arg Ser Ala Leu Ser Asn Arg Thr Pro Arg Arg Phe Arg Ser Arg
 1 5 10 15
 gat tgg ttc gat aac ccc gac cat atc gac atg acc gcg ctc tat ctg 96
 Asp Trp Phe Asp Asn Pro Asp His Ile Asp Met Thr Ala Leu Tyr Leu
 20 25 30
 gag cgc ttc atg aac tac ggg atc acg ccg gag gag ctg cgc agc ggc 144
 Glu Arg Phe Met Asn Tyr Gly Ile Thr Pro Glu Glu Leu Arg Ser Gly
 35 40 45
 aag ccg atc atc ggc atc gcc cag acc ggc agc gac atc tcg ccc tgc 192
 Lys Pro Ile Ile Gly Ile Ala Gln Thr Gly Ser Asp Ile Ser Pro Cys
 50 55 60
 aac cgc atc cac ctg gac ctg gtc cag cgg gtg cgg gac ggg atc cgc 240
 Asn Arg Ile His Leu Asp Leu Val Gln Arg Val Arg Asp Gly Ile Arg
 65 70 75 80
 gac gcc ggg ggc atc ccc atg gag ttc ccg gtc cat ccg atc ttc gag 288
 Asp Ala Gly Gly Ile Pro Met Glu Phe Pro Val His Pro Ile Phe Glu
 85 90 95
 aac tgc cgt cgc ccg acg gcg gcg ctg gac cgg aac ctc tcg tac ctg 336
 Asn Cys Arg Arg Pro Thr Ala Ala Leu Asp Arg Asn Leu Ser Tyr Leu
 100 105 110
 ggt ctc gtc gag acc ctg cac ggc tat ccg atc gac gcc gtg gtt ctg 384
 Gly Leu Val Glu Thr Leu His Gly Tyr Pro Ile Asp Ala Val Val Leu
 115 120 125
 acc acc ggc tgc gac aag acc acc ccg gcc ggg atc atg gcc gcc acc 432
 Thr Thr Gly Cys Asp Lys Thr Thr Pro Ala Gly Ile Met Ala Ala Thr
 130 135 140
 acg gtc aat atc ccg gcc atc gtg ctg tcg ggc ggc ccg atg ctg gac 480
 Thr Val Asn Ile Pro Ala Ile Val Leu Ser Gly Gly Pro Met Leu Asp
 145 150 155 160
 ggc tgg cac gag aac gag ctc gtg ggc tcg ggc acc gtg atc tgg cgc 528
 Gly Trp His Glu Asn Glu Leu Val Gly Ser Gly Thr Val Ile Trp Arg
 165 170 175
 tcg cgc cgc aag ctg gcg gcc ggc gag atc acc gag gaa gag ttc atc 576
 Ser Arg Arg Lys Leu Ala Ala Gly Glu Ile Thr Glu Glu Glu Phe Ile
 180 185 190
 gac cgc gcc gcc agc tcg gcg ccg tcg gcg ggc cac tgc aac acc atg 624
 Asp Arg Ala Ala Ser Ser Ala Pro Ser Ala Gly His Cys Asn Thr Met
 195 200 205
 ggc acg gcc tcg acc atg aac gcc gtg gcc gag gcg ctg ggc ctg tcg 672
 Gly Thr Ala Ser Thr Met Asn Ala Val Ala Glu Ala Leu Gly Leu Ser
 210 215 220
 ctg acc ggc tgc gcg gcc atc ccc gcc ccc tac cgc gag cgc ggc cag 720

Leu Thr Gly Cys Ala Ala Ile Pro Ala Pro Tyr Arg Glu Arg Gly Gln	
225 230 235 240	
atg gcc tac aag acc ggc cag cgc atc gtc gat ctg gcc tat gac gac	768
Met Ala Tyr Lys Thr Gly Gln Arg Ile Val Asp Leu Ala Tyr Asp Asp	
245 250 255	
gtc aaa ccg ctc gac atc ctg acc aag caa gcc ttc gag aac gcc atc	816
Val Lys Pro Leu Asp Ile Leu Thr Lys Gln Ala Phe Glu Asn Ala Ile	
260 265 270	
gcc ctg gtg gcg gcg gcc ggc ggc tcg acc aac gcc cag ccg cac atc	864
Ala Leu Val Ala Ala Ala Gly Gly Ser Thr Asn Ala Gln Pro His Ile	
275 280 285	
gtg gcc atg gcc cgt cac gcc ggc gtc gag atc acc gcc gac gac tgg	912
Val Ala Met Ala Arg His Ala Gly Val Glu Ile Thr Ala Asp Asp Trp	
290 295 300	
cgc gcg gcc tat gac atc ccg ctg atc gtc aac atg cag ccg gcc ggc	960
Arg Ala Ala Tyr Asp Ile Pro Leu Ile Val Asn Met Gln Pro Ala Gly	
305 310 315 320	
aag tat ctg ggc gag cgc ttc cac cga gcc ggc ggc gcg ccg gcg gtg	1008
Lys Tyr Leu Gly Glu Arg Phe His Arg Ala Gly Gly Ala Pro Ala Val	
325 330 335	
ctg tgg gag ctg ttg cag caa ggc cgc ctg cac ggc gac gtg ctg acc	1056
Leu Trp Glu Leu Leu Gln Gln Gly Arg Leu His Gly Asp Val Leu Thr	
340 345 350	
gtc acc ggc aag acg atg agc gag aac ctg caa ggc cgc gaa acc agc	1104
Val Thr Gly Lys Thr Met Ser Glu Asn Leu Gln Gly Arg Glu Thr Ser	
355 360 365	
gac cgc gag gtg atc ttc ccg tac cac gag ccg ctg gcc gag aag gcc	1152
Asp Arg Glu Val Ile Phe Pro Tyr His Glu Pro Leu Ala Glu Lys Ala	
370 375 380	
ggg ttc ctg gtt ctc aag ggc aac ctc ttc gac ttc gcg atc atg aag	1200
Gly Phe Leu Val Leu Lys Gly Asn Leu Phe Asp Phe Ala Ile Met Lys	
385 390 395 400	
tcc agc gtg atc ggc gag gag ttc cgc aag cgc tac ctg tcg cag ccc	1248
Ser Ser Val Ile Gly Glu Glu Phe Arg Lys Arg Tyr Leu Ser Gln Pro	
405 410 415	
ggc cag gaa ggc gtg ttc gaa gcc cgc gcc atc gtg ttc gac ggc tcg	1296
Gly Gln Glu Gly Val Phe Glu Ala Arg Ala Ile Val Phe Asp Gly Ser	
420 425 430	
gac gac tat cac aag cgg atc aac gat ccg gcc ctg gag atc gac gag	1344
Asp Asp Tyr His Lys Arg Ile Asn Asp Pro Ala Leu Glu Ile Asp Glu	
435 440 445	
cgc tgc atc ctg gtg atc cgc ggc gcg ggt ccg atc ggc tgg ccc ggc	1392
Arg Cys Ile Leu Val Ile Arg Gly Ala Gly Pro Ile Gly Trp Pro Gly	
450 455 460	
tcg gcc gag gtc gtc aac atg cag ccg ccg gat cac ctt ctg aag aag	1440
Ser Ala Glu Val Val Asn Met Gln Pro Pro His Leu Leu Lys Lys	
465 470 475 480	
ggg atc atg agc ctg ccc acc ctg ggc gat ggc cgt cag tcg ggc acc	1488
Gly Ile Met Ser Leu Pro Thr Leu Gly Asp Gly Arg Gln Ser Gly Thr	
485 490 495	
gcc gac agc ccc tcg atc ctg aac gcc tcg ccc gaa agc gcg atc ggc	1536
Ala Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Ile Gly	
500 505 510	
ggc ggc ctg tcg tgg ctg cgc acc ggc gac acc atc cgc atc gac ctc	1584
Gly Gly Leu Ser Trp Leu Arg Thr Gly Asp Thr Ile Arg Ile Asp Leu	
515 520 525	
aac acc ggc cgc tgc gac gcc ctg gtc gac gag gcg acg atc gcc gcg	1632

```

Asn Thr Gly Arg Cys Asp Ala Leu Val Asp Glu Ala Thr Ile Ala Ala
530                               535                               540
cgc aag cag gac ggc atc ccg gcg gtt ccc gcc acc atg acg ccc tgg      1680
Arg Lys Gln Asp Gly Ile Pro Ala Val Pro Ala Thr Met Thr Pro Trp
545                               550                               555                               560
cag gaa atc tac cgc gcc cac gcc agt cag ctc gac acc ggc ggc gtg      1728
Gln Glu Ile Tyr Arg Ala His Ala Ser Gln Leu Asp Thr Gly Gly Val
                               565                               570                               575
ctg gag ttc gcg gtc aag tac cag gac ctg gcg gcc aag ctg ccc cgc      1776
Leu Glu Phe Ala Val Lys Tyr Gln Asp Leu Ala Ala Lys Leu Pro Arg
                               580                               585                               590
cac aac cac tga      1788
His Asn His
                               595

```

```

<210> 29
<211> 595
<212> PRT
<213> Caulobacter crescentus

```

```

<400> 29
Met Arg Ser Ala Leu Ser Asn Arg Thr Pro Arg Arg Phe Arg Ser Arg
1                               5                               10                               15
Asp Trp Phe Asp Asn Pro Asp His Ile Asp Met Thr Ala Leu Tyr Leu
20                               25                               30
Glu Arg Phe Met Asn Tyr Gly Ile Thr Pro Glu Glu Leu Arg Ser Gly
35                               40                               45
Lys Pro Ile Ile Gly Ile Ala Gln Thr Gly Ser Asp Ile Ser Pro Cys
50                               55                               60
Asn Arg Ile His Leu Asp Leu Val Gln Arg Val Arg Asp Gly Ile Arg
65                               70                               75                               80
Asp Ala Gly Gly Ile Pro Met Glu Phe Pro Val His Pro Ile Phe Glu
85                               90                               95
Asn Cys Arg Arg Pro Thr Ala Ala Leu Asp Arg Asn Leu Ser Tyr Leu
100                               105                               110
Gly Leu Val Glu Thr Leu His Gly Tyr Pro Ile Asp Ala Val Val Leu
115                               120                               125
Thr Thr Gly Cys Asp Lys Thr Thr Pro Ala Gly Ile Met Ala Ala Thr
130                               135                               140
Thr Val Asn Ile Pro Ala Ile Val Leu Ser Gly Gly Pro Met Leu Asp
145                               150                               155                               160
Gly Trp His Glu Asn Glu Leu Val Gly Ser Gly Thr Val Ile Trp Arg
165                               170                               175
Ser Arg Arg Lys Leu Ala Ala Gly Glu Ile Thr Glu Glu Glu Phe Ile
180                               185                               190
Asp Arg Ala Ala Ser Ser Ala Pro Ser Ala Gly His Cys Asn Thr Met
195                               200                               205
Gly Thr Ala Ser Thr Met Asn Ala Val Ala Glu Ala Leu Gly Leu Ser
210                               215                               220
Leu Thr Gly Cys Ala Ala Ile Pro Ala Pro Tyr Arg Glu Arg Gly Gln
225                               230                               235                               240
Met Ala Tyr Lys Thr Gly Gln Arg Ile Val Asp Leu Ala Tyr Asp Asp
245                               250                               255
Val Lys Pro Leu Asp Ile Leu Thr Lys Gln Ala Phe Glu Asn Ala Ile
260                               265                               270
Ala Leu Val Ala Ala Ala Gly Gly Ser Thr Asn Ala Gln Pro His Ile

```

275 280 285
 Val Ala Met Ala Arg His Ala Gly Val Glu Ile Thr Ala Asp Asp Trp
 290 295 300
 Arg Ala Ala Tyr Asp Ile Pro Leu Ile Val Asn Met Gln Pro Ala Gly
 305 310 315 320
 Lys Tyr Leu Gly Glu Arg Phe His Arg Ala Gly Gly Ala Pro Ala Val
 325 330 335
 Leu Trp Glu Leu Leu Gln Gln Gly Arg Leu His Gly Asp Val Leu Thr
 340 345 350
 Val Thr Gly Lys Thr Met Ser Glu Asn Leu Gln Gly Arg Glu Thr Ser
 355 360 365
 Asp Arg Glu Val Ile Phe Pro Tyr His Glu Pro Leu Ala Glu Lys Ala
 370 375 380
 Gly Phe Leu Val Leu Lys Gly Asn Leu Phe Asp Phe Ala Ile Met Lys
 385 390 395 400
 Ser Ser Val Ile Gly Glu Glu Phe Arg Lys Arg Tyr Leu Ser Gln Pro
 405 410 415
 Gly Gln Glu Gly Val Phe Glu Ala Arg Ala Ile Val Phe Asp Gly Ser
 420 425 430
 Asp Asp Tyr His Lys Arg Ile Asn Asp Pro Ala Leu Glu Ile Asp Glu
 435 440 445
 Arg Cys Ile Leu Val Ile Arg Gly Ala Gly Pro Ile Gly Trp Pro Gly
 450 455 460
 Ser Ala Glu Val Val Asn Met Gln Pro Pro Asp His Leu Leu Lys Lys
 465 470 475 480
 Gly Ile Met Ser Leu Pro Thr Leu Gly Asp Gly Arg Gln Ser Gly Thr
 485 490 495
 Ala Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Ile Gly
 500 505 510
 Gly Gly Leu Ser Trp Leu Arg Thr Gly Asp Thr Ile Arg Ile Asp Leu
 515 520 525
 Asn Thr Gly Arg Cys Asp Ala Leu Val Asp Glu Ala Thr Ile Ala Ala
 530 535 540
 Arg Lys Gln Asp Gly Ile Pro Ala Val Pro Ala Thr Met Thr Pro Trp
 545 550 555 560
 Gln Glu Ile Tyr Arg Ala His Ala Ser Gln Leu Asp Thr Gly Gly Val
 565 570 575
 Leu Glu Phe Ala Val Lys Tyr Gln Asp Leu Ala Ala Lys Leu Pro Arg
 580 585 590
 His Asn His
 595

<210> 30
 <211> 1476
 <212> ADN
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)..(1476)

<400> 30
 atg aat acc cag tat aat tcc agt tat ata ttt tcg att acc tta gtc 48
 Met Asn Thr Gln Tyr Asn Ser Ser Tyr Ile Phe Ser Ile Thr Leu Val
 1 5 10 15
 gct aca tta ggt ggt tta tta ttt ggc tac gac acc gcc gtt att tcc 96

Ala	Thr	Leu	Gly	Gly	Leu	Leu	Phe	Gly	Tyr	Asp	Thr	Ala	Val	Ile	Ser	
			20					25					30			
ggt	act	gtt	gag	tca	ctc	aat	acc	gtc	ttt	gtt	gct	cca	caa	aac	tta	144
Gly	Thr	Val	Glu	Ser	Leu	Asn	Thr	Val	Phe	Val	Ala	Pro	Gln	Asn	Leu	
		35					40					45				
agt	gaa	tcc	gct	gcc	aac	tcc	ctg	tta	ggg	ttt	tgc	gtg	gcc	agc	gct	192
Ser	Glu	Ser	Ala	Ala	Asn	Ser	Leu	Leu	Gly	Phe	Cys	Val	Ala	Ser	Ala	
	50					55					60					
ctg	att	ggg	tgc	atc	atc	ggc	ggg	gcc	ctc	ggg	ggg	tat	tgc	agt	aac	240
Leu	Ile	Gly	Cys	Ile	Ile	Gly	Gly	Ala	Leu	Gly	Gly	Tyr	Cys	Ser	Asn	
65					70					75					80	
cgc	ttc	ggg	cgt	cgt	gat	tca	ctt	aag	att	gct	gct	gtc	ctg	ttt	ttt	288
Arg	Phe	Gly	Arg	Arg	Asp	Ser	Leu	Lys	Ile	Ala	Ala	Val	Leu	Phe	Phe	
				85					90					95		
att	tct	ggg	gta	ggg	tct	gcc	tgg	cca	gaa	ctt	ggg	ttt	acc	tct	ata	336
Ile	Ser	Gly	Val	Gly	Ser	Ala	Trp	Pro	Glu	Leu	Gly	Phe	Thr	Ser	Ile	
		100					105					110				
aac	cgg	gac	aac	act	gtg	cct	gtt	tat	ctg	gca	ggg	tat	gtc	cgg	gaa	384
Asn	Pro	Asp	Asn	Thr	Val	Pro	Val	Tyr	Leu	Ala	Gly	Tyr	Val	Pro	Glu	
	115						120					125				
ttt	gtt	att	tat	cgc	att	att	ggc	ggg	att	ggc	gtt	ggg	tta	gcc	tca	432
Phe	Val	Ile	Tyr	Arg	Ile	Ile	Gly	Gly	Ile	Gly	Val	Gly	Leu	Ala	Ser	
	130					135					140					
atg	ctc	tgc	cca	atg	tat	att	gcg	gaa	ctg	gct	cca	gct	cat	att	cgc	480
Met	Leu	Ser	Pro	Met	Tyr	Ile	Ala	Glu	Leu	Ala	Pro	Ala	His	Ile	Arg	
145					150					155					160	
ggg	aaa	ctg	gtc	tct	ttt	aac	cag	ttt	gog	att	att	ttc	ggg	caa	ctt	528
Gly	Lys	Leu	Val	Ser	Phe	Asn	Gln	Phe	Ala	Ile	Ile	Phe	Gly	Gln	Leu	
				165				170						175		
tta	gtt	tac	tgc	gta	aac	tat	ttt	att	gcc	cgt	tcc	ggg	gat	gcc	agc	576
Leu	Val	Tyr	Cys	Val	Asn	Tyr	Phe	Ile	Ala	Arg	Ser	Gly	Asp	Ala	Ser	
			180				185					190				
tgg	ctg	aat	act	gac	ggc	tgg	cgt	tat	atg	ttt	gcc	tgc	gaa	tgt	atc	624
Trp	Leu	Asn	Thr	Asp	Gly	Trp	Arg	Tyr	Met	Phe	Ala	Ser	Glu	Cys	Ile	
		195					200					205				
cct	gca	ctg	ctg	ttc	tta	atg	ctg	ctg	tat	acc	gtg	cca	gaa	agt	cct	672
Pro	Ala	Leu	Leu	Phe	Leu	Met	Leu	Leu	Tyr	Thr	Val	Pro	Glu	Ser	Pro	
	210					215					220					
cgc	tgg	ctg	atg	tgc	cgc	ggc	aag	caa	gaa	cag	gcg	gaa	ggg	atc	ctg	720
Arg	Trp	Leu	Met	Ser	Arg	Gly	Lys	Gln	Glu	Gln	Ala	Glu	Gly	Ile	Leu	
225					230					235					240	
cgc	aaa	att	atg	ggc	aac	acg	ctt	gca	act	cag	gca	gta	cag	gaa	att	768
Arg	Lys	Ile	Met	Gly	Asn	Thr	Leu	Ala	Thr	Gln	Ala	Val	Gln	Glu	Ile	
				245					250					255		
aaa	cac	tcc	ctg	gat	cat	ggc	cgc	aaa	acc	ggg	ggg	cgt	ctg	ctg	atg	816
Lys	His	Ser	Leu	Asp	His	Gly	Arg	Lys	Thr	Gly	Gly	Arg	Leu	Leu	Met	
			260					265					270			
ttt	ggc	gtg	ggc	gtg	att	gta	atc	ggc	gta	atg	ctc	tcc	atc	ttc	cag	864
Phe	Gly	Val	Gly	Val	Ile	Val	Ile	Gly	Val	Met	Leu	Ser	Ile	Phe	Gln	
		275					280					285				
caa	ttt	gtc	ggc	atc	aat	gtg	gtg	ctg	tac	tac	gcg	cgg	gaa	gtg	ttc	912
Gln	Phe	Val	Gly	Ile	Asn	Val	Val	Leu	Tyr	Tyr	Ala	Pro	Glu	Val	Phe	
	290					295					300					
aaa	acg	ctg	ggg	gcc	agc	acg	gat	atc	gcg	ctg	ttg	cag	acc	att	att	960
Lys	Thr	Leu	Gly	Ala	Ser	Thr	Asp	Ile	Ala	Leu	Leu	Gln	Thr	Ile	Ile	
305					310					315					320	
gtc	gga	gtt	atc	aac	ctc	acc	ttc	acc	gtt	ctg	gca	att	atg	acg	gtg	1008

Val	Gly	Val	Ile	Asn	Leu	Thr	Phe	Thr	Val	Leu	Ala	Ile	Met	Thr	Val		
				325					330					335			
gat	aaa	ttt	ggt	cgt	aag	cca	ctg	caa	att	atc	ggc	gca	ctc	gga	atg	1056	
Asp	Lys	Phe	Gly	Arg	Lys	Pro	Leu	Gln	Ile	Ile	Gly	Ala	Leu	Gly	Met		
			340					345					350				
gca	atc	ggt	atg	ttt	agc	ctc	ggt	acc	gcg	ttt	tac	act	cag	gca	ccg	1104	
Ala	Ile	Gly	Met	Phe	Ser	Leu	Gly	Thr	Ala	Phe	Tyr	Thr	Gln	Ala	Pro		
		355					360					365					
ggt	att	gtg	gcg	cta	ctg	tcg	atg	ctg	ttc	tat	gtt	gcc	gcc	ttt	gcc	1152	
Gly	Ile	Val	Ala	Leu	Leu	Ser	Met	Leu	Phe	Tyr	Val	Ala	Ala	Phe	Ala		
	370					375					380						
atg	tcc	tgg	ggt	ccg	gta	tgc	tgg	gta	ctg	ctg	tcg	gaa	atc	ttc	ccg	1200	
Met	Ser	Trp	Gly	Pro	Val	Cys	Trp	Val	Leu	Leu	Ser	Glu	Ile	Phe	Pro		
	385				390				395					400			
aat	gct	att	cgt	ggt	aaa	gcg	ctg	gca	atc	gcg	gtg	gcg	gcc	cag	tgg	1248	
Asn	Ala	Ile	Arg	Gly	Lys	Ala	Leu	Ala	Ile	Ala	Val	Ala	Ala	Gln	Trp		
			405					410						415			
ctg	gcg	aac	tac	ttc	gtc	tcc	tgg	acc	ttc	ccg	atg	atg	gac	aaa	aac	1296	
Leu	Ala	Asn	Tyr	Phe	Val	Ser	Trp	Thr	Phe	Pro	Met	Met	Asp	Lys	Asn		
		420					425				430						
tcc	tgg	ctg	gtg	gcc	cat	ttc	cac	aac	ggt	ttc	tcc	tac	tgg	att	tac	1344	
Ser	Trp	Leu	Val	Ala	His	Phe	His	Asn	Gly	Phe	Ser	Tyr	Trp	Ile	Tyr		
	435					440					445						
ggt	tgt	atg	ggc	gtt	ctg	gca	gca	ctg	ttt	atg	tgg	aaa	ttt	gtc	ccg	1392	
Gly	Cys	Met	Gly	Val	Leu	Ala	Ala	Leu	Phe	Met	Trp	Lys	Phe	Val	Pro		
	450					455				460							
gaa	acc	aaa	ggt	aaa	acc	ctt	gag	gag	ctg	gaa	gcg	ctc	tgg	gaa	ccg	1440	
Glu	Thr	Lys	Gly	Lys	Thr	Leu	Glu	Glu	Leu	Glu	Ala	Leu	Trp	Glu	Pro		
	465				470				475					480			
gaa	acg	aag	aaa	aca	caa	caa	act	gct	acg	ctg	taa					1476	
Glu	Thr	Lys	Lys	Thr	Gln	Gln	Thr	Ala	Thr	Leu							
			485					490									

<210> 31
 <211> 491
 <212> PRT
 <213> Escherichia coli

<400> 31

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

130	135	140
Met Leu Ser Pro Met Tyr Ile Ala Glu Leu Ala Pro Ala His Ile Arg		
145	150	155
Gly Lys Leu Val Ser Phe Asn Gln Phe Ala Ile Ile Phe Gly Gln Leu		160
	165	170
Leu Val Tyr Cys Val Asn Tyr Phe Ile Ala Arg Ser Gly Asp Ala Ser		175
	180	185
Trp Leu Asn Thr Asp Gly Trp Arg Tyr Met Phe Ala Ser Glu Cys Ile		190
	195	200
Pro Ala Leu Leu Phe Leu Met Leu Leu Tyr Thr Val Pro Glu Ser Pro		205
	210	215
Arg Trp Leu Met Ser Arg Gly Lys Gln Glu Gln Ala Glu Gly Ile Leu		220
225	230	235
Arg Lys Ile Met Gly Asn Thr Leu Ala Thr Gln Ala Val Gln Glu Ile		240
	245	250
Lys His Ser Leu Asp His Gly Arg Lys Thr Gly Gly Arg Leu Leu Met		255
	260	265
Phe Gly Val Gly Val Ile Val Ile Gly Val Met Leu Ser Ile Phe Gln		270
	275	280
Gln Phe Val Gly Ile Asn Val Val Leu Tyr Tyr Ala Pro Glu Val Phe		285
	290	295
Lys Thr Leu Gly Ala Ser Thr Asp Ile Ala Leu Leu Gln Thr Ile Ile		300
305	310	315
Val Gly Val Ile Asn Leu Thr Phe Thr Val Leu Ala Ile Met Thr Val		320
	325	330
Asp Lys Phe Gly Arg Lys Pro Leu Gln Ile Ile Gly Ala Leu Gly Met		335
	340	345
Ala Ile Gly Met Phe Ser Leu Gly Thr Ala Phe Tyr Thr Gln Ala Pro		350
	355	360
Gly Ile Val Ala Leu Leu Ser Met Leu Phe Tyr Val Ala Ala Phe Ala		365
	370	375
Met Ser Trp Gly Pro Val Cys Trp Val Leu Leu Ser Glu Ile Phe Pro		380
385	390	395
Asn Ala Ile Arg Gly Lys Ala Leu Ala Ile Ala Val Ala Ala Gln Trp		400
	405	410
Leu Ala Asn Tyr Phe Val Ser Trp Thr Phe Pro Met Met Asp Lys Asn		415
	420	425
Ser Trp Leu Val Ala His Phe His Asn Gly Phe Ser Tyr Trp Ile Tyr		430
	435	440
Gly Cys Met Gly Val Leu Ala Ala Leu Phe Met Trp Lys Phe Val Pro		445
	450	455
Glu Thr Lys Gly Lys Thr Leu Glu Glu Leu Glu Ala Leu Trp Glu Pro		460
465	470	475
Glu Thr Lys Lys Thr Gln Gln Thr Ala Thr Leu		480
	485	490

<210> 32
 <211> 399
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp PtacTtrp

<400> 32
 ggtaccagat ctccctgttg acaattaatc atcggtctcta taatgtgtgg aatcgtgagc

```

ggataacaat ttcacacaag gagactcccg ggagccgcca gttccgctgg cggcatttta 120
actttcttta atgaagccgg aaaaatccta aattcattta atatttatct ttttaccggt 180
tcgcttaacc cggtcgaacg tcaacttacg tcatTTTTcc gcccaacagt aatataatca 240
aacaatttaa tcccgcaaca taacaccagt aaaatcaata attttctcta agtcacttat 300
tcctcaggta attgttaata tatccagaat gttcctcaaa atatatTTtc cctctatctt 360
ctcgttgcgc ttaatttgac taatttcat tagggatcc 399

```

<210> 33
 <211> 362
 <212> ADN
 <213> *Brevibacterium flavum*

```

<400> 33
agatcgTTta gatcgaagg aaaacgtcga aaagcaattt gcttttcgac gccccacccc 60
gcgcgtTTta gcgtgtcagt aggcgcgtag ggtaagtggg gtagcggctt gttagatata 120
ttgaaatcgg ctttcaacag cattgatttc gatgtattta gctggccggt accctgcgaa 180
tgtccacagg gtagctggta gtttgaaaat caacgccggt gcccttagga ttcagtaact 240
ggcacatttt gtaatgcgct agatctgtgt gctcagtcct ccaggctgct tatcacagt 300
aaagcaaaac caattcgtgg ctgcgaaagt cgtagccacc acgaagtcca ggaggacata 360
ca 362

```

<210> 34
 <211> 1968
 <212> ADN
 <213> *Escherichia coli*

<220>
 <221> CDS
 <222> (1)..(1968)

```

<400> 34
atg tct gtt cgc aat att ttt gct gac gag agc cac gat att tac acc 48
Met Ser Val Arg Asn Ile Phe Ala Asp Glu Ser His Asp Ile Tyr Thr
1 5 10 15
gtc aga acg cac gcc gat ggc ccg gac ggc gaa ctc cca tta acc gca 96
Val Arg Thr His Ala Asp Gly Pro Asp Gly Glu Leu Pro Leu Thr Ala
20 25 30
gag atg ctt atc aac cgc ccg agc ggg gat ctg ttc ggt atg acc atg 144
Glu Met Leu Ile Asn Arg Pro Ser Gly Asp Leu Phe Gly Met Thr Met
35 40 45
aat gcc gga atg ggt tgg tct ccg gac gag ctg gat cgg gac ggt att 192
Asn Ala Gly Met Gly Trp Ser Pro Asp Glu Leu Asp Arg Asp Gly Ile
50 55 60
tta ctg ctc agt aca ctc ggt ggc tta cgc ggc gca gac ggt aaa ccc 240
Leu Leu Leu Ser Thr Leu Gly Gly Leu Arg Gly Ala Asp Gly Lys Pro
65 70 75 80
gtg gcg ctg gcg ttg cac cag ggg cat tac gaa ctg gac atc cag atg 288
Val Ala Leu Ala Leu His Gln Gly His Tyr Glu Leu Asp Ile Gln Met
85 90 95
aaa gcg gcg gcc gag gtt att aaa gcc aac cat gcc ctg ccc tat gcc 336
Lys Ala Ala Ala Glu Val Ile Lys Ala Asn His Ala Leu Pro Tyr Ala
100 105 110
gtg tac gtc tcc gat cct tgt gac ggg cgt act cag ggt aca acg ggg 384
Val Tyr Val Ser Asp Pro Cys Asp Gly Arg Thr Gln Gly Thr Thr Gly
115 120 125

```

atg ttt gat tgc cta cca tac cga aat gac gca tgc atg gta atg cgc	432
Met Phe Asp Ser Leu Pro Tyr Arg Asn Asp Ala Ser Met Val Met Arg	
130 135 140	
cgc ctt att cgc tct ctg ccc gac gcg aaa gca gtt att ggt gtg gcg	480
Arg Leu Ile Arg Ser Leu Pro Asp Ala Lys Ala Val Ile Gly Val Ala	
145 150 155 160	
agt tgc gat aag ggg ctt ccg gcc acc atg atg gca ctc gcc gcg cag	528
Ser Cys Asp Lys Gly Leu Pro Ala Thr Met Met Ala Leu Ala Ala Gln	
165 170 175	
cac aac atc gca acc gtg ctg gtc ccc ggc ggc gcg acg ctg ccc gca	576
His Asn Ile Ala Thr Val Leu Val Pro Gly Gly Ala Thr Leu Pro Ala	
180 185 190	
aag gat gga gaa gac aac ggc aag gtg caa acc att ggc gca cgc ttc	624
Lys Asp Gly Glu Asp Asn Gly Lys Val Gln Thr Ile Gly Ala Arg Phe	
195 200 205	
gcc aat ggc gaa tta tct cta cag gac gca cgc cgt gcg ggc tgt aaa	672
Ala Asn Gly Glu Leu Ser Leu Gln Asp Ala Arg Arg Ala Gly Cys Lys	
210 215 220	
gcc tgt gcc tct tcc ggc ggc ggc tgt caa ttt ttg ggc act gcc ggg	720
Ala Cys Ala Ser Ser Gly Gly Gly Cys Gln Phe Leu Gly Thr Ala Gly	
225 230 235 240	
aca tct cag gtg gtg gcc gaa gga ttg gga ctg gca atc cca cat tca	768
Thr Ser Gln Val Val Ala Glu Gly Leu Gly Leu Ala Ile Pro His Ser	
245 250 255	
gcc ctg gcc cct tcc ggt gag cct gtg tgg cgg gag atc gcc aga gct	816
Ala Leu Ala Pro Ser Gly Glu Pro Val Trp Arg Glu Ile Ala Arg Ala	
260 265 270	
tcc gcg cga gct gcg ctg aac ctg agt caa aaa ggc atc acc acc cgg	864
Ser Ala Arg Ala Ala Leu Asn Leu Ser Gln Lys Gly Ile Thr Thr Arg	
275 280 285	
gaa att ctc acc gat aaa gcg ata gag aat gcg atg acg gtc cat gcc	912
Glu Ile Leu Thr Asp Lys Ala Ile Glu Asn Ala Met Thr Val His Ala	
290 295 300	
gcg ttc ggt ggt tca aca aac ctg ctg tta cac atc ccg gca att gct	960
Ala Phe Gly Gly Ser Thr Asn Leu Leu Leu His Ile Pro Ala Ile Ala	
305 310 315 320	
cac cag gca ggt tgc cat atc ccg acc gtt gat gac tgg atc cgc atc	1008
His Gln Ala Gly Cys His Ile Pro Thr Val Asp Asp Trp Ile Arg Ile	
325 330 335	
aac aag cgc gtg ccc cga ctg gtg agc gta ctg cct aat ggc ccg gtt	1056
Asn Lys Arg Val Pro Arg Leu Val Ser Val Leu Pro Asn Gly Pro Val	
340 345 350	
tat cat cca acg gtc aat gcc ttt atg gca ggt ggt gtg ccg gaa gtc	1104
Tyr His Pro Thr Val Asn Ala Phe Met Ala Gly Gly Val Pro Glu Val	
355 360 365	
atg ttg cat ctg cgc agc ctc gga ttg ttg cat gaa gac gtt atg acg	1152
Met Leu His Leu Arg Ser Leu Gly Leu Leu His Glu Asp Val Met Thr	
370 375 380	
gtt acc ggc agc acg ctg aaa gaa aac ctc gac tgg tgg gag cac tcc	1200
Val Thr Gly Ser Thr Leu Lys Glu Asn Leu Asp Trp Trp Glu His Ser	
385 390 395 400	
gaa cgg cgt cag cgg ttc aag caa ctc ctg ctc gat cag gaa caa atc	1248
Glu Arg Arg Gln Arg Phe Lys Gln Leu Leu Leu Asp Gln Glu Gln Ile	
405 410 415	
aac gct gac gaa gtg atc atg tct ccg cag caa gca aaa gcg cgc gga	1296
Asn Ala Asp Glu Val Ile Met Ser Pro Gln Gln Ala Lys Ala Arg Gly	
420 425 430	

```

tta acc tca act atc acc ttc ccg gtg ggc aat att gcg cca gaa ggt      1344
Leu Thr Ser Thr Ile Thr Phe Pro Val Gly Asn Ile Ala Pro Glu Gly
      435      440      445
tcg gtg atc aaa tcc acc gcc att gac ccc tcg atg att gat gag caa      1392
Ser Val Ile Lys Ser Thr Ala Ile Asp Pro Ser Met Ile Asp Glu Gln
      450      455      460
ggg atc tat tac cat aaa ggt gtg gcg aag gtt tat ctg tcc gag aaa      1440
Gly Ile Tyr Tyr His Lys Gly Val Ala Lys Val Tyr Leu Ser Glu Lys
      465      470      475      480
agt gcg att tac gat atc aaa cat gac aag atc aag gcg ggc gat att      1488
Ser Ala Ile Tyr Asp Ile Lys His Asp Lys Ile Lys Ala Gly Asp Ile
      485      490      495
ctg gtc att att ggc gtt gga cct tca ggt aca ggg atg gaa gaa acc      1536
Leu Val Ile Ile Gly Val Gly Pro Ser Gly Thr Gly Met Glu Glu Thr
      500      505      510
tac cag gtt acc agt gcc ctg aag cat ctg tca tac ggt aag cat gtt      1584
Tyr Gln Val Thr Ser Ala Leu Lys His Leu Ser Tyr Gly Lys His Val
      515      520      525
tcg tta atc acc gat gca cgt ttc tcg ggc gtt tct act ggc gcg tgc      1632
Ser Leu Ile Thr Asp Ala Arg Phe Ser Gly Val Ser Thr Gly Ala Cys
      530      535      540
atc ggc cat gtg ggg cca gaa gcg ctg gcc gga ggc ccc atc ggt aaa      1680
Ile Gly His Val Gly Pro Glu Ala Leu Ala Gly Gly Pro Ile Gly Lys
      545      550      555      560
tta cgc acc ggg gat tta att gaa att aaa att gat tgt cgc gag ctt      1728
Leu Arg Thr Gly Asp Leu Ile Glu Ile Lys Ile Asp Cys Arg Glu Leu
      565      570      575
cac ggc gaa gtc aat ttc ctc gga acc cgt agc gat gaa caa tta cct      1776
His Gly Glu Val Asn Phe Leu Gly Thr Arg Ser Asp Glu Gln Leu Pro
      580      585      590
tca cag gag gag gca act gca ata tta aat gcc aga ccc agc cat cag      1824
Ser Gln Glu Glu Ala Thr Ala Ile Leu Asn Ala Arg Pro Ser His Gln
      595      600      605
gat tta ctt ccc gat cct gaa ttg cca gat gat acc cgg cta tgg gca      1872
Asp Leu Leu Pro Asp Pro Glu Leu Pro Asp Asp Thr Arg Leu Trp Ala
      610      615      620
atg ctt cag gcc gtg agt ggt ggg aca tgg acc ggt tgt att tat gat      1920
Met Leu Gln Ala Val Ser Gly Gly Thr Trp Thr Gly Cys Ile Tyr Asp
      625      630      635      640
gta aac aaa att ggc gcg gct ttg cgc gat ttt atg aat aaa aac tga      1968
Val Asn Lys Ile Gly Ala Ala Leu Arg Asp Phe Met Asn Lys Asn
      645      650      655

```

```

<210> 35
<211> 655
<212> PRT
<213> Escherichia coli

```

```

<400> 35
Met Ser Val Arg Asn Ile Phe Ala Asp Glu Ser His Asp Ile Tyr Thr
1      5      10      15
Val Arg Thr His Ala Asp Gly Pro Asp Gly Glu Leu Pro Leu Thr Ala
20      25      30
Glu Met Leu Ile Asn Arg Pro Ser Gly Asp Leu Phe Gly Met Thr Met
35      40      45
Asn Ala Gly Met Gly Trp Ser Pro Asp Glu Leu Asp Arg Asp Gly Ile

```

50	Leu 65	Leu	Leu	Ser	Thr	Leu 70	Gly	Gly	Leu	Arg	Gly 75	Ala	Asp	Gly	Lys	Pro 80
	Val	Ala	Leu	Ala	Leu 85	His	Gln	Gly	His	Tyr 90	Glu	Leu	Asp	Ile	Gln 95	Met
	Lys	Ala	Ala	Ala 100	Glu	Val	Ile	Lys	Ala 105	Asn	His	Ala	Leu	Pro 110	Tyr	Ala
	Val	Tyr	Val	Ser 115	Asp	Pro	Cys	Asp	Gly 120	Arg	Thr	Gln	Gly 125	Thr	Thr	Gly
	Met	Phe 130	Asp	Ser	Leu	Pro	Tyr 135	Arg	Asn	Asp	Ala	Ser 140	Met	Val	Met	Arg
	Arg 145	Leu	Ile	Arg	Ser	Leu 150	Pro	Asp	Ala	Lys	Ala 155	Val	Ile	Gly	Val	Ala 160
	Ser	Cys	Asp	Lys 165	Gly	Leu	Pro	Ala	Thr	Met 170	Met	Ala	Leu	Ala 175	Ala	Gln
	His	Asn	Ile	Ala 180	Thr	Val	Leu	Val	Pro 185	Gly	Gly	Ala	Thr	Leu 190	Pro	Ala
	Lys	Asp	Gly 195	Glu	Asp	Asn	Gly	Lys 200	Val	Gln	Thr	Ile	Gly 205	Ala	Arg	Phe
	Ala	Asn	Gly 210	Glu	Leu	Ser	Leu 215	Gln	Asp	Ala	Arg 220	Arg	Ala	Gly	Cys	Lys
	Ala 225	Cys	Ala	Ser	Ser	Gly 230	Gly	Gly	Cys	Gln	Phe 235	Leu	Gly	Thr	Ala	Gly 240
	Thr	Ser	Gln	Val 245	Val	Ala	Glu	Gly	Leu	Gly 250	Leu	Ala	Ile	Pro	His	Ser
	Ala	Leu	Ala 260	Pro	Ser	Gly	Glu	Pro	Val 265	Trp	Arg	Glu	Ile	Ala 270	Arg	Ala
	Ser	Ala	Arg 275	Ala	Ala	Leu	Asn	Leu 280	Ser	Gln	Lys	Gly	Ile	Thr	Thr	Arg
	Glu	Ile 290	Leu	Thr	Asp	Lys	Ala 295	Ile	Glu	Asn	Ala 300	Met	Thr	Val	His	Ala
	Ala 305	Phe	Gly	Gly	Ser	Thr 310	Asn	Leu	Leu	Leu	His 315	Ile	Pro	Ala	Ile	Ala 320
	His	Gln	Ala	Gly 325	Cys	His	Ile	Pro	Thr	Val 330	Asp	Asp	Trp	Ile	Arg	Ile
	Asn	Lys	Arg	Val 340	Pro	Arg	Leu	Val	Ser 345	Val	Leu	Pro	Asn	Gly 350	Pro	Val
	Tyr	His	Pro 355	Thr	Val	Asn	Ala	Phe 360	Met	Ala	Gly	Gly	Val	Pro	Glu	Val
	Met	Leu 370	His	Leu	Arg	Ser	Leu 375	Gly	Leu	Leu	His 380	Glu	Asp	Val	Met	Thr
	Val 385	Thr	Gly	Ser	Thr	Leu 390	Lys	Glu	Asn	Leu	Asp 395	Trp	Trp	Glu	His	Ser
	Glu	Arg	Arg	Gln 405	Arg	Phe	Lys	Gln	Leu	Leu 410	Leu	Asp	Gln	Glu	Gln	Ile
	Asn	Ala	Asp	Glu 420	Val	Ile	Met	Ser	Pro 425	Gln	Gln	Ala	Lys	Ala 430	Arg	Gly
	Leu	Thr	Ser 435	Thr	Ile	Thr	Phe	Pro	Val 440	Gly	Asn	Ile	Ala 445	Pro	Glu	Gly
	Ser	Val	Ile 450	Lys	Ser	Thr	Ala 455	Ile	Asp	Pro	Ser	Met 460	Ile	Asp	Glu	Gln
	Gly 465	Ile	Tyr	Tyr	His	Lys 470	Gly	Val	Ala	Lys	Val 475	Tyr	Leu	Ser	Glu	Lys
	Ser	Ala	Ile	Tyr 485	Asp	Ile	Lys	His	Asp	Lys 490	Ile	Lys	Ala	Gly	Asp	Ile
	Leu	Val	Ile	Ile 500	Gly	Val	Gly	Pro	Ser 505	Gly	Thr	Gly	Met	Glu	Glu	Thr

Tyr Gln Val Thr Ser Ala Leu Lys His Leu Ser Tyr Gly Lys His Val
 515 520 525
 Ser Leu Ile Thr Asp Ala Arg Phe Ser Gly Val Ser Thr Gly Ala Cys
 530 535 540
 Ile Gly His Val Gly Pro Glu Ala Leu Ala Gly Gly Pro Ile Gly Lys
 545 550 555 560
 Leu Arg Thr Gly Asp Leu Ile Glu Ile Lys Ile Asp Cys Arg Glu Leu
 565 570 575
 His Gly Glu Val Asn Phe Leu Gly Thr Arg Ser Asp Glu Gln Leu Pro
 580 585 590
 Ser Gln Glu Glu Ala Thr Ala Ile Leu Asn Ala Arg Pro Ser His Gln
 595 600 605
 Asp Leu Leu Pro Asp Pro Glu Leu Pro Asp Asp Thr Arg Leu Trp Ala
 610 615 620
 Met Leu Gln Ala Val Ser Gly Gly Thr Trp Thr Gly Cys Ile Tyr Asp
 625 630 635 640
 Val Asn Lys Ile Gly Ala Ala Leu Arg Asp Phe Met Asn Lys Asn
 645 650 655

<210> 36
 <211> 1968
 <212> ADN
 <213> Escherichia coli

<220>
 <221> CDS
 <222> (1)..(1968)

<400> 36
 atg acc att gag aaa att ttc acc ccg cag gac gac gcg ttt tat gcg 48
 Met Thr Ile Glu Lys Ile Phe Thr Pro Gln Asp Asp Ala Phe Tyr Ala
 1 5 10 15
 gtg atc acc cac gcg gcg ggg ccg cag ggc gct ctg ccg ctg acc ccg 96
 Val Ile Thr His Ala Ala Gly Pro Gln Gly Ala Leu Pro Leu Thr Pro
 20 25 30
 cag atg ctg atg gaa tct ccc agc ggc aac ctg ttc ggc atg acg cag 144
 Gln Met Leu Met Glu Ser Pro Ser Gly Asn Leu Phe Gly Met Thr Gln
 35 40 45
 aac gcc ggg atg ggc tgg gac gcc aac aag ctc acc ggc aaa gag gtg 192
 Asn Ala Gly Met Gly Trp Asp Ala Asn Lys Leu Thr Gly Lys Glu Val
 50 55 60
 ctg att atc ggc act cag ggc ggc atc cgc gcc gga gac gga cgc cca 240
 Leu Ile Ile Gly Thr Gln Gly Gly Ile Arg Ala Gly Asp Gly Arg Pro
 65 70 75 80
 atc gcg ctg ggc tac cac acc ggg cat tgg gag atc ggc atg cag atg 288
 Ile Ala Leu Gly Tyr His Thr Gly His Trp Glu Ile Gly Met Gln Met
 85 90 95
 cag gcg gcg gcg aag gag atc acc cgc aat ggc ggc atc ccg ttc gcg 336
 Gln Ala Ala Ala Lys Glu Ile Thr Arg Asn Gly Gly Ile Pro Phe Ala
 100 105 110
 gcc ttc gtc agc gat ccg tgc gac ggg cgc tcg cag ggc acg cac ggt 384
 Ala Phe Val Ser Asp Pro Cys Asp Gly Arg Ser Gln Gly Thr His Gly
 115 120 125
 atg ttc gat tcc ctg ccg tac cgc aac gac gcg gcg atc gtg ttt cgc 432
 Met Phe Asp Ser Leu Pro Tyr Arg Asn Asp Ala Ile Val Phe Arg
 130 135 140

cgc ctg atc cgc tcc ctg ccg acg cgg cgg gcg gtg atc ggc gta gcg	480
Arg Leu Ile Arg Ser Leu Pro Thr Arg Arg Ala Val Ile Gly Val Ala	
145 150 155 160	
acc tgc gat aaa ggg ctg ccc gcc acc atg att gcg ctg gcc gcg atg	528
Thr Cys Asp Lys Gly Leu Pro Ala Thr Met Ile Ala Leu Ala Ala Met	
165 170 175	
cac gac ctg ccg act att ctg gtg ccg ggc ggg gcg acg ctg ccg ccg	576
His Asp Leu Pro Thr Ile Leu Val Pro Gly Gly Ala Thr Leu Pro Pro	
180 185 190	
acc gtc ggg gaa gac gcg ggc aag gtg cag acc atc ggc gcg cgt ttc	624
Thr Val Gly Glu Asp Ala Gly Lys Val Gln Thr Ile Gly Ala Arg Phe	
195 200 205	
gcc aac cac gaa ctc tcc ctg cag gag gcc gcc gaa ctg ggc tgt cgc	672
Ala Asn His Glu Leu Ser Leu Gln Glu Ala Ala Glu Leu Gly Cys Arg	
210 215 220	
gcc tgc gcc tcg ccg ggc ggc ggg tgt cag ttc ctc ggc acg gcg ggc	720
Ala Cys Ala Ser Pro Gly Gly Gly Cys Gln Phe Leu Gly Thr Ala Gly	
225 230 235 240	
acc tcg cag gtg gtc gcg gag gcg ctg ggt ctg gcg ctg ccg cac tcc	768
Thr Ser Gln Val Val Ala Glu Ala Leu Gly Leu Ala Leu Pro His Ser	
245 250 255	
gcg ctg gcg ccg tcc ggg cag gcg gtg tgg ctg gag atc gcc cgc cag	816
Ala Leu Ala Pro Ser Gly Gln Ala Val Trp Leu Glu Ile Ala Arg Gln	
260 265 270	
tcg gcg cgc gcg gtc agc gag ctg gat agc cgc ggc atc acc acg ccg	864
Ser Ala Arg Ala Val Ser Glu Leu Asp Ser Arg Gly Ile Thr Thr Arg	
275 280 285	
gat atc ctc tcc gat aaa gcc atc gaa aac gcg atg gtg atc cac gcg	912
Asp Ile Leu Ser Asp Lys Ala Ile Glu Asn Ala Met Val Ile His Ala	
290 295 300	
gcg ttc ggc ggc tcc acc aat tta ctg ctg cac att ccg gcc atc gcc	960
Ala Phe Gly Gly Ser Thr Asn Leu Leu Leu His Ile Pro Ala Ile Ala	
305 310 315 320	
cac gcg gcg ggc tgc acg atc ccg gac gtt gag cac tgg acg cgc atc	1008
His Ala Ala Gly Cys Thr Ile Pro Asp Val Glu His Trp Thr Arg Ile	
325 330 335	
aac cgt aaa gtg ccg cgt ctg gtg agc gtg ctg ccc aac ggc ccg gac	1056
Asn Arg Lys Val Pro Arg Leu Val Ser Val Leu Pro Asn Gly Pro Asp	
340 345 350	
tat cac ccg acc gtg cgc gcc ttc ctc gcg ggc ggc gtg ccg gag gtg	1104
Tyr His Pro Thr Val Arg Ala Phe Leu Ala Gly Gly Val Pro Glu Val	
355 360 365	
atg ctc cac ctg cgc gac ctc ggc ctg ctg cat ctg gac gcc atg acc	1152
Met Leu His Leu Arg Asp Leu Gly Leu Leu His Leu Asp Ala Met Thr	
370 375 380	
gtg acc ggc cag acg gtg ggc gag aac ctt gaa tgg tgg cag gcg tcc	1200
Val Thr Gly Gln Thr Val Gly Glu Asn Leu Glu Trp Trp Gln Ala Ser	
385 390 395 400	
gag cgc ccg gcg cgc ttc cgc cag tgc ctg cgc gag cag gac ggc gta	1248
Glu Arg Arg Ala Arg Phe Arg Gln Cys Leu Arg Glu Gln Asp Gly Val	
405 410 415	
gag ccg gat gac gtg atc ctg ccg ccg gag aag gca aaa gcg aaa ggg	1296
Glu Pro Asp Asp Val Ile Leu Pro Pro Glu Lys Ala Lys Ala Lys Gly	
420 425 430	
ctg acc tcg acg gtc tgc ttc ccg acg ggc aac atc gct ccg gaa ggt	1344
Leu Thr Ser Thr Val Cys Phe Pro Thr Gly Asn Ile Ala Pro Glu Gly	
435 440 445	

tcg gtg atc aag gcc acg gcg atc gac ccg tcg gtg gtg ggc gaa gat	1392
Ser Val Ile Lys Ala Thr Ala Ile Asp Pro Ser Val Val Gly Glu Asp	
450 455 460	
ggc gta tac cac cac acc ggc cgg gtg cgg gtg ttt gtc tcg gaa gcg	1440
Gly Val Tyr His His Thr Gly Arg Val Arg Val Phe Val Ser Glu Ala	
465 470 475 480	
cag gcg atc aag gcg atc aag cgg gaa gag att gtg cag ggc gat atc	1488
Gln Ala Ile Lys Ala Ile Lys Arg Glu Glu Ile Val Gln Gly Asp Ile	
485 490 495	
atg gtg gtg atc ggc ggc ggg ccg tcc ggc acc ggc atg gaa gag acc	1536
Met Val Val Ile Gly Gly Gly Pro Ser Gly Thr Gly Met Glu Glu Thr	
500 505 510	
tac cag ctc acc tcc gcg cta aag cat atc tcg tgg ggc aag acg gtg	1584
Tyr Gln Leu Thr Ser Ala Leu Lys His Ile Ser Trp Gly Lys Thr Val	
515 520 525	
tcg ctc atc acc gat gcg cgc ttc tcg ggc gtg tcg acg ggc gcc tgc	1632
Ser Leu Ile Thr Asp Ala Arg Phe Ser Gly Val Ser Thr Gly Ala Cys	
530 535 540	
ttc ggc cac gtg tcg ccg gag gcg ctg gcg ggc ggg ccg att ggc aag	1680
Phe Gly His Val Ser Pro Glu Ala Leu Ala Gly Gly Pro Ile Gly Lys	
545 550 555 560	
ctg cgc gat aac gac atc atc gag att gcc gtg gat cgt ctg acg tta	1728
Leu Arg Asp Asn Asp Ile Ile Glu Ile Ala Val Asp Arg Leu Thr Leu	
565 570 575	
act ggc agc gtg aac ttc atc ggc acc gcg gac aac ccg ctg acg ccg	1776
Thr Gly Ser Val Asn Phe Ile Gly Thr Ala Asp Asn Pro Leu Thr Pro	
580 585 590	
gaa gag ggc gcg cgc gag ctg gcg cgg cgg cag acg cac ccg gac ctg	1824
Glu Glu Gly Ala Arg Glu Leu Ala Arg Arg Gln Thr His Pro Asp Leu	
595 600 605	
cac gcc cac gac ttt ttg ccg gac gac acc cgg ctg tgg gcg gca ctg	1872
His Ala His Asp Phe Leu Pro Asp Asp Thr Arg Leu Trp Ala Ala Leu	
610 615 620	
cag tcg gtg agc ggc ggc acc tgg aaa ggc tgt att tat gac acc gat	1920
Gln Ser Val Ser Gly Gly Thr Trp Lys Gly Cys Ile Tyr Asp Thr Asp	
625 630 635 640	
aaa att atc gag gta att aac gcc ggt aaa aaa gcg ctc gga att taa	1968
Lys Ile Ile Glu Val Ile Asn Ala Gly Lys Lys Ala Leu Gly Ile	
645 650 655	

<210> 37
 <211> 655
 <212> PRT
 <213> Escherichia coli

<400> 37
 Met Thr Ile Glu Lys Ile Phe Thr Pro Gln Asp Asp Ala Phe Tyr Ala
 1 5 10 15
 Val Ile Thr His Ala Ala Gly Pro Gln Gly Ala Leu Pro Leu Thr Pro
 20 25 30
 Gln Met Leu Met Glu Ser Pro Ser Gly Asn Leu Phe Gly Met Thr Gln
 35 40 45
 Asn Ala Gly Met Gly Trp Asp Ala Asn Lys Leu Thr Gly Lys Glu Val
 50 55 60
 Leu Ile Ile Gly Thr Gln Gly Gly Ile Arg Ala Gly Asp Gly Arg Pro
 65 70 75 80

Ile	Ala	Leu	Gly	Tyr	His	Thr	Gly	His	Trp	Glu	Ile	Gly	Met	Gln	Met	85	90	95
Gln	Ala	Ala	Ala	Lys	Glu	Ile	Thr	Arg	Asn	Gly	Gly	Ile	Pro	Phe	Ala	100	105	110
Ala	Phe	Val	Ser	Asp	Pro	Cys	Asp	Gly	Arg	Ser	Gln	Gly	Thr	His	Gly	115	120	125
Met	Phe	Asp	Ser	Leu	Pro	Tyr	Arg	Asn	Asp	Ala	Ala	Ile	Val	Phe	Arg	130	135	140
Arg	Leu	Ile	Arg	Ser	Leu	Pro	Thr	Arg	Arg	Ala	Val	Ile	Gly	Val	Ala	145	150	155
Thr	Cys	Asp	Lys	Gly	Leu	Pro	Ala	Thr	Met	Ile	Ala	Leu	Ala	Ala	Met	165	170	175
His	Asp	Leu	Pro	Thr	Ile	Leu	Val	Pro	Gly	Gly	Ala	Thr	Leu	Pro	Pro	180	185	190
Thr	Val	Gly	Glu	Asp	Ala	Gly	Lys	Val	Gln	Thr	Ile	Gly	Ala	Arg	Phe	195	200	205
Ala	Asn	His	Glu	Leu	Ser	Leu	Gln	Glu	Ala	Ala	Glu	Leu	Gly	Cys	Arg	210	215	220
Ala	Cys	Ala	Ser	Pro	Gly	Gly	Gly	Cys	Gln	Phe	Leu	Gly	Thr	Ala	Gly	225	230	235
Thr	Ser	Gln	Val	Val	Ala	Glu	Ala	Leu	Gly	Leu	Ala	Leu	Pro	His	Ser	245	250	255
Ala	Leu	Ala	Pro	Ser	Gly	Gln	Ala	Val	Trp	Leu	Glu	Ile	Ala	Arg	Gln	260	265	270
Ser	Ala	Arg	Ala	Val	Ser	Glu	Leu	Asp	Ser	Arg	Gly	Ile	Thr	Thr	Arg	275	280	285
Asp	Ile	Leu	Ser	Asp	Lys	Ala	Ile	Glu	Asn	Ala	Met	Val	Ile	His	Ala	290	295	300
Ala	Phe	Gly	Gly	Ser	Thr	Asn	Leu	Leu	Leu	His	Ile	Pro	Ala	Ile	Ala	305	310	315
His	Ala	Ala	Gly	Cys	Thr	Ile	Pro	Asp	Val	Glu	His	Trp	Thr	Arg	Ile	325	330	335
Asn	Arg	Lys	Val	Pro	Arg	Leu	Val	Ser	Val	Leu	Pro	Asn	Gly	Pro	Asp	340	345	350
Tyr	His	Pro	Thr	Val	Arg	Ala	Phe	Leu	Ala	Gly	Gly	Val	Pro	Glu	Val	355	360	365
Met	Leu	His	Leu	Arg	Asp	Leu	Gly	Leu	Leu	His	Leu	Asp	Ala	Met	Thr	370	375	380
Val	Thr	Gly	Gln	Thr	Val	Gly	Glu	Asn	Leu	Glu	Trp	Trp	Gln	Ala	Ser	385	390	395
Glu	Arg	Arg	Ala	Arg	Phe	Arg	Gln	Cys	Leu	Arg	Glu	Gln	Asp	Gly	Val	405	410	415
Glu	Pro	Asp	Asp	Val	Ile	Leu	Pro	Pro	Glu	Lys	Ala	Lys	Ala	Lys	Gly	420	425	430
Leu	Thr	Ser	Thr	Val	Cys	Phe	Pro	Thr	Gly	Asn	Ile	Ala	Pro	Glu	Gly	435	440	445
Ser	Val	Ile	Lys	Ala	Thr	Ala	Ile	Asp	Pro	Ser	Val	Val	Gly	Glu	Asp	450	455	460
Gly	Val	Tyr	His	His	Thr	Gly	Arg	Val	Arg	Val	Phe	Val	Ser	Glu	Ala	465	470	475
Gln	Ala	Ile	Lys	Ala	Ile	Lys	Arg	Glu	Glu	Ile	Val	Gln	Gly	Asp	Ile	485	490	495
Met	Val	Val	Ile	Gly	Gly	Gly	Pro	Ser	Gly	Thr	Gly	Met	Glu	Glu	Thr	500	505	510
Tyr	Gln	Leu	Thr	Ser	Ala	Leu	Lys	His	Ile	Ser	Trp	Gly	Lys	Thr	Val	515	520	525
Ser	Leu	Ile	Thr	Asp	Ala	Arg	Phe	Ser	Gly	Val	Ser	Thr	Gly	Ala	Cys			

530					535					540					
Phe	Gly	His	Val	Ser	Pro	Glu	Ala	Leu	Ala	Gly	Gly	Pro	Ile	Gly	Lys
545					550					555					560
Leu	Arg	Asp	Asn	Asp	Ile	Ile	Glu	Ile	Ala	Val	Asp	Arg	Leu	Thr	Leu
				565					570					575	
Thr	Gly	Ser	Val	Asn	Phe	Ile	Gly	Thr	Ala	Asp	Asn	Pro	Leu	Thr	Pro
				580					585					590	
Glu	Glu	Gly	Ala	Arg	Glu	Leu	Ala	Arg	Arg	Gln	Thr	His	Pro	Asp	Leu
				595					600					605	
His	Ala	His	Asp	Phe	Leu	Pro	Asp	Asp	Thr	Arg	Leu	Trp	Ala	Ala	Leu
				610					615					620	
Gln	Ser	Val	Ser	Gly	Gly	Thr	Trp	Lys	Gly	Cys	Ile	Tyr	Asp	Thr	Asp
625					630					635					640
Lys	Ile	Ile	Glu	Val	Ile	Asn	Ala	Gly	Lys	Lys	Ala	Leu	Gly	Ile	
				645					650					655	

```
<210> 38
<211> 40
<212> ADN
<213> Trình tự nhân tạo
```

<220>
<223> đoạn mỗi xylX IFS 5742-10-1

<400> 38
acacaaggag actcccatgg gcgtgagtga attcctgccg

```
<210> 39
<211> 40
<212> ADN
<213> Trình tự nhân tạo
```

<220>
<223> đoạn mỗi xylX IFA 5742-10-2

<400> 39
gqaactggcg gctcccttag aggaggccgc ggccggccag

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

<220>
<223> đoạn mỗi xylA IFS 5742-10-3

```
<400> 40
acacaaggag actcccatga cgcacaccct gcgccattac
```

<210>	41
<211>	40
<212>	ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi xylA_IFA_5742-10-4

<400> 41

ggaactggcg gctcccttac gaccacgagt aggaggtttt

40

<210> 42

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi xylD_IFS_5742-10-5

<400> 42

acacaaggag actcccatgt ctaaccgcac gccccgccgg

40

<210> 43

<211> 40

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi xylD-IFA_5742-10-6

<400> 43

ggaactggcg gctccctcag tggttgtggc ggggcagctt

40

<210> 44

<211> 44

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0823-01F_4691-87-1

<400> 44

gtcgactcta gaggatcccc gtgggcgtga gtgaattcct gccg

44

<210> 45

<211> 24

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0823-01R_4691-87-2

<400> 45

tatgcctgtc ctgccagcac tgcc

24

<210> 46
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0823-02F_4691-87-3

<400> 46
 ggcagtgctg gcaggacagg cata

24

<210> 47
 <211> 44
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0823-02R_4691-87-4

<400> 47
 cgaattcgag ctcggtaccc ttagaggagg ccgcggccgg ccag

44

<210> 48
 <211> 44
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-01F_4691-87-5

<400> 48
 gtgcactcta gaggatcccc atgaccgaca ccctgcgcca ttac

44

<210> 49
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-01R_5659-8-7

<400> 49
 gcgtcgcccc aggccgggaa tgcc

24

<210> 50
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-02F_5659-8-8

<400> 50

gtcgacgccg cgcgcaaggc attc 24

<210> 51
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-02R_5659-8-9

<400> 51
 gagcgccggg gcggccttcc atgc 24

<210> 52
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-03F_5659-8-10

<400> 52
 cttcccgatc gccatcccgg catg 24

<210> 53
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-03R_5659-8-11

<400> 53
 cagctgcacg cgggcctgac gtgc 24

<210> 54
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi CC0822-04F_5659-8-12

<400> 54
 ggtagccgcc gccgccgtgg cacg 24

<210> 55
 <211> 24
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0822-04R_5659-8-13

<400> 55

ggcgaccttc tcggccagca gtgc

24

<210> 56

<211> 24

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0822-05F_5659-8-14

<400> 56

ggattcacga caagttcgtg gcac

24

<210> 57

<211> 44

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi CC0822-05R_4691-87-14

<400> 57

ogaattcgag ctcggtaccc ttacgaccac gagtaggagg tttt

44

<210> 58

<211> 35

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi Peftu(Pst)

<400> 58

ccaagcttgc atgccagatc gtttagatcc gaagg

35

<210> 59

<211> 22

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> đoạn mỗi Peftu_Rv

<400> 59

tgtatgtcct cctggacttc gt

22

<210> 60

<211> 52

<212> ADN

<213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi Pefu_xylXABCD_fw
 <400> 60
 ccaggaggac atacaatggg cgtgagtga ttcctgccgg aagattggaa ag 52
 <210> 61
 <211> 39
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi eftu_xylXABCD_rv
 <400> 61
 cggtaaccgg ggatctcagt ggttgtggcg gggcagctt 39
 <210> 62
 <211> 38
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi PmsrA(Pst)
 <400> 62
 ccaagcttgc atgccatttg cgcctgcaac gtaggttg 38
 <210> 63
 <211> 23
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi PmsrAR
 <400> 63
 aacaggaatg ttccttttoga aaa 23
 <210> 64
 <211> 41
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi PmsrA_xylD_fw
 <400> 64
 aggaacattc ctgttatgtc taaccgcacg ccccgccggt t 41

<210> 65
 <211> 25
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi SpcR-F

 <400> 65
 ctaataacgt aacgtgactg gcaag 25

 <210> 66
 <211> 23
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi SpcR-R

 <400> 66
 ataagtaaga ttaaccatta gtc 23

 <210> 67
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi spc(GTG start)-F

 <400> 67
 gtgaggagga tatatttgaa tacatacgaa 30

 <210> 68
 <211> 35
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi spc(stop)-R

 <400> 68
 taattttttt aatctgttat ttaaataagtt tatag 35

 <210> 69
 <211> 42
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi Spc-pVC7-Cm-F

 <400> 69

atatatcctc ctcacttttag cttccttagc tcctgaaaat ct 42

<210> 70
 <211> 45
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi Spc-pVC7-Cm-R

<400> 70
 agattaaaaa aattataatt tttttaaggc agttattggt gccct 45

<210> 71
 <211> 54
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi ME_Spc_fw

<400> 71
 ctgtctctta tacacatctc caagcttgca tgcgcgggcc tggttggttg ggtt 54

<210> 72
 <211> 58
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn mỗi ME_Peftu_xylXABCD_rv

<400> 72
 ctgtctctta tacacatctc ggtaccggg gatctcagtg gttgtggcgg ggcagctt 58

<210> 73
 <211> 1785
 <212> ADN
 <213> Agrobacterium tumefaciens

<220>
 <221> CDS
 <222> (1)..(1785)

<400> 73
 atg agc gac gac aag aca ccc gtg cgg cgg ctg cga tcc cag gac tgg 48
 Met Ser Asp Asp Lys Thr Pro Val Arg Arg Leu Arg Ser Gln Asp Trp
 1 5 10 15
 ttc gac aac cct gac cat ctc gac atg acg gcg ctc tat ctg gag cgc 96
 Phe Asp Asn Pro Asp His Leu Asp Met Thr Ala Leu Tyr Leu Glu Arg
 20 25 30
 ttc atg aat tat ggc gtg acg ccg gaa gag ctg cgt tcc ggc aag ccg 144
 Phe Met Asn Tyr Gly Val Thr Pro Glu Glu Leu Arg Ser Gly Lys Pro

gtc	atc	ggc	atc	gcg	cag	agc	ggc	agc	gac	ctg	acg	ccc	tgc	aac	cgc	192
Val	Ile	Gly	Ile	Ala	Gln	Ser	Gly	Ser	Asp	Leu	Thr	Pro	Cys	Asn	Arg	
50						55					60					
gtc	cat	gtc	gaa	ctg	gtc	aag	cgg	gtg	cgc	gat	ggg	atc	cgt	gat	gcg	240
Val	His	Val	Glu	Leu	Val	Lys	Arg	Val	Arg	Asp	Gly	Ile	Arg	Asp	Ala	
65						70				75					80	
ggc	ggc	atc	ccc	atc	gag	ttt	ccg	acc	cat	ccg	atg	ttc	gaa	aac	tgc	288
Gly	Gly	Ile	Pro	Ile	Glu	Phe	Pro	Thr	His	Pro	Met	Phe	Glu	Asn	Cys	
				85					90					95		
aag	cga	ccg	aca	gcg	gcg	ctt	gac	cgc	aat	ctt	gcc	tat	ctc	agc	ctg	336
Lys	Arg	Pro	Thr	Ala	Ala	Leu	Asp	Arg	Asn	Leu	Ala	Tyr	Leu	Ser	Leu	
			100					105					110			
gtg	gaa	gtg	ctt	tac	ggg	tat	cca	ctc	gat	ggc	gtc	gtg	ttg	acc	acg	384
Val	Glu	Val	Leu	Tyr	Gly	Tyr	Pro	Leu	Asp	Gly	Val	Val	Leu	Thr	Thr	
		115					120					125				
ggc	tgc	gac	aag	acc	acg	cct	tcg	gcg	ctg	atg	gcg	gca	agc	acg	gtt	432
Gly	Cys	Asp	Lys	Thr	Thr	Pro	Ser	Ala	Leu	Met	Ala	Ala	Ser	Thr	Val	
	130					135				140						
gat	att	ccg	gct	atc	gtc	tta	tcc	ggc	ggg	ccg	atg	ctc	gac	gga	tat	480
Asp	Ile	Pro	Ala	Ile	Val	Leu	Ser	Gly	Gly	Pro	Met	Leu	Asp	Gly	Tyr	
145					150					155					160	
cat	gac	ggc	gat	ctg	gtt	ggg	tcc	ggc	acg	gtc	atc	tgg	cgc	atg	cgc	528
His	Asp	Gly	Asp	Leu	Val	Gly	Ser	Gly	Thr	Val	Ile	Trp	Arg	Met	Arg	
				165					170					175		
cgc	aaa	tac	ggc	gcc	ggc	gaa	atc	acc	cgc	gag	gaa	ttc	ctg	cag	gcg	576
Arg	Lys	Tyr	Gly	Ala	Gly	Glu	Ile	Thr	Arg	Glu	Glu	Phe	Leu	Gln	Ala	
			180					185					190			
gcg	ctc	gaa	tcc	gct	cct	tcg	gtc	ggg	cac	tgc	aac	acc	atg	ggc	acg	624
Ala	Leu	Glu	Ser	Ala	Pro	Ser	Val	Gly	His	Cys	Asn	Thr	Met	Gly	Thr	
		195				200						205				
gcc	tct	acc	atg	aat	gcc	att	gcc	gaa	gcg	ctc	ggc	atg	tcg	ctg	acc	672
Ala	Ser	Thr	Met	Asn	Ala	Ile	Ala	Glu	Ala	Leu	Gly	Met	Ser	Leu	Thr	
		210				215					220					
ggc	tgc	ggc	gca	atc	ccg	gcc	gct	tac	cgc	gaa	cgc	ggc	cag	atg	gcc	720
Gly	Cys	Gly	Ala	Ile	Pro	Ala	Ala	Tyr	Arg	Glu	Arg	Gly	Gln	Met	Ala	
225					230					235					240	
tat	cgc	acc	ggc	cgc	cgc	gcc	gtc	gaa	ctg	gtt	gtc	gag	aat	atc	aag	768
Tyr	Arg	Thr	Gly	Arg	Arg	Ala	Val	Glu	Leu	Val	Val	Glu	Asn	Ile	Lys	
				245					250					255		
cct	tcg	gat	atc	atg	acg	cgc	gag	gcg	ttt	ttg	aac	gcg	atc	cgc	gtc	816
Pro	Ser	Asp	Ile	Met	Thr	Arg	Glu	Ala	Phe	Leu	Asn	Ala	Ile	Arg	Val	
			260					265					270			
aat	tcg	gca	atc	ggc	ggg	tcg	aca	aat	gcc	cag	ccg	cat	ctg	gcg	gcg	864
Asn	Ser	Ala	Ile	Gly	Gly	Ser	Thr	Asn	Ala	Gln	Pro	His	Leu	Ala	Ala	
		275				280						285				
atg	gcg	aaa	cat	gcg	ggc	gtc	gaa	ctt	cgc	gag	gaa	gac	tgg	cag	gtt	912
Met	Ala	Lys	His	Ala	Gly	Val	Glu	Leu	Arg	Glu	Glu	Asp	Trp	Gln	Val	
		290				295					300					
cac	ggg	tat	gac	att	ccg	ctc	atc	gcc	aat	gtc	cag	ccc	gcc	gga	aaa	960
His	Gly	Tyr	Asp	Ile	Pro	Leu	Ile	Ala	Asn	Val	Gln	Pro	Ala	Gly	Lys	
305					310					315					320	
tgg	ctg	ggc	gag	aaa	tat	cac	cgt	gcc	ggc	ggg	acg	ccc	gcc	atc	atg	1008
Trp	Leu	Gly	Glu	Lys	Tyr	His	Arg	Ala	Gly	Gly	Thr	Pro	Ala	Ile	Met	
				325					330					335		
tgg	gag	ctt	ctg	aag	gca	ggc	aag	ctt	gat	ggg	agc	tgt	ccg	acc	gtg	1056
Trp	Glu	Leu	Leu	Lys	Ala	Gly	Lys	Leu	Asp	Gly	Ser	Cys	Pro	Thr	Val	

acc ggc aag acc gtg gcg gaa aat ctg gac ggg cgg gaa tcg acc gac	1104
Thr Gly Lys Thr Val Ala Glu Asn Leu Asp Gly Arg Glu Ser Thr Asp	
355 360 365	
cgc gac gtg atc ctt cct tac gat aag ccg ctc aag gaa cgg gcc ggt	1152
Arg Asp Val Ile Leu Pro Tyr Asp Lys Pro Leu Lys Glu Arg Ala Gly	
370 375 380	
ttc ctc gtc ctc aag ggc aat ctg ttc gat ttc gcc atc atg aag acg	1200
Phe Leu Val Leu Lys Gly Asn Leu Phe Asp Phe Ala Ile Met Lys Thr	
385 390 395 400	
agt gtg atc tcg gcg gaa ttc cgg cag cgt tat ctg agt gag ccg ggt	1248
Ser Val Ile Ser Ala Glu Phe Arg Gln Arg Tyr Leu Ser Glu Pro Gly	
405 410 415	
cga gag ggg att ttc gag ggc aaa tgt gtg gtt ttc gat ggt tcg gaa	1296
Arg Glu Gly Ile Phe Glu Gly Lys Cys Val Val Phe Asp Gly Ser Glu	
420 425 430	
gac tat cac gcc cgt atc aac gat ccg tct ctc gat atc gat gag cgc	1344
Asp Tyr His Ala Arg Ile Asn Asp Pro Ser Leu Asp Ile Asp Glu Arg	
435 440 445	
acc att ctc gtc att cgc ggg gcc gga ccg ctc ggc tgg ccg ggt tcg	1392
Thr Ile Leu Val Ile Arg Gly Ala Gly Pro Leu Gly Trp Pro Gly Ser	
450 455 460	
gct gag gtc gtc aac atg cag ccg ccg gac gcg ctg ttg aaa aaa ggc	1440
Ala Glu Val Val Asn Met Gln Pro Pro Asp Ala Leu Leu Lys Lys Gly	
465 470 475 480	
ata acc agc ctg ccg aca atc ggc gac ggc cgg caa tcg gga acc gcg	1488
Ile Thr Ser Leu Pro Thr Ile Gly Asp Gly Arg Gln Ser Gly Thr Ala	
485 490 495	
gac agc ccg tcg atc ctc aat gcc tcg cca gaa agt gcc gcc ggt gga	1536
Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Ala Gly Gly	
500 505 510	
gga ctg gca tgg ctt cgc acg ggt gat gta atc cgc atc gac ttc aat	1584
Gly Leu Ala Trp Leu Arg Thr Gly Asp Val Ile Arg Ile Asp Phe Asn	
515 520 525	
cag ggc aag tgc gac gcg ttg gta ccg gat gca gag ctt gct gcc cga	1632
Gln Gly Lys Cys Asp Ala Leu Val Pro Asp Ala Glu Leu Ala Ala Arg	
530 535 540	
aag gcc gat ggc att cct gcg gtg cct gcg gat gcc acg ccc tgg cag	1680
Lys Ala Asp Gly Ile Pro Ala Val Pro Ala Asp Ala Thr Pro Trp Gln	
545 550 555 560	
cgc atc tat cgt caa tcg gtg acc cag ctt tcg gac ggt gcg gtt ctg	1728
Arg Ile Tyr Arg Gln Ser Val Thr Gln Leu Ser Asp Gly Ala Val Leu	
565 570 575	
gaa ggg gcc gcg gat ttc cgc cgg att gcc gag aag atg ccc cgg cac	1776
Glu Gly Ala Ala Asp Phe Arg Arg Ile Ala Glu Lys Met Pro Arg His	
580 585 590	
aac cat tga	1785
Asn His	

<210> 74
 <211> 594
 <212> PRT
 <213> Agrobacterium tumefaciens

<400> 74
 Met Ser Asp Asp Lys Thr Pro Val Arg Arg Leu Arg Ser Gln Asp Trp

1				5					10				15		
Phe	Asp	Asn	Pro	Asp	His	Leu	Asp	Met	Thr	Ala	Leu	Tyr	Leu	Glu	Arg
			20					25					30		
Phe	Met	Asn	Tyr	Gly	Val	Thr	Pro	Glu	Glu	Leu	Arg	Ser	Gly	Lys	Pro
		35					40					45			
Val	Ile	Gly	Ile	Ala	Gln	Ser	Gly	Ser	Asp	Leu	Thr	Pro	Cys	Asn	Arg
	50					55					60				
Val	His	Val	Glu	Leu	Val	Lys	Arg	Val	Arg	Asp	Gly	Ile	Arg	Asp	Ala
65					70				75					80	
Gly	Gly	Ile	Pro	Ile	Glu	Phe	Pro	Thr	His	Pro	Met	Phe	Glu	Asn	Cys
				85					90					95	
Lys	Arg	Pro	Thr	Ala	Ala	Leu	Asp	Arg	Asn	Leu	Ala	Tyr	Leu	Ser	Leu
			100					105					110		
Val	Glu	Val	Leu	Tyr	Gly	Tyr	Pro	Leu	Asp	Gly	Val	Val	Leu	Thr	Thr
	115						120						125		
Gly	Cys	Asp	Lys	Thr	Thr	Pro	Ser	Ala	Leu	Met	Ala	Ala	Ser	Thr	Val
	130					135					140				
Asp	Ile	Pro	Ala	Ile	Val	Leu	Ser	Gly	Gly	Pro	Met	Leu	Asp	Gly	Tyr
145					150					155					160
His	Asp	Gly	Asp	Leu	Val	Gly	Ser	Gly	Thr	Val	Ile	Trp	Arg	Met	Arg
			165						170					175	
Arg	Lys	Tyr	Gly	Ala	Gly	Glu	Ile	Thr	Arg	Glu	Glu	Phe	Leu	Gln	Ala
			180					185					190		
Ala	Leu	Glu	Ser	Ala	Pro	Ser	Val	Gly	His	Cys	Asn	Thr	Met	Gly	Thr
	195						200					205			
Ala	Ser	Thr	Met	Asn	Ala	Ile	Ala	Glu	Ala	Leu	Gly	Met	Ser	Leu	Thr
	210				215						220				
Gly	Cys	Gly	Ala	Ile	Pro	Ala	Ala	Tyr	Arg	Glu	Arg	Gly	Gln	Met	Ala
225					230					235					240
Tyr	Arg	Thr	Gly	Arg	Arg	Ala	Val	Glu	Leu	Val	Val	Glu	Asn	Ile	Lys
			245					250						255	
Pro	Ser	Asp	Ile	Met	Thr	Arg	Glu	Ala	Phe	Leu	Asn	Ala	Ile	Arg	Val
			260					265					270		
Asn	Ser	Ala	Ile	Gly	Gly	Ser	Thr	Asn	Ala	Gln	Pro	His	Leu	Ala	Ala
	275						280						285		
Met	Ala	Lys	His	Ala	Gly	Val	Glu	Leu	Arg	Glu	Glu	Asp	Trp	Gln	Val
	290					295				300					
His	Gly	Tyr	Asp	Ile	Pro	Leu	Ile	Ala	Asn	Val	Gln	Pro	Ala	Gly	Lys
305					310					315					320
Trp	Leu	Gly	Glu	Lys	Tyr	His	Arg	Ala	Gly	Gly	Thr	Pro	Ala	Ile	Met
			325						330					335	
Trp	Glu	Leu	Leu	Lys	Ala	Gly	Lys	Leu	Asp	Gly	Ser	Cys	Pro	Thr	Val
			340					345					350		
Thr	Gly	Lys	Thr	Val	Ala	Glu	Asn	Leu	Asp	Gly	Arg	Glu	Ser	Thr	Asp
	355						360					365			
Arg	Asp	Val	Ile	Leu	Pro	Tyr	Asp	Lys	Pro	Leu	Lys	Glu	Arg	Ala	Gly
	370					375					380				
Phe	Leu	Val	Leu	Lys	Gly	Asn	Leu	Phe	Asp	Phe	Ala	Ile	Met	Lys	Thr
385					390					395					400
Ser	Val	Ile	Ser	Ala	Glu	Phe	Arg	Gln	Arg	Tyr	Leu	Ser	Glu	Pro	Gly
			405						410					415	
Arg	Glu	Gly	Ile	Phe	Glu	Gly	Lys	Cys	Val	Val	Phe	Asp	Gly	Ser	Glu
	420						425					430			
Asp	Tyr	His	Ala	Arg	Ile	Asn	Asp	Pro	Ser	Leu	Asp	Ile	Asp	Glu	Arg
	435					440						445			
Thr	Ile	Leu	Val	Ile	Arg	Gly	Ala	Gly	Pro	Leu	Gly	Trp	Pro	Gly	Ser
	450					455					460				

Ala Glu Val Val Asn Met Gln Pro Pro Asp Ala Leu Leu Lys Lys Gly
 465 470 475 480
 Ile Thr Ser Leu Pro Thr Ile Gly Asp Gly Arg Gln Ser Gly Thr Ala
 485 490 495
 Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Ala Gly Gly
 500 505 510
 Gly Leu Ala Trp Leu Arg Thr Gly Asp Val Ile Arg Ile Asp Phe Asn
 515 520 525
 Gln Gly Lys Cys Asp Ala Leu Val Pro Asp Ala Glu Leu Ala Ala Arg
 530 535 540
 Lys Ala Asp Gly Ile Pro Ala Val Pro Ala Asp Ala Thr Pro Trp Gln
 545 550 555 560
 Arg Ile Tyr Arg Gln Ser Val Thr Gln Leu Ser Asp Gly Ala Val Leu
 565 570 575
 Glu Gly Ala Ala Asp Phe Arg Arg Ile Ala Glu Lys Met Pro Arg His
 580 585 590
 Asn His

<210> 75
 <211> 1785
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 75
 atgagcgcgatg ataaaaacccc ggtgcgctgc ctgcgtttctc aggattgggtt tgataaccgcg 60
 gatcatctcgg atatgacggc gctgtatctg gaacgcttta tgaattatgg tgtgaccccg 120
 gaagaactgc gcagcggtaa accggttatt ggcacgcgcc agagcgggtc tgatctgacc 180
 ccgtgcaacc gtgtgcatgt tgaactgggtg aaacgtgttc gcgatgggtat tcgcgatgca 240
 ggcggtattc cgatcgaatt tccgaacgcac ccgatgttcg aaaactgtaa acgtccgacc 300
 gcggccctgg atcgtaatct ggocctatctg agcctgggtg aagtctctgta tggctacccg 360
 ctggatgggtg tggttctgac cacgggctgc gataaaacca ccccgagtgc cctgatggca 420
 gccagcaccg tggatattcc ggcaatcgtt ctgagcggcg gtccgatgct ggatgggtat 480
 catgatggcg atctgggtggg ttctggcacg gttatttggc gtatgcgtcg caaatacggt 540
 gccggcgaaa tcaccgcgca agaatttctg caggcgcgcac tggaatctgc gccgagtgtg 600
 ggccactgca acaccatggg tacggcctct accatgaatg ccattgcaga agccctgggt 660
 atgagtctga ccggttgtgg tgccatcccg gccgcataatc gtgaacgtgg ccagatggcc 720
 tatcgacccg gtgcgcgcgc ggtggaactg gtggttgaaa acattaaacc gagcgatctc 780
 atgacgcgtg aagcatttct gaacgcgatt cgcgttaatt ctgcaatcgg cggtagtacc 840
 aatgcgcagc cgcacatctggc agcgatggcg aaacacgcgc gcgtggaact gcgtgaagaa 900
 gattggcagg ttcattggtta tgatattccg ctgatcgcca atgtgcagcc ggccggtaaa 960
 tggctgggag aaaaatatca ccgtgcgggc ggcaccccg caattatgtg ggaactgctg 1020
 aaagccggta aactggatgg cagttgtccg accgtgacgg gcaaaacggt tgcagaaaac 1080
 ctggatggtc gtgaaagcac cgatcgcgat gttatcctgc cgtatgataa accgctgaaa 1140
 gaacgtgcgg gctttctggt gctgaaagggt aatctgtttg atttcgcgat tatgaaaacc 1200
 agcgttatct ctgccgaatt tcgtcagcgt tatctgtctg aaccgggtcg tgaaggcatt 1260
 tttgaaggca aatgcgtggt tttcgatgggt agtgaagatt accacgcccg cattaacgat 1320
 ccgagcctgg atatcgatga acgtaccatt ctgggtgattc gcggtgcagg tccgctgggt 1380
 tggccgggca gcgcggaagt ggttaatatg cagccgccag atgcgctgct gaaaaaggc 1440
 attacgtctc tgccgaccat cgggtgatggc cgccagagtg gtacggccga tagtccgagc 1500
 atcctgaatg caagtccgga aagcgccgca ggcggtggcc tggcctggct gcgtaccggt 1560
 gatgtgattc gcatcgattt taatcagggt aaatgtgatg cactgggtcc ggatgcggaa 1620
 ctggcgcccc gtaaagccga tggcattccg gcagttccgg cggatgcaac cccgtggcag 1680
 cgtatctacc gccagagcgt gaccagctg tctgatggcg cagttctgga aggtgcagcg 1740

gatttccgtc gcattgcgga aaaaatgccg cgccataacc actga

1785

<210> 76

<211> 1785

<212> ADN

<213> Herbaspirillum seropedicae

<220>

<221> CDS

<222> (1)..(1785)

<400> 76

atg aac aag cca aac gcc acc ccc cgt cgc ttc cgc tcc cag gac tgg	48
Met Asn Lys Pro Asn Ala Thr Pro Arg Arg Phe Arg Ser Gln Asp Trp	
1 5 10 15	
ttc gac aac ccg gac cac atc gac atg acc gcg ctc tac ctg gag cgc	96
Phe Asp Asn Pro Asp His Ile Asp Met Thr Ala Leu Tyr Leu Glu Arg	
20 25 30	
ttc atg aac tac gcc atc acg gcc gag gag ctg cgc tcg gga cgt ccc	144
Phe Met Asn Tyr Gly Ile Thr Ala Glu Glu Leu Arg Ser Gly Arg Pro	
35 40 45	
atc atc gcc atc gcc cag agc gcc agc gac atc agc ccc tgc aac cgc	192
Ile Ile Gly Ile Ala Gln Ser Gly Ser Asp Ile Ser Pro Cys Asn Arg	
50 55 60	
atc cac ctg gag ctg gcc aag cgg gtg cgc gat gcc atc cgc gat gcc	240
Ile His Leu Glu Leu Ala Lys Arg Val Arg Asp Gly Ile Arg Asp Ala	
65 70 75 80	
ggc ggc att ccc atg gaa ttc ccg ctg cat ccc atc ttc gag aac tgc	288
Gly Gly Ile Pro Met Glu Phe Pro Leu His Pro Ile Phe Glu Asn Cys	
85 90 95	
cgc cgc ccc acc gcc gcc atc gac cgc aac ctg gcc tac ctg ggc ctg	336
Arg Arg Pro Thr Ala Ala Ile Asp Arg Asn Leu Ala Tyr Leu Gly Leu	
100 105 110	
gtg gaa atc ctg cac gcc tat ccc atc gac gcc gtg gtg ctc acc acc	384
Val Glu Ile Leu His Gly Tyr Pro Ile Asp Ala Val Val Leu Thr Thr	
115 120 125	
ggc tgc gac aag acc acg ccc tcg cag atc atg gcc gca gcc acc gtc	432
Gly Cys Asp Lys Thr Thr Pro Ser Gln Ile Met Ala Ala Ala Thr Val	
130 135 140	
gat atc ccc gcc atc gtg ctc tcc gcc gcc ccc atg ctg gac ggc tgg	480
Asp Ile Pro Ala Ile Val Leu Ser Gly Gly Pro Met Leu Asp Gly Trp	
145 150 155 160	
atg gat ggc gaa ctg gtc gcc tcc gcc tcg gcc atc tgg aag ggc cgc	528
Met Asp Gly Glu Leu Val Gly Ser Gly Ser Ala Ile Trp Lys Gly Arg	
165 170 175	
aag ctg ctc tcg gca gcc agc atc gac aac gag aaa ttc ctg gag atc	576
Lys Leu Leu Ser Ala Gly Ser Ile Asp Asn Glu Lys Phe Leu Glu Ile	
180 185 190	
gcg gcg gcc tcc gcg ccc tcg tcc gcc cac tgc aac acc atg ggc acc	624
Ala Ala Ala Ser Ala Pro Ser Ser Gly His Cys Asn Thr Met Gly Thr	
195 200 205	
gcc tcc acc atg aac gcc atg gcc gag gcc ctg gcc atg tcg ctg acc	672
Ala Ser Thr Met Asn Ala Met Ala Glu Ala Leu Gly Met Ser Leu Thr	
210 215 220	
ggc tgc tcg gcc att ccc gcg ccc tac cgc gaa cgc gcc cag atg gcc	720
Gly Cys Ser Ala Ile Pro Ala Pro Tyr Arg Glu Arg Gly Gln Met Ala	

225	230	235	240	
tat gag acg ggc cgg cgc atc gtc ggc atg gcc tac gaa gac ctg cgt				768
Tyr Glu Thr Gly Arg Arg Ile Val Gly Met Ala Tyr Glu Asp Leu Arg				
	245	250	255	
oct tcc gcc atc ctc acg cgc gac gcc ttc ctc gac gcc atc gtg gtc				816
Pro Ser Ala Ile Leu Thr Arg Asp Ala Phe Leu Asp Ala Ile Val Val				
	260	265	270	
aat gcc gcc atc ggc ggc tcc acc aat gcc cag ccg cac atc atg gcc				864
Asn Ala Ala Ile Gly Gly Ser Thr Asn Ala Gln Pro His Ile Met Ala				
	275	280	285	
atg gcg cgc cat gcg ggc gtg gaa ctg cag tcg gaa gac tgg atg aag				912
Met Ala Arg His Ala Gly Val Glu Leu Gln Ser Glu Asp Trp Met Lys				
	290	295	300	
tac ggc tac gac gtg cca ctg ctg ctg aac atg cag ccg gcc ggc aag				960
Tyr Gly Tyr Asp Val Pro Leu Leu Leu Asn Met Gln Pro Ala Gly Lys				
	305	310	315	
tac ctg ggc gaa cgc ttc cac cgc gcc ggc ggc gtg ccg gcc atc atg				1008
Tyr Leu Gly Glu Arg Phe His Arg Ala Gly Gly Val Pro Ala Ile Met				
	325	330	335	
tgg gaa ctg cag cag gcc ggc aag ctg cgc gcc gag cgc atc acc gct				1056
Trp Glu Leu Gln Gln Ala Gly Lys Leu Arg Ala Glu Arg Ile Thr Ala				
	340	345	350	
acc ggc aag acc atg gcc gag aac ctg cag ggc ccg gca tcg aac gac				1104
Thr Gly Lys Thr Met Ala Glu Asn Leu Gln Gly Arg Ala Ser Asn Asp				
	355	360	365	
cgg gaa atg att tac cca ttt gcc gcc ccg ctg cgc gag cgc gcc ggc				1152
Arg Glu Met Ile Tyr Pro Phe Ala Ala Pro Leu Arg Glu Arg Ala Gly				
	370	375	380	
ttc ctg gtg ttg aag ggc aac ctg ttc gac ttc gcc atc atg aag acc				1200
Phe Leu Val Leu Lys Gly Asn Leu Phe Asp Phe Ala Ile Met Lys Thr				
	385	390	395	
agc gtc atc tcg gaa acc ttc cgc gaa cgc tat ctc agc acg ccc ggc				1248
Ser Val Ile Ser Glu Thr Phe Arg Glu Arg Tyr Leu Ser Thr Pro Gly				
	405	410	415	
cag gag aac atc ttc gaa tgc cgc gcc gtg gtc ttc gat ggt tcg gac				1296
Gln Glu Asn Ile Phe Glu Cys Arg Ala Val Val Phe Asp Gly Ser Asp				
	420	425	430	
gac tat cac gcc cgc atc aac gac ccg gcg ctc aag atc gac gag aac				1344
Asp Tyr His Ala Arg Ile Asn Asp Pro Ala Leu Lys Ile Asp Glu Asn				
	435	440	445	
acc ctg ctg gcc atc cgt ggc gcc ggc ccg gtg ggc tgg ccc ggt tcg				1392
Thr Leu Leu Ala Ile Arg Gly Ala Gly Pro Val Gly Trp Pro Gly Ser				
	450	455	460	
gcc gag gta gtc aac atg caa ccg ccc gac gcc ctg atc aag cgc ggt				1440
Ala Glu Val Val Asn Met Gln Pro Pro Asp Ala Leu Ile Lys Arg Gly				
	465	470	475	
gta tcg acc ctg ccg acg ctg ggc gat ggc cgc cag tcg ggc acc tcc				1488
Val Ser Thr Leu Pro Thr Leu Gly Asp Gly Arg Gln Ser Gly Thr Ser				
	485	490	495	
gac agc ccc tcc atc ctc aac gcc tcc ccc gag agc gcc gtc ggt ggc				1536
Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Val Gly Gly				
	500	505	510	
ggc ctg gcc tac ctg cgc gat ggc gac cgc gtg cgc atc gac ctc aac				1584
Gly Leu Ala Tyr Leu Arg Asp Gly Asp Arg Val Arg Ile Asp Leu Asn				
	515	520	525	
acc ggc gag tgc aac atg ctg gtc agc gag gaa gag ctg gcg cgc cgc				1632
Thr Gly Glu Cys Asn Met Leu Val Ser Glu Glu Glu Leu Ala Arg Arg				

530	535	540	
aag agc gag ggc atc ccg ccg gtg ccg ccc agc cag acg ccc tgg cag			1680
Lys Ser Glu Gly Ile Pro Pro Val Pro Pro Ser Gln Thr Pro Trp Gln			
545	550	555	560
gaa atc tat cgc agc acg gtc ggc cag ctg gag acc ggc gcc tgc atg			1728
Glu Ile Tyr Arg Ser Thr Val Gly Gln Leu Glu Thr Gly Ala Cys Met			
	565	570	575
gaa ctg gcc ttg aag tac cag ggt gtg gcc cag acc ctg ccc cgg cac			1776
Glu Leu Ala Leu Lys Tyr Gln Gly Val Ala Gln Thr Leu Pro Arg His			
	580	585	590
aat cac tga			1785
Asn His			

<210> 77
 <211> 594
 <212> PRT
 <213> Herbaspirillum seropedicae

<400> 77

Met Asn Lys Pro Asn Ala Thr Pro Arg Arg Phe Arg Ser Gln Asp Trp	
1 5 10 15	
Phe Asp Asn Pro Asp His Ile Asp Met Thr Ala Leu Tyr Leu Glu Arg	
20 25 30	
Phe Met Asn Tyr Gly Ile Thr Ala Glu Glu Leu Arg Ser Gly Arg Pro	
35 40 45	
Ile Ile Gly Ile Ala Gln Ser Gly Ser Asp Ile Ser Pro Cys Asn Arg	
50 55 60	
Ile His Leu Glu Leu Ala Lys Arg Val Arg Asp Gly Ile Arg Asp Ala	
65 70 75 80	
Gly Gly Ile Pro Met Glu Phe Pro Leu His Pro Ile Phe Glu Asn Cys	
85 90 95	
Arg Arg Pro Thr Ala Ala Ile Asp Arg Asn Leu Ala Tyr Leu Gly Leu	
100 105 110	
Val Glu Ile Leu His Gly Tyr Pro Ile Asp Ala Val Val Leu Thr Thr	
115 120 125	
Gly Cys Asp Lys Thr Thr Pro Ser Gln Ile Met Ala Ala Ala Thr Val	
130 135 140	
Asp Ile Pro Ala Ile Val Leu Ser Gly Gly Pro Met Leu Asp Gly Trp	
145 150 155 160	
Met Asp Gly Glu Leu Val Gly Ser Gly Ser Ala Ile Trp Lys Gly Arg	
165 170 175	
Lys Leu Leu Ser Ala Gly Ser Ile Asp Asn Glu Lys Phe Leu Glu Ile	
180 185 190	
Ala Ala Ala Ser Ala Pro Ser Ser Gly His Cys Asn Thr Met Gly Thr	
195 200 205	
Ala Ser Thr Met Asn Ala Met Ala Glu Ala Leu Gly Met Ser Leu Thr	
210 215 220	
Gly Cys Ser Ala Ile Pro Ala Pro Tyr Arg Glu Arg Gly Gln Met Ala	
225 230 235 240	
Tyr Glu Thr Gly Arg Arg Ile Val Gly Met Ala Tyr Glu Asp Leu Arg	
245 250 255	
Pro Ser Ala Ile Leu Thr Arg Asp Ala Phe Leu Asp Ala Ile Val Val	
260 265 270	
Asn Ala Ala Ile Gly Gly Ser Thr Asn Ala Gln Pro His Ile Met Ala	
275 280 285	
Met Ala Arg His Ala Gly Val Glu Leu Gln Ser Glu Asp Trp Met Lys	

290 295 300
 Tyr Gly Tyr Asp Val Pro Leu Leu Leu Asn Met Gln Pro Ala Gly Lys
 305 310 315 320
 Tyr Leu Gly Glu Arg Phe His Arg Ala Gly Gly Val Pro Ala Ile Met
 325 330 335
 Trp Glu Leu Gln Gln Ala Gly Lys Leu Arg Ala Glu Arg Ile Thr Ala
 340 345 350
 Thr Gly Lys Thr Met Ala Glu Asn Leu Gln Gly Arg Ala Ser Asn Asp
 355 360 365
 Arg Glu Met Ile Tyr Pro Phe Ala Ala Pro Leu Arg Glu Arg Ala Gly
 370 375 380
 Phe Leu Val Leu Lys Gly Asn Leu Phe Asp Phe Ala Ile Met Lys Thr
 385 390 395 400
 Ser Val Ile Ser Glu Thr Phe Arg Glu Arg Tyr Leu Ser Thr Pro Gly
 405 410 415
 Gln Glu Asn Ile Phe Glu Cys Arg Ala Val Val Phe Asp Gly Ser Asp
 420 425 430
 Asp Tyr His Ala Arg Ile Asn Asp Pro Ala Leu Lys Ile Asp Glu Asn
 435 440 445
 Thr Leu Leu Ala Ile Arg Gly Ala Gly Pro Val Gly Trp Pro Gly Ser
 450 455 460
 Ala Glu Val Val Asn Met Gln Pro Pro Asp Ala Leu Ile Lys Arg Gly
 465 470 475 480
 Val Ser Thr Leu Pro Thr Leu Gly Asp Gly Arg Gln Ser Gly Thr Ser
 485 490 495
 Asp Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ser Ala Val Gly Gly
 500 505 510
 Gly Leu Ala Tyr Leu Arg Asp Gly Asp Arg Val Arg Ile Asp Leu Asn
 515 520 525
 Thr Gly Glu Cys Asn Met Leu Val Ser Glu Glu Glu Leu Ala Arg Arg
 530 535 540
 Lys Ser Glu Gly Ile Pro Pro Val Pro Pro Ser Gln Thr Pro Trp Gln
 545 550 555 560
 Glu Ile Tyr Arg Ser Thr Val Gly Gln Leu Glu Thr Gly Ala Cys Met
 565 570 575
 Glu Leu Ala Leu Lys Tyr Gln Gly Val Ala Gln Thr Leu Pro Arg His
 580 585 590
 Asn His

<210> 78
 <211> 1785
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 78
 atgaacaaac cgaatgccac cccgcgtcgc ttctgtagcc aggattgggt cgataaccgc 60
 gatcatattg atatgacggc gctgtatctg gaacgcttta tgaattacgg cattaccgcc 120
 gaagaactgc gtagcggctc cccgattatc ggcacgcac agagtggtag cgatatttct 180
 ccgtgcaacc gtatccatct ggaactggca aaacgtgttc gcgatgggat tcgcgatgcg 240
 ggcggtatcc cgatggaatt tccgctgcac ccgattttcg aaaattgccg tcgcccgcac 300
 gccgcaattg atcgtaattt ggcataatct ggcctgggtg aaattctgca tggttaccgc 360
 atcgatgcgg tggttctgac cacgggctgc gataaaacca cgccgagtca gattatggca 420
 gcggccaccg tggatattcc ggcgatcgtt ctgagcggcg gtccgatgct ggatggttgg 480

```

atggatggcg aactgggttg ttctggcagt gccatctgga aaggccgcaa actgctgagc 540
gcaggttcta tcgataacga aaagttcctg gaaattgcag ccgcatctgc cccgagctct 600
ggtcactgca acaccatggg tacggcaagt accatgaatg cgatggccga agcactgggc 660
atgagcctga cgggttggtc tgcaattccg gcgccgtatc gtgaacgcgg ccagatggcc 720
tacgaaaccg gccgtcgcat tgtgggtatg gcgtatgaag atctgcgcc gagcgccatc 780
ctgaccogtg atgcctttct ggatgcaatt gtggttaacg cagcgatcgg cggtagcacc 840
aatgcacagc cgcataattat ggogatggcc cgtcacgcgg gcgttgaact gcagtctgaa 900
gattggatga aatatggtta cgatgtgccg ctgctgctga acatgcagcc ggccggtaaa 960
tatctgggcg aacgctttca togcgcaggc ggtgttcggc cgattatgtg ggaactgcag 1020
caggccggta aactgcgtgc agaacgcac accgcgacgg gcaaaaccat ggccgaaaac 1080
ctgcagggtc gtgcaagcaa tgatcgcgaa atgatttacc cgtttgccgc accgctgcgt 1140
gaacgtgccg gcttctctgt tctgaaaggt aacctgtttg atttcgccat tatgaaaacg 1200
agtgtgatca gcgaaacctt tctgaaacgc tatctgtcta cgccgggcca ggaaaatatc 1260
tttgaatgtc gtgcggtggt ttctgatggt agtcatgatt accacgcacg cattaacgat 1320
ccggcgctga aaatcgatga aaataacctg ctggccattc gtggtgccgg tccggttggt 1380
tggccgggta gcgcggaagt ggttaacatg cagccgccag atgccctgat caaacgtggc 1440
gtgagtacgc tgccgacctt ggtgatggc cgccagagcg gcaccagcga ttctccgagt 1500
attctgaatg cctctccgga aagtgcagtg ggcggtggcc tggcatatct gcgtgatggc 1560
gatcgtgttc gcatcgatct gaacacgggt gaatgtaata tgctggtgag tgaagaagaa 1620
ctggcgctc gcaaaagcga aggcattccg ccggttccgc cgtctcagac cccgtggcag 1680
gaaatctatc gtagcacggt gggccagctg gaaaccggtg cgtgcatgga actggccctg 1740
aaatatcagg gtgtggccca gacctgccc cgtcataatc actga 1785

```

<210> 79
 <211> 1755
 <212> ADN
 <213> *Actinoplanes missouriensis*

<220>
 <221> CDS
 <222> (1)..(1755)

```

<400> 79
atg gac cta cac agg ccc cgg aag tta cga agg agg ctg gcg atg cag 48
Met Asp Leu His Arg Pro Arg Lys Leu Arg Arg Arg Leu Ala Met Gln
1 5 10 15
cga cgc agt gca cag tgg tac gcc ggt gac gac cgg aac agc tac atc 96
Arg Arg Ser Ala Gln Trp Tyr Ala Gly Asp Asp Arg Asn Ser Tyr Ile
20 25 30
cac cgc gcc tgg atg cgc cgg ggc ctg ccg gcc gac gcg ttc gac ggg 144
His Arg Ala Trp Met Arg Arg Gly Leu Pro Ala Asp Ala Phe Asp Gly
35 40 45
cgc ccg cac atc gcc atc gcc aac acc gcc tcg gac ctg acc ccg tgc 192
Arg Pro His Ile Ala Ile Ala Asn Thr Ala Ser Asp Leu Thr Pro Cys
50 55 60
aac gcc cac ttc gac gag gtc gcc cgc agc gtc gcc gac ggc atc cac 240
Asn Ala His Phe Asp Glu Val Ala Arg Ser Val Ala Asp Gly Ile His
65 70 75 80
cgg gcg ggc ggc gtc gcg ctg aac ctg ccc gtc gtg tcg atc ggc gag 288
Arg Ala Gly Gly Val Ala Leu Asn Leu Pro Val Val Ser Ile Gly Glu
85 90 95
acc cag gtc cgg ccc acc gcg atg ctg tgg cga aac atg gcc gcg atg 336
Thr Gln Val Arg Pro Thr Ala Met Leu Trp Arg Asn Met Ala Ala Met
100 105 110
gcg atc gag gag atg ctg cgc gcc aac ccg atc gac ggc gtc gtc ctg 384
Ala Ile Glu Glu Met Leu Arg Ala Asn Pro Ile Asp Gly Val Val Leu

```

ctc	ggc	ggc	tgc	gac	aag	acc	atc	ccc	gcg	ctg	ctc	atg	ggc	gcc	gcc	432
Leu	Gly	Gly	Cys	Asp	Lys	Thr	Ile	Pro	Ala	Leu	Leu	Met	Gly	Ala	Ala	
130							135				140					
tgc	gtc	gac	ctg	ccg	gcc	gtc	gtg	atg	ccc	ggc	ggc	ccg	atg	ctg	acc	480
Ser	Val	Asp	Leu	Pro	Ala	Val	Val	Met	Pro	Gly	Gly	Pro	Met	Leu	Thr	
145					150					155					160	
ggc	acc	ttc	ccg	ggc	gtg	ccg	ctc	ggc	tgc	ggc	acc	gac	gtg	tgg	aag	528
Gly	Thr	Phe	Arg	Gly	Val	Pro	Leu	Gly	Cys	Gly	Thr	Asp	Val	Trp	Lys	
			165					170						175		
ctg	agc	gag	gag	gtc	cgg	gcc	ggc	acc	ctg	agc	gcc	gcc	gag	ttc	acc	576
Leu	Ser	Glu	Glu	Val	Arg	Ala	Gly	Thr	Leu	Ser	Ala	Ala	Glu	Phe	Thr	
			180					185					190			
cgc	tcc	gag	tca	tgc	atg	atc	agg	agc	aag	ggc	cac	tgc	aac	acc	atg	624
Arg	Ser	Glu	Ser	Ser	Met	Ile	Arg	Ser	Lys	Gly	His	Cys	Asn	Thr	Met	
	195						200				205					
ggc	acg	gcg	tgc	acc	atg	ggc	ctg	ctc	gcc	gaa	gtg	ctc	ggc	atg	acc	672
Gly	Thr	Ala	Ser	Thr	Met	Gly	Leu	Leu	Ala	Glu	Val	Leu	Gly	Met	Thr	
	210					215					220					
ctg	ccg	ggg	gtg	gcc	ggc	acg	ccc	gcc	ccg	gac	agc	ccg	ctg	ctg	gag	720
Leu	Pro	Gly	Val	Ala	Gly	Thr	Pro	Ala	Pro	Asp	Ser	Arg	Leu	Leu	Glu	
225					230					235					240	
gcc	gcc	cac	gcg	acc	ggg	gtg	ctc	gcc	gtc	ggc	ctg	gtc	gac	gcg	gac	768
Ala	Ala	His	Ala	Thr	Gly	Val	Leu	Ala	Val	Gly	Leu	Val	Asp	Ala	Asp	
			245					250					255			
cgc	cgc	ccg	agc	cag	gtg	atg	acc	cgc	ggc	tgc	ttc	ctc	aac	gcg	atc	816
Arg	Arg	Pro	Ser	Gln	Val	Met	Thr	Arg	Gly	Ser	Phe	Leu	Asn	Ala	Ile	
		260						265				270				
gtc	gcg	ctc	ggc	gcc	ctg	ggc	ggc	tcc	acc	aac	gcc	gtc	gtg	cac	ctg	864
Val	Ala	Leu	Ala	Ala	Leu	Gly	Gly	Ser	Thr	Asn	Ala	Val	Val	His	Leu	
	275					280					285					
ctg	gcc	atc	gcc	ggc	cgg	ctc	ggc	gtg	ccg	ctg	tcc	cag	gac	gac	ttc	912
Leu	Ala	Ile	Ala	Gly	Arg	Leu	Gly	Val	Pro	Leu	Ser	Gln	Asp	Asp	Phe	
	290					295					300					
gac	acc	acc	ggc	gcc	gac	gtt	ccg	ctg	ctg	gtc	gac	ctg	ctc	ccg	gcc	960
Asp	Thr	Thr	Gly	Ala	Asp	Val	Pro	Leu	Leu	Val	Asp	Leu	Leu	Pro	Ala	
305					310					315					320	
ggc	cgc	ttc	ctg	atg	gac	gac	ctc	tac	cgc	ggc	ggc	ggc	ctg	cac	gcc	1008
Gly	Arg	Phe	Leu	Met	Asp	Asp	Leu	Tyr	Arg	Ala	Gly	Gly	Leu	His	Ala	
			325					330					335			
gtc	ctc	gcc	gag	gtc	cgt	gac	ctg	ctc	gac	ccg	tcc	gcg	atc	acg	gtc	1056
Val	Leu	Ala	Glu	Val	Arg	Asp	Leu	Leu	Asp	Pro	Ser	Ala	Ile	Thr	Val	
		340						345				350				
acc	ggc	ccg	ccg	ctg	acc	gag	cac	ctc	ggc	gac	gcc	cgc	gtc	cac	gac	1104
Thr	Gly	Arg	Pro	Leu	Thr	Glu	His	Leu	Gly	Asp	Ala	Arg	Val	His	Asp	
		355					360					365				
cgc	gag	gtg	atc	cgg	ccg	cgg	gcc	gaa	ccg	ttg	ctg	ccg	cac	gcc	ggg	1152
Arg	Glu	Val	Ile	Arg	Pro	Arg	Ala	Glu	Pro	Leu	Leu	Pro	His	Ala	Gly	
	370					375					380					
atc	gcg	gtc	ctc	tac	ggc	aac	ctg	gcg	ccg	gac	ggc	gcg	gtc	gtc	aaa	1200
Ile	Ala	Val	Leu	Tyr	Gly	Asn	Leu	Ala	Pro	Asp	Gly	Ala	Val	Val	Lys	
385					390					395					400	
ccg	gcc	gcc	gcg	tcc	gag	cac	ttg	ctg	cgc	cac	ccg	ggc	ccg	gcc	gtg	1248
Pro	Ala	Ala	Ala	Ser	Glu	His	Leu	Leu	Arg	His	Arg	Gly	Pro	Ala	Val	
			405					410				415				
gtc	ttc	gac	tcc	gtc	gag	gac	ctg	cac	gcc	cgg	ctc	gac	gac	ccg	gac	1296
Val	Phe	Asp	Ser	Val	Glu	Asp	Leu	His	Ala	Arg	Leu	Asp	Asp	Pro	Asp	

ctg gac gtc acc gcc gac tgc gtg ctg gtg ctg cgc ggc tgc ggt ccc	420	425	430	1344
Leu Asp Val Thr Ala Asp Ser Val Leu Val Leu Arg Gly Cys Gly Pro				
	435	440	445	
aag ggc tat ccg ggc atg ccg gag gtg tcc aac atg ccg ctg ccg gcg				1392
Lys Gly Tyr Pro Gly Met Pro Glu Val Ser Asn Met Pro Leu Pro Ala				
	450	455	460	
aaa ctc ctc gaa cag ggg gtc cgc gac atg gtc cgg gtc tgc gac ggg				1440
Lys Leu Leu Glu Gln Gly Val Arg Asp Met Val Arg Val Cys Asp Gly				
	465	470	475	480
cgg atg tgc ggt acg gcg tac ggc acg gtg gtc ctg cac gtc gcc ccg				1488
Arg Met Ser Gly Thr Ala Tyr Gly Thr Val Val Leu His Val Ala Pro				
	485	490	495	
gaa gcc gcg gcg ggc ggg ccg ctc gcc cgg gtc cgc acc ggc gac atg				1536
Glu Ala Ala Ala Gly Gly Pro Leu Ala Arg Val Arg Thr Gly Asp Met				
	500	505	510	
atc atc ctc gac gtc gcg aac cgg cgc ctc gac gcc gac gtc ccg gcc				1584
Ile Ile Leu Asp Val Ala Asn Arg Arg Leu Asp Ala Asp Val Pro Ala				
	515	520	525	
gag gag tgg gcc gcc cgc gag ccg tca ccg gag gcg gcg aaa gcc tac				1632
Glu Glu Trp Ala Ala Arg Glu Pro Ser Pro Glu Ala Ala Lys Ala Tyr				
	530	535	540	
gcg gcg ccg tcc cgc ggc tgg gag cgt ctc tac gtc gac acc gtc ggc				1680
Ala Ala Pro Ser Arg Gly Trp Glu Arg Leu Tyr Val Asp Thr Val Gly				
	545	550	555	560
cag gcc gac acc ggc gcc gac tgc gac ttc ctg cgc ggc gcg agc ggc				1728
Gln Ala Asp Thr Gly Ala Asp Cys Asp Phe Leu Arg Gly Ala Ser Gly				
	565	570	575	
gac cgc gtc tcc cgc gag tcc cac tga				1755
Asp Arg Val Ser Arg Glu Ser His				
	580			

<210> 80

<211> 584

<212> PRT

<213> Actinoplanes missouriensis

<400> 80

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

145					150					155					160
Gly	Thr	Phe	Arg	Gly	Val	Pro	Leu	Gly	Cys	Gly	Thr	Asp	Val	Trp	Lys
				165					170					175	
Leu	Ser	Glu	Glu	Val	Arg	Ala	Gly	Thr	Leu	Ser	Ala	Ala	Glu	Phe	Thr
			180					185					190		
Arg	Ser	Glu	Ser	Ser	Met	Ile	Arg	Ser	Lys	Gly	His	Cys	Asn	Thr	Met
		195					200					205			
Gly	Thr	Ala	Ser	Thr	Met	Gly	Leu	Leu	Ala	Glu	Val	Leu	Gly	Met	Thr
	210					215					220				
Leu	Pro	Gly	Val	Ala	Gly	Thr	Pro	Ala	Pro	Asp	Ser	Arg	Leu	Leu	Glu
225					230					235					240
Ala	Ala	His	Ala	Thr	Gly	Val	Leu	Ala	Val	Gly	Leu	Val	Asp	Ala	Asp
				245					250					255	
Arg	Arg	Pro	Ser	Gln	Val	Met	Thr	Arg	Gly	Ser	Phe	Leu	Asn	Ala	Ile
			260					265					270		
Val	Ala	Leu	Ala	Ala	Leu	Gly	Gly	Ser	Thr	Asn	Ala	Val	Val	His	Leu
		275					280					285			
Leu	Ala	Ile	Ala	Gly	Arg	Leu	Gly	Val	Pro	Leu	Ser	Gln	Asp	Asp	Phe
	290					295					300				
Asp	Thr	Thr	Gly	Ala	Asp	Val	Pro	Leu	Leu	Val	Asp	Leu	Leu	Pro	Ala
305					310					315					320
Gly	Arg	Phe	Leu	Met	Asp	Asp	Leu	Tyr	Arg	Ala	Gly	Gly	Leu	His	Ala
				325					330					335	
Val	Leu	Ala	Glu	Val	Arg	Asp	Leu	Leu	Asp	Pro	Ser	Ala	Ile	Thr	Val
			340					345					350		
Thr	Gly	Arg	Pro	Leu	Thr	Glu	His	Leu	Gly	Asp	Ala	Arg	Val	His	Asp
		355					360					365			
Arg	Glu	Val	Ile	Arg	Pro	Arg	Ala	Glu	Pro	Leu	Leu	Pro	His	Ala	Gly
	370					375					380				
Ile	Ala	Val	Leu	Tyr	Gly	Asn	Leu	Ala	Pro	Asp	Gly	Ala	Val	Val	Lys
385					390					395					400
Pro	Ala	Ala	Ala	Ser	Glu	His	Leu	Leu	Arg	His	Arg	Gly	Pro	Ala	Val
				405					410					415	
Val	Phe	Asp	Ser	Val	Glu	Asp	Leu	His	Ala	Arg	Leu	Asp	Asp	Pro	Asp
			420					425					430		
Leu	Asp	Val	Thr	Ala	Asp	Ser	Val	Leu	Val	Leu	Arg	Gly	Cys	Gly	Pro
		435					440					445			
Lys	Gly	Tyr	Pro	Gly	Met	Pro	Glu	Val	Ser	Asn	Met	Pro	Leu	Pro	Ala
	450					455					460				
Lys	Leu	Leu	Glu	Gln	Gly	Val	Arg	Asp	Met	Val	Arg	Val	Cys	Asp	Gly
465					470					475					480
Arg	Met	Ser	Gly	Thr	Ala	Tyr	Gly	Thr	Val	Val	Leu	His	Val	Ala	Pro
				485					490					495	
Glu	Ala	Ala	Ala	Gly	Gly	Pro	Leu	Ala	Arg	Val	Arg	Thr	Gly	Asp	Met
			500					505					510		
Ile	Ile	Leu	Asp	Val	Ala	Asn	Arg	Arg	Leu	Asp	Ala	Asp	Val	Pro	Ala
		515					520					525			
Glu	Glu	Trp	Ala	Ala	Arg	Glu	Pro	Ser	Pro	Glu	Ala	Ala	Lys	Ala	Tyr
	530					535					540				
Ala	Ala	Pro	Ser	Arg	Gly	Trp	Glu	Arg	Leu	Tyr	Val	Asp	Thr	Val	Gly
545					550					555					560
Gln	Ala	Asp	Thr	Gly	Ala	Asp	Cys	Asp	Phe	Leu	Arg	Gly	Ala	Ser	Gly
				565					570					575	
Asp	Arg	Val	Ser	Arg	Glu	Ser	His								
				580											

<210> 81
 <211> 1755
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 81
 atggatctgc atcgtccgcg taaactgcgt cgccgtctgg ccatgcagcg ccgtagcgcc 60
 cagtggatag caggcgatga tcgcaactct tacattcatc gtgcatggat gcgccgtggg 120
 ctgccggcag atgcatttga tggccgtccg cacattgcga tcgccaacac ggcgctctgat 180
 ctgaccccggt gcaatgccca tttcgatgaa gttgcgcgca gtgtggccga tggatttcac 240
 cgtgcaggcg gtgttgcgct gaatctgcgc gtggttagca ttggtgaaac ccaagtgcgt 300
 ccgaccgcca tgctgtggcg taacatggcg gccatggcaa ttgaagaaat gctgcgtgcg 360
 aatccgatcg atgggtgtgt tctgctgggc ggttgtgata aaacgattcc ggccctgctg 420
 atgggcgcgag cgagtgttga tctgccggca gtggttatgc cgggcggtcc gatgctgacc 480
 ggcacgtttc gcggtgtgcc gctgggttgt ggcaccgatg tttgaaaact gagcgaagaa 540
 gtgcgtgcgg gtaçgctgtc tgccgcagaa tttaccgcga gtgaaagctc tatgatccgt 600
 agcaaaggtc attgtaacac catgggcacg gcaagcacca tgggtctgct ggcagaagtt 660
 ctgggcatga cgctgccggg tgtggcgggc accccggcac cggattctcg tctgctggaa 720
 gccgcacatg ccaccgggtg gctggcggtt ggccctgggtg atgccgatcg ccgtccgtct 780
 caggttatga cgcgtggcag ttttctgaat gcaattgtgg cactggcggc gctggcgggc 840
 agtaccaatg cagtggttca tctgctggca attgccggcc gcctgggtgt gccgtgagc 900
 caggatgatt ttgataccac ggggtgccga gttccgctgc tggttgatct gctgccggca 960
 ggccgcttcc tgatggatga tctgtatcgt gcgggcggtc tgcatgcggt tctggcagaa 1020
 gtgcgcgatc tgctggatcc gagcgcgatc accgtgaccg gtcgtccgct gaccgaacat 1080
 ctgggcgatg cgcgcgttca cgatcgtgaa gtgattcgtc cgcgtgcgga accgtgctg 1140
 ccgcatgcag gtattgcggt gctgtacggt aatctggccc cggatgggtc ggtggttaaa 1200
 ccggcggcag ccagcgaaca tctgctgcgt caccgtggtc cggcggtggg tttcgattct 1260
 gttgaagatc tgcacgcacg cctggatgat ccggatctgg atgtgaccgc ggattctgtg 1320
 ctggttctgc gtggctgcgg tccgaaaggc tatccgggta tgccggaagt tagtaacatg 1380
 ccgctgcggc cgaaactgct ggaacagggt gtgcgcgata tgggtgcgtgt ttgtgatggc 1440
 cgtatgtctg gtacggcata cggcaccgtg gttctgcata ttgccccgga agcggcagcg 1500
 ggcggtccgc tggcacgcgt gcgtaccggc gatatgatta tcctggatgt tgccaatcgc 1560
 cgtctggatg cagatgtgcc ggccgaagaa tggggcgcac gtgaaccgag tccggaagcc 1620
 gcaaaagcct atgcggcccc gagccgtggt tgggaacgtc tgtacgttga tacggtgggt 1680
 caggccgata ccggcgcgaga ttgcgatttt ctgctggcg cgagcggtga tcgctgaggt 1740
 cgtgaaagcc actga 1755

<210> 82
 <211> 1962
 <212> ADN
 <213> *Aspergillus oryzae*

<220>
 <221> CDS
 <222> (1)..(1962)

<400> 82
 atg tgc tgt pag tca cgc acc tgc tgt gaa ggc tgc tcc tgc agt gac 48
 Met Ser Cys Gln Ser Arg Thr Ser Cys Glu Gly Cys Ser Cys Ser Asp
 1 5 10 15
 ggg ggc agc cgg cca cca gtg aac atc gag gac tgt gag agc gaa ctg 96
 Gly Gly Ser Arg Pro Pro Val Asn Ile Glu Asp Cys Glu Ser Glu Leu
 20 25 30

ctt gcc ctt cgg agg cgt aca gtg gag ctc gag aaa act cta gca tca	144
Leu Ala Leu Arg Arg Arg Thr Val Glu Leu Glu Lys Thr Leu Ala Ser	
35 40 45	
atg caa gat ggc cgg cca cat gct aat gca tca agg gcc cgg aag tta	192
Met Gln Asp Gly Arg Pro His Ala Asn Ala Ser Arg Ala Arg Lys Leu	
50 55 60	
cgg tcg gcc aac tgg ttc aac tgc gag agc gat ccg ggc atg atg gcc	240
Arg Ser Ala Asn Trp Phe Asn Cys Glu Ser Asp Pro Gly Met Met Ala	
65 70 75 80	
ctc tac att gag cgg tac ctc aac tat ggt atc acc cgg gaa gaa ctg	288
Leu Tyr Ile Glu Arg Tyr Leu Asn Tyr Gly Ile Thr Arg Glu Glu Leu	
85 90 95	
atg tct ggt aaa cca ata atc gga ata gca cag tcc ggg tcc gac ttg	336
Met Ser Gly Lys Pro Ile Ile Gly Ile Ala Gln Ser Gly Ser Asp Leu	
100 105 110	
tct cca tgc aac cgc cat cac ctc gag ttg gcg aaa cgg gtt cga gaa	384
Ser Pro Cys Asn Arg His His Leu Glu Leu Ala Lys Arg Val Arg Glu	
115 120 125	
ggt atc agg tcc gcc gga ggg atc gcc ttt gaa ttc ccc acg cat cct	432
Gly Ile Arg Ser Ala Gly Gly Ile Ala Phe Glu Phe Pro Thr His Pro	
130 135 140	
att cag gag act agc agg aga ccc act gct tgt att gac cga aat cta	480
Ile Gln Glu Thr Ser Arg Arg Pro Thr Ala Cys Ile Asp Arg Asn Leu	
145 150 155 160	
tcg tat ctg ggc ctg gta gaa att ctc ttc gga tat ccc ttg gac ggt	528
Ser Tyr Leu Gly Leu Val Glu Ile Leu Phe Gly Tyr Pro Leu Asp Gly	
165 170 175	
gta gtc ctc ctg acg ggc tgc gac aag act act ccc gcc gct ctg atg	576
Val Val Leu Leu Thr Gly Cys Asp Lys Thr Thr Pro Ala Ala Leu Met	
180 185 190	
gcc gcg gct aca gtg aat atc cca gca ata tgc ttg aat gtc ggc cca	624
Ala Ala Ala Thr Val Asn Ile Pro Ala Ile Cys Leu Asn Val Gly Pro	
195 200 205	
atg ctt aac ggt tac gtg aag aag gac ctg gcg ggg tcc gga atg gta	672
Met Leu Asn Gly Tyr Val Lys Lys Asp Leu Ala Gly Ser Gly Met Val	
210 215 220	
ttg tgg acg ggt aga gag atg tat gcg gca ggt gag atc aac aag gag	720
Leu Trp Thr Gly Arg Glu Met Tyr Ala Ala Gly Glu Ile Asn Lys Glu	
225 230 235 240	
gaa ttc att gac tat gtg tcc aaa ggt aca cca tca gtg gga cat tgt	768
Glu Phe Ile Asp Tyr Val Ser Lys Gly Thr Pro Ser Val Gly His Cys	
245 250 255	
aac acg atg gga aca gca tca act atg aat gct ttg gca gaa gcg ctt	816
Asn Thr Met Gly Thr Ala Ser Thr Met Asn Ala Leu Ala Glu Ala Leu	
260 265 270	
gga atg gcc cta cct ggg tcg gct gcc ata ccg gct ccg tat cgt gaa	864
Gly Met Ala Leu Pro Gly Ser Ala Ala Ile Pro Ala Pro Tyr Arg Glu	
275 280 285	
cgt ggc caa tgt gct tac gaa aca ggt ttg cgt att gtc gaa atg gtt	912
Arg Gly Gln Cys Ala Tyr Glu Thr Gly Leu Arg Ile Val Glu Met Val	
290 295 300	
cat tcc gat cgg aag ccc agc gat att atg act cga gag gct ttc gag	960
His Ser Asp Arg Lys Pro Ser Asp Ile Met Thr Arg Glu Ala Phe Glu	
305 310 315 320	
aat gtt att gta gtt aat act gct atc ggc ggc agt acc aat gcc cct	1008
Asn Val Ile Val Val Asn Thr Ala Ile Gly Gly Ser Thr Asn Ala Pro	
325 330 335	

att cat atc aac gcc att gcc aaa cat att ggg gtt gag gtt tca ctg	1056
Ile His Ile Asn Ala Ile Ala Lys His Ile Gly Val Glu Val Ser Leu	
340 345 350	
gat gac tgg gac cga ctt gga ttt cat att cca cta ctg cta aac atg	1104
Asp Asp Trp Asp Arg Leu Gly Phe His Ile Pro Leu Leu Leu Asn Met	
355 360 365	
caa cca gcc ggt gaa ctg ctg gga gag gaa tac tat cgg gcg ggc ggg	1152
Gln Pro Ala Gly Glu Leu Leu Gly Glu Glu Tyr Tyr Arg Ala Gly Gly	
370 375 380	
cta ccc gcg att atg gcg gag ctg cta gat gct agg aag ctg aac ccg	1200
Leu Pro Ala Ile Met Ala Glu Leu Leu Asp Ala Arg Lys Leu Asn Pro	
385 390 395 400	
gat gct tta aca tgc aat ggt tac aca gta gcc gag aat gtc cgg gac	1248
Asp Ala Leu Thr Cys Asn Gly Tyr Thr Val Ala Glu Asn Val Arg Asp	
405 410 415	
aaa cat acc tgg gac cgt cga atg atc aag ccg tac aat gag cca ctt	1296
Lys His Thr Trp Asp Arg Arg Met Ile Lys Pro Tyr Asn Glu Pro Leu	
420 425 430	
ctg gaa gat gcc ggt ttt ctc cat ctt caa ggc agt ctg ttc cgg tct	1344
Leu Glu Asp Ala Gly Phe Leu His Leu Gln Gly Ser Leu Phe Arg Ser	
435 440 445	
gct atc atg aag aca tgt gtg ata tcg gag cct ttc cga caa aag ttc	1392
Ala Ile Met Lys Thr Cys Val Ile Ser Glu Pro Phe Arg Gln Lys Phe	
450 455 460	
ttg gag aat ccc aag gac ccg aat gca ttt gaa ggt acg gtt gta gta	1440
Leu Glu Asn Pro Lys Asp Pro Asn Ala Phe Glu Gly Thr Val Val Val	
465 470 475 480	
ttt gat gga ccg gag gat tat cat cac cga ctt gag gac cct tcc acc	1488
Phe Asp Gly Pro Glu Asp Tyr His His Arg Leu Glu Asp Pro Ser Thr	
485 490 495	
ccc atc gac gac aga agt atc ctg gtg atg cgc ggt gct ggc cca ttg	1536
Pro Ile Asp Asp Arg Ser Ile Leu Val Met Arg Gly Ala Gly Pro Leu	
500 505 510	
gga tat cca ggt gcc gcc gag gtc gtc aac atg cac cca cct gga cgg	1584
Gly Tyr Pro Gly Ala Ala Glu Val Val Asn Met His Pro Pro Gly Arg	
515 520 525	
ctt tta cga caa ggg gtc aaa tcg ctt ccg tgc att gga gac ggg cga	1632
Leu Leu Arg Gln Gly Val Lys Ser Leu Pro Cys Ile Gly Asp Gly Arg	
530 535 540	
caa tcg gga act tcc ggc tca cca tca atc ctc aat gct agt ccc gag	1680
Gln Ser Gly Thr Ser Gly Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu	
545 550 555 560	
gca gcg gcg ggt ggt aat cta gcc ctt ctt caa gat ggg gac aga ctc	1728
Ala Ala Ala Gly Gly Asn Leu Ala Leu Leu Gln Asp Gly Asp Arg Leu	
565 570 575	
cgt gtt gat cta aat aaa cgg cgc gtt gac atc ctt gtt tcc acg gag	1776
Arg Val Asp Leu Asn Lys Arg Arg Val Asp Ile Leu Val Ser Thr Glu	
580 585 590	
gaa ctg gaa aag cgg aga aag aca cta gaa gcc caa gga ggt tat gat	1824
Glu Leu Glu Lys Arg Arg Lys Thr Leu Glu Ala Gln Gly Gly Tyr Asp	
595 600 605	
gtg ccg gaa agt caa act cca tgg cag gaa ctg ttc agg agg gag acg	1872
Val Pro Glu Ser Gln Thr Pro Trp Gln Glu Leu Phe Arg Arg Glu Thr	
610 615 620	
aca cag ttg agt gat ggt atg gtc ctc cga gac gcg gta aaa tac caa	1920
Thr Gln Leu Ser Asp Gly Met Val Leu Arg Asp Ala Val Lys Tyr Gln	
625 630 635 640	

cga ctt gcc cag cga tat gag aac ccc cgg cac aat cat tga
 Arg Leu Ala Gln Arg Tyr Glu Asn Pro Arg His Asn His
 645 650

1962

<210> 83
 <211> 653
 <212> PRT
 <213> *Aspergillus oryzae*

<400> 83
 Met Ser Cys Gln Ser Arg Thr Ser Cys Glu Gly Cys Ser Cys Ser Asp
 1 5 10 15
 Gly Gly Ser Arg Pro Pro Val Asn Ile Glu Asp Cys Glu Ser Glu Leu
 20 25 30
 Leu Ala Leu Arg Arg Arg Thr Val Glu Leu Glu Lys Thr Leu Ala Ser
 35 40 45
 Met Gln Asp Gly Arg Pro His Ala Asn Ala Ser Arg Ala Arg Lys Leu
 50 55 60
 Arg Ser Ala Asn Trp Phe Asn Cys Glu Ser Asp Pro Gly Met Met Ala
 65 70 75 80
 Leu Tyr Ile Glu Arg Tyr Leu Asn Tyr Gly Ile Thr Arg Glu Glu Leu
 85 90 95
 Met Ser Gly Lys Pro Ile Ile Gly Ile Ala Gln Ser Gly Ser Asp Leu
 100 105 110
 Ser Pro Cys Asn Arg His His Leu Glu Leu Ala Lys Arg Val Arg Glu
 115 120 125
 Gly Ile Arg Ser Ala Gly Gly Ile Ala Phe Glu Phe Pro Thr His Pro
 130 135 140
 Ile Gln Glu Thr Ser Arg Pro Thr Ala Cys Ile Asp Arg Asn Leu
 145 150 155 160
 Ser Tyr Leu Gly Leu Val Glu Ile Leu Phe Gly Tyr Pro Leu Asp Gly
 165 170 175
 Val Val Leu Leu Thr Gly Cys Asp Lys Thr Thr Pro Ala Ala Leu Met
 180 185 190
 Ala Ala Ala Thr Val Asn Ile Pro Ala Ile Cys Leu Asn Val Gly Pro
 195 200 205
 Met Leu Asn Gly Tyr Val Lys Lys Asp Leu Ala Gly Ser Gly Met Val
 210 215 220
 Leu Trp Thr Gly Arg Glu Met Tyr Ala Ala Gly Glu Ile Asn Lys Glu
 225 230 235 240
 Glu Phe Ile Asp Tyr Val Ser Lys Gly Thr Pro Ser Val Gly His Cys
 245 250 255
 Asn Thr Met Gly Thr Ala Ser Thr Met Asn Ala Leu Ala Glu Ala Leu
 260 265 270
 Gly Met Ala Leu Pro Gly Ser Ala Ile Pro Ala Pro Tyr Arg Glu
 275 280 285
 Arg Gly Gln Cys Ala Tyr Glu Thr Gly Leu Arg Ile Val Glu Met Val
 290 295 300
 His Ser Asp Arg Lys Pro Ser Asp Ile Met Thr Arg Glu Ala Phe Glu
 305 310 315 320
 Asn Val Ile Val Val Asn Thr Ala Ile Gly Gly Ser Thr Asn Ala Pro
 325 330 335
 Ile His Ile Asn Ala Ile Ala Lys His Ile Gly Val Glu Val Ser Leu
 340 345 350
 Asp Asp Trp Asp Arg Leu Gly Phe His Ile Pro Leu Leu Leu Asn Met
 355 360 365

Gln Pro Ala Gly Glu Leu Leu Gly Glu Glu Tyr Tyr Arg Ala Gly Gly
 370 375 380
 Leu Pro Ala Ile Met Ala Glu Leu Leu Asp Ala Arg Lys Leu Asn Pro
 385 390 395 400
 Asp Ala Leu Thr Cys Asn Gly Tyr Thr Val Ala Glu Asn Val Arg Asp
 405 410 415
 Lys His Thr Trp Asp Arg Arg Met Ile Lys Pro Tyr Asn Glu Pro Leu
 420 425 430
 Leu Glu Asp Ala Gly Phe Leu His Leu Gln Gly Ser Leu Phe Arg Ser
 435 440 445
 Ala Ile Met Lys Thr Cys Val Ile Ser Glu Pro Phe Arg Gln Lys Phe
 450 455 460
 Leu Glu Asn Pro Lys Asp Pro Asn Ala Phe Glu Gly Thr Val Val Val
 465 470 475 480
 Phe Asp Gly Pro Glu Asp Tyr His His Arg Leu Glu Asp Pro Ser Thr
 485 490 495
 Pro Ile Asp Asp Arg Ser Ile Leu Val Met Arg Gly Ala Gly Pro Leu
 500 505 510
 Gly Tyr Pro Gly Ala Ala Glu Val Val Asn Met His Pro Pro Gly Arg
 515 520 525
 Leu Leu Arg Gln Gly Val Lys Ser Leu Pro Cys Ile Gly Asp Gly Arg
 530 535 540
 Gln Ser Gly Thr Ser Gly Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu
 545 550 555 560
 Ala Ala Ala Gly Gly Asn Leu Ala Leu Leu Gln Asp Gly Asp Arg Leu
 565 570 575
 Arg Val Asp Leu Asn Lys Arg Arg Val Asp Ile Leu Val Ser Thr Glu
 580 585 590
 Glu Leu Glu Lys Arg Arg Lys Thr Leu Glu Ala Gln Gly Gly Tyr Asp
 595 600 605
 Val Pro Glu Ser Gln Thr Pro Trp Gln Glu Leu Phe Arg Arg Glu Thr
 610 615 620
 Thr Gln Leu Ser Asp Gly Met Val Leu Arg Asp Ala Val Lys Tyr Gln
 625 630 635 640
 Arg Leu Ala Gln Arg Tyr Glu Asn Pro Arg His Asn His
 645 650

<210> 84
 <211> 1962
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 84
 atgtcttgcc agagtgcac cagctgtgaa ggttgacgt gttctgatgg cggttctcgt 60
 ccgcccgtta atattgaaga ttgcgaaagt gaactgctgg ccctgcgtcg ccgtacgggtg 120
 gaactggaâa aaacctggc atctatgcag gatggccgtc cgcattgcaa tgcctctcgt 180
 gcacgtaaac tgcgtagtgc aaactggttt aattgtgaaa gtgatccggg tatgatggcg 240
 ctgtatatatt aacgctatct gaattacggc atcaccctg aagaactgat gattggcaaa 300
 ccgattatcg gtattgcaca gattggcagc gatctgagcc cgtgcaaccg ccatcacctg 360
 gaactggcga aacgtgtgcg tgaaggtatt cgtagcgcag gcggtatcgc gtttgaattt 420
 ccgaccacc cgattcagga aacgagccgc cgtccgaccg cgtgtatcga tcgtaattctg 480
 tcttatctgg gtctggtgga aatcctgttt gggtaccgc tggatggcgt ggttctgctg 540
 accggttggtg ataaaaccac ccggggcgcc ctgatggcag ccgcaaccgt taacattccg 600

```

gccatctgtc tgaacgtggg tccgatgctg aatggctatg tgaaaaaaga tctggcaggg 660
agcgggtatg tgctgtggac gggtcgcgaa atgtatgcag cgggcgaaat caacaaagaa 720
gaatttatcg attacgttag caaaggcacc cgtctgtgg gccattgcaa taccatgggc 780
acggcgctcta ccatgaacgc cctggcagaa gccctgggta tggccctgcc gggtagtgcc 840
gcaattccgg ccccgatatg cgaacgtggc cagtgtgcat acgaaacggg cctgcgcatt 900
gttgaaatgg tgcacagtga tcgcaaacgc agcgatatca tgaccctga agcctttgaa 960
aacgtgattg tggttaatac ggcgatcggc ggtagtacca acgccccgat tcatatcaat 1020
gccattgcaa aacacatcgg tgttgaagtg agcctggatg attgggatcg tctgggcttc 1080
catattccgc tgctgtgaa tatgcagccg gcgggcgaac tgctgggtga agaataattat 1140
cgtgccggcg gtctgcggc catcatggc gaactgctgg atgcacgtaa actgaaccgc 1200
gatgcgctga cgtgcaatgg ttataccgtt gcggaaaacg tgccgcgataa acatacgtgg 1260
gatgcgcgta tgatcaaacc gtacaacgaa ccgctgctgg aagatgcggg tttctgcac 1320
ctgcagggct ctctgttccg cagtgccatt atgaaaacct gcgtgatcag cgaaccgttt 1380
cgtcagaaat tcctggaaaa ccgaaagat ccgaatgcgt ttgaaggtag ggtggtggtg 1440
tttgatggcc cgggaagatta tcatcaccgc ctggaagatc cgagcaccgc gattgatgat 1500
cgctctattc tggttatgcg tggcgccggc ccgctgggct acccggtgc ggccgaagtt 1560
gtgaatatgc atccgccggg togcctgctg cgtcagggtg tgaaaagtct gccgtgcatt 1620
ggcgatggtc gtcagagcgg cacctctggt agtccgagca tcctgaacgc gagcccgaa 1680
gcagccgcgg gcggtaacct ggcgctgctg caggatggtg atgcctgcg tgttgatctg 1740
aacaacgcc gtgttgatat tctggtgtct acggaagaac tggaaaaacg ccgtaaaacc 1800
ctggaagcgc agggcggtta tgatgtgccc gaaagccaga ccccggtgca ggaactgttt 1860
cgcggtgaaa ccacgcagct gtctgatggc atggttctgc gcgatgcgct gaaatatcag 1920
cgctggcac agcgttacga aaaccgcgct cataatcact ga 1962

```

<210> 85
 <211> 1161
 <212> ADN
 <213> *Agrobacterium tumefaciens*

<220>
 <221> CDS
 <222> (1)..(1161)

```

<400> 85
atg acc ggc tca agg tcc atc gac gct acc gct att ttg cct gaa gac 48
Met Thr Gly Ser Arg Ser Ile Asp Ala Thr Ala Ile Leu Pro Glu Asp
1 5 10 15
ttt gaa aac gcc ctg ctc gtc ggc cgc gta tgg tcg aaa agc gaa ggc 96
Phe Glu Asn Ala Leu Leu Val Gly Arg Val Trp Ser Lys Ser Glu Gly
20 25 30
ggc ccc tgc ccc gtt ctg cta aaa ggc ggc gtg ctt tac gac ctc acc 144
Gly Pro Cys Pro Val Leu Leu Lys Gly Gly Val Leu Tyr Asp Leu Thr
35 40 45
tct att tcg ccg acc atg tct gaa ctt ctg gaa aag acg gat ctg gtg 192
Ser Ile Ser Pro Thr Met Ser Glu Leu Leu Glu Lys Thr Asp Leu Val
50 55 60
gag ctg ctt gca gac acc ggc aca ttc gtg gcg ctt ggc gcg ctc gag 240
Glu Leu Leu Ala Asp Thr Gly Thr Phe Val Ala Leu Gly Ala Leu Glu
65 70 75 80
gcc ttt ctg gat ggc tcc gct ggt gaa ctt ctc gcc ccc aac gat att 288
Ala Phe Leu Asp Gly Ser Ala Gly Glu Leu Leu Ala Pro Asn Asp Ile
85 90 95
cag gcc gtg aag gct gca ggc gtg acc ttc gcc gac agc atg ctg gaa 336
Gln Ala Val Lys Ala Ala Gly Val Thr Phe Ala Asp Ser Met Leu Glu
100 105 110
cgc gtg atc gag gag cag gcc aag ggc gat cca ctt cgc gcg cag gaa 384

```

Arg Val Ile Glu Glu Gln Ala Lys Gly Asp Pro Leu Arg Ala Gln Glu	
115 120 125	
ata cgc gga cgg ctg gcg ccg gtt ctc ggc gac aat ctc aag gga ctt	432
Ile Arg Gly Arg Leu Ala Pro Val Leu Gly Asp Asn Leu Lys Gly Leu	
130 135 140	
gag gcc ggt tcg gaa aaa gca gcc gaa gtg aag aaa ctg ctg cag gaa	480
Glu Ala Gly Ser Glu Lys Ala Ala Glu Val Lys Lys Leu Leu Gln Glu	
145 150 155 160	
atg ggc ctg tgg tcg caa tat ctg gaa gtt ggc atc ggt ccg gat gcg	528
Met Gly Leu Trp Ser Gln Tyr Leu Glu Val Gly Ile Gly Pro Asp Ala	
165 170 175	
gag ata ttc tcc aag gcg cag gcc atg gcc tcc gtc ggc tgc ggc gcg	576
Glu Ile Phe Ser Lys Ala Gln Ala Met Ala Ser Val Gly Cys Gly Ala	
180 185 190	
ctg atc ggc gtt cac ccg aaa tcg cag tgg aac aac ccg gag ccg gaa	624
Leu Ile Gly Val His Pro Lys Ser Gln Trp Asn Asn Pro Glu Pro Glu	
195 200 205	
gtg gtg ctg gcc att gcc tcc gac ggc cgt atc gtt ggc gcg acg ctt	672
Val Val Leu Ala Ile Ala Ser Asp Gly Arg Ile Val Gly Ala Thr Leu	
210 215 220	
ggc aac gac gtc aac ctg cgt gac ttc gag ggc cgc tcc gcc ctc ctg	720
Gly Asn Asp Val Asn Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu	
225 230 235 240	
ctc agc aag gca aaa gac aac aac gca tcc tgc gca atc ggt ccc ttc	768
Leu Ser Lys Ala Lys Asp Asn Asn Ala Ser Cys Ala Ile Gly Pro Phe	
245 250 255	
atc cgc ctg ttt gat gga cgc ttt acc atc gaa gac gtg aag aag gcg	816
Ile Arg Leu Phe Asp Gly Arg Phe Thr Ile Glu Asp Val Lys Lys Ala	
260 265 270	
cag ata tcg ctt ctg gtc gaa ggc gag gac ggc ttc acc atg acc ggt	864
Gln Ile Ser Leu Leu Val Glu Gly Glu Asp Gly Phe Thr Met Thr Gly	
275 280 285	
gca agc gca atg cag gcg atc agc cgc acg ccg gaa aac ctc gct tcc	912
Ala Ser Ala Met Gln Ala Ile Ser Arg Thr Pro Glu Asn Leu Ala Ser	
290 295 300	
cag ctc ttg aac cgc aac cat cag tat cct gac ggc gcc gtc ttt ttc	960
Gln Leu Leu Asn Arg Asn His Gln Tyr Pro Asp Gly Ala Val Phe Phe	
305 310 315 320	
ctt ggt acg atg ttc gcg ccg gta aag gac cgc cat ggc ccg ggt ctc	1008
Leu Gly Thr Met Phe Ala Pro Val Lys Asp Arg His Gly Pro Gly Leu	
325 330 335	
ggc ttc acc cac tcc aag ggc gac cgt gtc gaa atc tcc acg ccc aag	1056
Gly Phe Thr His Ser Lys Gly Asp Arg Val Glu Ile Ser Thr Pro Lys	
340 345 350	
ctc ggc aag ctg atc aac tgg gtc acg acg aca gac gaa tgc ccg gaa	1104
Leu Gly Lys Leu Ile Asn Trp Val Thr Thr Thr Asp Glu Cys Pro Glu	
355 360 365	
tgg aca ttc ggc acc gcg gcc ctc atg cgc aac ctg gcg aaa cgc ggg	1152
Trp Thr Phe Gly Thr Ala Ala Leu Met Arg Asn Leu Ala Lys Arg Gly	
370 375 380	
ctg ctc tag	1161
Leu Leu	
385	

<210> 86
<211> 386

<212> PRT

<213> *Agrobacterium tumefaciens*

<400> 86

```

Met Thr Gly Ser Arg Ser Ile Asp Ala Thr Ala Ile Leu Pro Glu Asp
1      5      10      15
Phe Glu Asn Ala Leu Leu Val Gly Arg Val Trp Ser Lys Ser Glu Gly
20      25      30
Gly Pro Cys Pro Val Leu Leu Lys Gly Gly Val Leu Tyr Asp Leu Thr
35      40      45
Ser Ile Ser Pro Thr Met Ser Glu Leu Leu Glu Lys Thr Asp Leu Val
50      55      60
Glu Leu Leu Ala Asp Thr Gly Thr Phe Val Ala Leu Gly Ala Leu Glu
65      70      75      80
Ala Phe Leu Asp Gly Ser Ala Gly Glu Leu Leu Ala Pro Asn Asp Ile
85      90      95
Gln Ala Val Lys Ala Ala Gly Val Thr Phe Ala Asp Ser Met Leu Glu
100     105     110
Arg Val Ile Glu Glu Gln Ala Lys Gly Asp Pro Leu Arg Ala Gln Glu
115     120     125
Ile Arg Gly Arg Leu Ala Pro Val Leu Gly Asp Asn Leu Lys Gly Leu
130     135     140
Glu Ala Gly Ser Glu Lys Ala Ala Glu Val Lys Lys Leu Leu Gln Glu
145     150     155     160
Met Gly Leu Trp Ser Gln Tyr Leu Glu Val Gly Ile Gly Pro Asp Ala
165     170     175
Glu Ile Phe Ser Lys Ala Gln Ala Met Ala Ser Val Gly Cys Gly Ala
180     185     190
Leu Ile Gly Val His Pro Lys Ser Gln Trp Asn Asn Pro Glu Pro Glu
195     200     205
Val Val Leu Ala Ile Ala Ser Asp Gly Arg Ile Val Gly Ala Thr Leu
210     215     220
Gly Asn Asp Val Asn Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu
225     230     235     240
Leu Ser Lys Ala Lys Asp Asn Asn Ala Ser Cys Ala Ile Gly Pro Phe
245     250     255
Ile Arg Leu Phe Asp Gly Arg Phe Thr Ile Glu Asp Val Lys Lys Ala
260     265     270
Gln Ile Ser Leu Leu Val Glu Gly Glu Asp Gly Phe Thr Met Thr Gly
275     280     285
Ala Ser Ala Met Gln Ala Ile Ser Arg Thr Pro Glu Asn Leu Ala Ser
290     295     300
Gln Leu Leu Asn Arg Asn His Gln Tyr Pro Asp Gly Ala Val Phe Phe
305     310     315     320
Leu Gly Thr Met Phe Ala Pro Val Lys Asp Arg His Gly Pro Gly Leu
325     330     335
Gly Phe Thr His Ser Lys Gly Asp Arg Val Glu Ile Ser Thr Pro Lys
340     345     350
Leu Gly Lys Leu Ile Asn Trp Val Thr Thr Thr Asp Glu Cys Pro Glu
355     360     365
Trp Thr Phe Gly Thr Ala Ala Leu Met Arg Asn Leu Ala Lys Arg Gly
370     375     380
Leu Leu
385

```

<210> 87

<211> 1161
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 87.
 atgacgggca gccgttctat tgatgcgacc gccatcctgc cggaagattt cgaaaacgcg 60
 ctgctgggtt gccgcgtttg gagtaaaagc gaaggcggtc cgtgcccggt gctgctgaaa 120
 ggcggtgttc tgtatgatct gaccagcatt agtccgacca tgagcgaact gctggaaaaa 180
 accgatctgg tggaactgct ggccgatacc ggcaaccttc ttgcgctggg tgccctggaa 240
 gcatttctgg atggtagcgc aggcgaactg ctggcgccga acgatatcca ggcagtgaaa 300
 gcggccggcg ttaccttcgc ggattctatg ctggaacgtg tgattgaaga acaggcgaaa 360
 ggtgatccgc tgcgtgcaca ggaaattcgc ggccgcctgg caccggtgct gggtgataat 420
 ctgaaaggtc tggaagcggg ctctgaaaaa gcagcggaa tgaaaaaact gctgcaggaa 480
 atgggcctgt ggagtcagta cctggaagtt ggcattggtc cggatgccga aatcttttct 540
 aaagcacagg cgatggccag tgtgggctgc ggtgcaactga ttggtgttca tccgaaaagt 600
 cagtggaaaca atccggaacc ggaagtgggt ctggcaattg cgagcgatgg tcgtatcgtg 660
 ggccgcgacgc tgggtaacga tgtaaatctg cgtgatttgc aaggccgcag cgccctgctg 720
 ctgtctaaag caaaagataa caatgogagt tgtgccattg gcccgtttat ccgtctgttc 780
 gatggtcgct ttaccattga agatgtgaaa aaagcccaga tctctctgct ggttgaaggt 840
 gaagatggct ttacgatgac cggtgccagt gcaatgcagg ccattagccg tacgccggaa 900
 aatctggcaa gccagctgct gaaccgcaat caccagtatc cggatggcgc ggtgtttttc 960
 ctgggcacca tgttcgcccc ggttaaagat cgtcatggcc cgggtctggg ctttacgcac 1020
 agcaaaggcg atcgctgga aatttctacc ccgaaactgg gtaaactgat caactgggtt 1080
 accacgaccg atgaatgtcc ggaatggacc tttggcaccg ccgcaactgat gcgtaaatctg 1140
 gcgaaacgcg gtctgctgta g 1161

<210> 88
 <211> 1200
 <212> ADN
 <213> Cupriavidus necator

<220>
 <221> CDS
 <222> (1)..(1200)

<400> 88
 atg gca ttg ccg ctc acg ctc agc gct acg caa acg ctt cct gcc gac 48
 Met Ala Leu Pro Leu Thr Leu Ser Ala Thr Gln Thr Leu Pro Ala Asp
 1 5 10 15
 ggg ctg gcg ggg acg ctg gtc ggc cgc gca tgg atc ccg gcc ggt gac 96
 Gly Leu Ala Gly Thr Leu Val Gly Arg Ala Trp Ile Pro Ala Gly Asp
 20 25 30
 ggc gtt ccg gcc ggc ccg gct gtc gtg gtc ttg cgg ccg gat ggt gta 144
 Gly Val Pro Ala Gly Pro Ala Val Val Val Leu Arg Pro Asp Gly Val
 35 40 45
 ttc gac att tcc gat gtc gcc ccg acc atg agc acc ctg ctc gaa cag 192
 Phe Asp Ile Ser Asp Val Ala Pro Thr Met Ser Thr Leu Leu Glu Gln
 50 55 60
 gac gac ccg ctc acg gtg gtg cac aat gcg ccc ggg cgc tgg atc gcc 240
 Asp Asp Pro Leu Thr Val Val His Asn Ala Pro Gly Arg Trp Ile Gly
 65 70 75 80
 aag ctg gac gac ctg ctg gcc aac acg gcc gat ccg cac ggc agc aac 288
 Lys Leu Asp Asp Leu Leu Ala Asn Thr Ala Asp Pro His Gly Ser Asn

																85																	90																	95																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
ggc	gtg	gcc	cgc	ctg	ctg	gcg	cca	tgc	gac	ctg	cag	gtg	atc	aag	gcg	336																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									

385

390

395

<210> 89

<211> 399

<212> PRT

<213> *Cupriavidus necator*

<400> 89

```

Met Ala Leu Pro Leu Thr Leu Ser Ala Thr Gln Thr Leu Pro Ala Asp
1      5      10      15
Gly Leu Ala Gly Thr Leu Val Gly Arg Ala Trp Ile Pro Ala Gly Asp
20     25     30
Gly Val Pro Ala Gly Pro Ala Val Val Val Leu Arg Pro Asp Gly Val
35     40     45
Phe Asp Ile Ser Asp Val Ala Pro Thr Met Ser Thr Leu Leu Glu Gln
50     55     60
Asp Asp Pro Leu Thr Val Val His Asn Ala Pro Gly Arg Trp Ile Gly
65     70     75     80
Lys Leu Asp Asp Leu Leu Ala Asn Thr Ala Asp Pro His Gly Ser Asn
85     90     95
Gly Val Ala Arg Leu Leu Ala Pro Cys Asp Leu Gln Val Ile Lys Ala
100    105    110
Ala Gly Val Thr Phe Ala Gly Ser Leu Val Glu Arg Val Ile Glu Glu
115    120    125
Gln Thr Lys Gly Asp Pro Gln Gly Ala Ala Glu Val Arg Asn Arg Ile
130    135    140
Gln Ala Leu Val Gly Glu Arg Leu Ser Arg Ile Arg Pro Gly Ser Arg
145    150    155    160
Glu Ala Gly Glu Leu Lys Ala Leu Leu Ile Glu His Gly Met Trp Ser
165    170    175
Gln Tyr Leu Glu Val Gly Ile Gly Pro Asp Ala Glu Ile Phe Thr Lys
180    185    190
Ala Pro Leu Leu Ser Ala Leu Gly Thr Gly Thr Glu Ile Gly Leu His
195    200    205
Pro Gly Ser Ala Trp Asn Asn Pro Glu Pro Glu Ile Val Leu Ala Ile
210    215    220
Asn Ser Arg Gly Asp Val Leu Gly Ala Thr Leu Gly Asn Asp Val Asn
225    230    235    240
Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu Leu Gly Lys Ala Lys
245    250    255
Asp Asn Asn Gly Ser Cys Ala Ile Gly Pro Phe Leu Arg Leu Phe Asp
260    265    270
Gln Ser Phe Ser Leu Asp Asp Val Arg Arg Ala Thr Val Asp Leu Arg
275    280    285
Val Asp Gly Leu Asp Gly Phe Val Leu Ser Gly Thr Ser Ser Met Asp
290    295    300
Gln Ile Thr Arg Asp Pro Leu Glu Leu Ala Glu Gln Ala Met Gly Ala
305    310    315    320
Thr His Gln Tyr Pro Asp Gly Ala Met Leu Phe Leu Gly Thr Leu Phe
325    330    335
Ala Pro Val Glu Asp Arg Asp Thr Ala Gly Gly Gly Phe Thr His Lys
340    345    350
Gly Gly Asp Leu Val Thr Ile Ser Ser Arg Gln Leu Gly Ser Leu Val
355    360    365
Asn Arg Val Gly Arg Ser Asp Arg Ile Ala Pro Trp Thr Phe Gly Val
370    375    380

```

Arg Ala Leu Met Ala Asn Leu Ala Ala Arg Gly His Thr Thr Phe
385 390 395

<210> 90
<211> 1200
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> ADN tổng hợp

<400> 90
atggcgctgc cgctgaccct gagcgccacc cagacgctgc cggccgatgg tctggccggc 60
accctggtgg gccgtgcatg gattccggcc ggtgatggcg ttccggccgg tccggccgtg 120
gttgtgctgc gtccggatgg tgttttcgat atcagtgatg tggcgccgac catgagcacg 180
ctgctggaac aggatgatcc gctgaccgtt gtgcataacg ccccggtcg ctggattggc 240
aaactggatg atctgctggc aaacacggcg gatccgcacg gttctaattg tgtggcacgt 300
ctgctggcac cgtgcgatct gcaggttatt aaagcggccg gcgtgacctt tgccggtagt 360
ctggttgaac gcgtgatcga agaacagacg aaaggtgatc cgcagggcgc agcgggaagt 420
cgtaaatcgt ttcaggcgct ggttggcgaa cgctgagcc gtatccgccc gggttctcgt 480
gaagccggtg aactgaaagc cctgctgatt gaacatggca tgtggagcca gtatctggaa 540
gtgggtattg gcccgatgc cgaaatcttc accaaagcac cgctgctgag tgcgctgggc 600
accggcacgg aaatcggtct gcacccgggc agcgcatgga acaatccgga accggaaatt 660
gttctggcca tcaactctcg cggatgatgt ctggggcgaa ccctgggtaa cgatgttaat 720
ctgcgcgatt tcgaaggccg tagcgccctg ctgctgggta aagcaaaaga taacaatggc 780
tcttgtgcaa ttggtccgtt tctgcgcctg ttcatcagc cttttagtct ggatgatgtg 840
cgtcgcgca ccgttgatct gcgtgtggat ggtctggatg gctttgttct gageggcacc 900
agctctatgg atcagatcac gcgtgatccg ctggaactgg ccgaacaggc aatgggtgcg 960
acccatcagt acccgatgg cgcatgctg ttctctgggta cgtgtttgc accggtggaa 1020
gatcgcgata ccgcgggcgg tggctttacg cacaaggtg gcgatctggt taccattagt 1080
agccgtcagc tgggctctct ggttaaccgt gtgggtcgca gtgatcgtat tgccccgtgg 1140
accttcggcg tgcgcgcct gatggcaaat ctggcgcac gtggtcatac cacgttttga 1200

<210> 91
<211> 1113
<212> ADN
<213> Pseudomonas elodea

<220>
<221> CDS
<222> (1)..(1113)

<400> 91
atg ctg cct gcc gat cat gcg cag gcg att ctg gtc ggc cgt gtg cag 48
Met Leu Pro Ala Asp His Ala Gln Ala Ile Leu Val Gly Arg Val Gln
1 5 10 15
acc ccg gcg ggc ccg agc ccc gtt ctc ctc cgc gat ggc cag gtg atc 96
Thr Pro Ala Gly Pro Ser Pro Val Leu Leu Arg Asp Gly Gln Val Ile
20 25 30
gac gtt tcg gcg atc gcg ccg acc gtc gcc gac ctg ctg gaa cgc gac 144
Asp Val Ser Ala Ile Ala Pro Thr Val Ala Asp Leu Leu Glu Arg Asp
35 40 45
gac atc gcg acg ctg agc ggc acg gtg ctg tgc agc gtc gac gcg ctc 192
Asp Ile Ala Thr Leu Ser Gly Thr Val Leu Cys Ser Val Asp Ala Leu
50 55 60

ggc acc gag tcg gcg ccg cag gtg ctg gct ccg gtc gac ctg cag tgc	240
Gly Thr Glu Ser Ala Pro Gln Val Leu Ala Pro Val Asp Leu Gln Cys	
65 70 75 80	
gtg aag gcc gcc ggc gtc acc ttc gcc gtc tcg gcg ctg gag cgc gtg	288
Val Lys Ala Ala Gly Val Thr Phe Ala Val Ser Ala Leu Glu Arg Val	
85 90 95	
atc gag gaa cgc gcc cgc ggc gat tcc gcc aag gcc gcc gag att cgc	336
Ile Glu Glu Arg Ala Arg Gly Asp Ser Ala Lys Ala Ala Glu Ile Arg	
100 105 110	
ggc gac ctc gaa gcc aag gtg ggt tcg ggc atc cgc tcg gtc gtc ccc	384
Gly Asp Leu Glu Ala Lys Val Gly Ser Gly Ile Arg Ser Val Val Pro	
115 120 125	
ggt acc gcc gag gcc gcg gcg ctc aag gcc gcg ctg atc gag gcg ggc	432
Gly Thr Ala Glu Ala Ala Ala Leu Lys Ala Ala Leu Ile Glu Ala Gly	
130 135 140	
atg tgg tcg caa tat ctc gaa gtg gcg atc ggc ccg gac gcg gag gtg	480
Met Trp Ser Gln Tyr Leu Glu Val Ala Ile Gly Pro Asp Ala Glu Val	
145 150 155 160	
ttc acc aag gcg ccg gtt ctg tcg gcg atg ggc tgg ggc gcc gag atc	528
Phe Thr Lys Ala Pro Val Leu Ser Ala Met Gly Trp Gly Ala Glu Ile	
165 170 175	
ggc atc cgc tcg gac agc gac tgg aac aat ccg gag ccg gaa gtg gtg	576
Gly Ile Arg Ser Asp Ser Asp Trp Asn Asn Pro Glu Pro Glu Val Val	
180 185 190	
ctg gtg gtc gac cgc aat ggt gcg atc aag ggc gcg acg ctc ggc aac	624
Leu Val Val Asp Arg Asn Gly Ala Ile Lys Gly Ala Thr Leu Gly Asn	
195 200 205	
gac gtc aac ctg cgc gac ttc gag ggc cgc agc gcg ctg ctg ctg ggc	672
Asp Val Asn Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu Leu Gly	
210 215 220	
aag gcg aag gac aac aat gcc tct acc gcg atc ggc ccg ttc atc cgc	720
Lys Ala Lys Asp Asn Asn Ala Ser Thr Ala Ile Gly Pro Phe Ile Arg	
225 230 235 240	
ctg ttc gat gac ggc ttc acg atg gac gac gtg cgt agc gcg gtg gtc	768
Leu Phe Asp Asp Gly Phe Thr Met Asp Asp Val Arg Ser Ala Val Val	
245 250 255	
gac ctc acc atc gac ggc ccg gag ggc tat cgc ctc tcg ggc acc aac	816
Asp Leu Thr Ile Asp Gly Pro Glu Gly Tyr Arg Leu Ser Gly Thr Asn	
260 265 270	
aag atg agc gag atc agc cga gat ccg acc gag ctc gtg cgc cag acg	864
Lys Met Ser Glu Ile Ser Arg Asp Pro Thr Glu Leu Val Arg Gln Thr	
275 280 285	
ctg agc gag cac cag tat ccg gac ggc ttc gcg ctg ttc ctc ggc acg	912
Leu Ser Glu His Gln Tyr Pro Asp Gly Phe Ala Leu Phe Leu Gly Thr	
290 295 300	
ctg ttc gcg ccg gtg cag gat cgc gac cat ccc ggc cgc ggc ttc act	960
Leu Phe Ala Pro Val Gln Asp Arg Asp His Pro Gly Arg Gly Phe Thr	
305 310 315 320	
cac aag ccc ggc gat att gtc cgc att tcc acg ccg aag ctc ggc acg	1008
His Lys Pro Gly Asp Ile Val Arg Ile Ser Thr Pro Lys Leu Gly Thr	
325 330 335	
ctc gtc aac cgc gtc acc acg tcc aag gcc gcc gcg ccc tgg acg ttc	1056
Leu Val Asn Arg Val Thr Thr Ser Lys Ala Ala Ala Pro Trp Thr Phe	
340 345 350	
ggc atc cgc gat ctg atg cgc aat ctc gcc gcc cgc ggc ctt ctc tcg	1104
Gly Ile Arg Asp Leu Met Arg Asn Leu Ala Ala Arg Gly Leu Leu Ser	
355 360 365	

cat tcc taa
His Ser
370

1113

<210> 92
<211> 370
<212> PRT
<213> Pseudomonas elodea

<400> 92
Met Leu Pro Ala Asp His Ala Gln Ala Ile Leu Val Gly Arg Val Gln
1 5 10 15
Thr Pro Ala Gly Pro Ser Pro Val Leu Arg Asp Gly Gln Val Ile
20 25 30
Asp Val Ser Ala Ile Ala Pro Thr Val Ala Asp Leu Leu Glu Arg Asp
35 40 45
Asp Ile Ala Thr Leu Ser Gly Thr Val Leu Cys Ser Val Asp Ala Leu
50 55 60
Gly Thr Glu Ser Ala Pro Gln Val Leu Ala Pro Val Asp Leu Gln Cys
65 70 75 80
Val Lys Ala Ala Gly Val Thr Phe Ala Val Ser Ala Leu Glu Arg Val
85 90 95
Ile Glu Glu Arg Ala Arg Gly Asp Ser Ala Lys Ala Ala Glu Ile Arg
100 105 110
Gly Asp Leu Glu Ala Lys Val Gly Ser Gly Ile Arg Ser Val Val Pro
115 120 125
Gly Thr Ala Glu Ala Ala Ala Leu Lys Ala Ala Leu Ile Glu Ala Gly
130 135 140
Met Trp Ser Gln Tyr Leu Glu Val Ala Ile Gly Pro Asp Ala Glu Val
145 150 155 160
Phe Thr Lys Ala Pro Val Leu Ser Ala Met Gly Trp Gly Ala Glu Ile
165 170 175
Gly Ile Arg Ser Asp Ser Asp Trp Asn Asn Pro Glu Pro Glu Val Val
180 185 190
Leu Val Val Asp Arg Asn Gly Ala Ile Lys Gly Ala Thr Leu Gly Asn
195 200 205
Asp Val Asn Leu Arg Asp Phe Glu Gly Arg Ser Ala Leu Leu Leu Gly
210 215 220
Lys Ala Lys Asp Asn Asn Ala Ser Thr Ala Ile Gly Pro Phe Ile Arg
225 230 235 240
Leu Phe Asp Asp Gly Phe Thr Met Asp Asp Val Arg Ser Ala Val Val
245 250 255
Asp Leu Thr Ile Asp Gly Pro Glu Gly Tyr Arg Leu Ser Gly Thr Asn
260 265 270
Lys Met Ser Glu Ile Ser Arg Asp Pro Thr Glu Leu Val Arg Gln Thr
275 280 285
Leu Ser Glu His Gln Tyr Pro Asp Gly Phe Ala Leu Phe Leu Gly Thr
290 295 300
Leu Phe Ala Pro Val Gln Asp Arg Asp His Pro Gly Arg Gly Phe Thr
305 310 315 320
His Lys Pro Gly Asp Ile Val Arg Ile Ser Thr Pro Lys Leu Gly Thr
325 330 335
Leu Val Asn Arg Val Thr Thr Ser Lys Ala Ala Ala Pro Trp Thr Phe
340 345 350
Gly Ile Arg Asp Leu Met Arg Asn Leu Ala Ala Arg Gly Leu Leu Ser
355 360 365

His Ser
370

<210> 93
<211> 1113
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> ADN tổng hợp

<400> 93
atgctgccgg cggatcatgc ccaagcaatc ctggctgggc ggcgtccaaac cccggctggc 60
cgcgtcccgccg tccgtgctgcg tgatgggcag gtcattgatg tgtccgcaat cgcaccgacc 120
gtggcagacc tgcctggaacg tgatgacatt gccaccctga gcggtaacgg gctgtgctct 180
gttgatgccc tgggcacgga aagcgcaccg caggctcctgg caccgggtgga cctgcaatgt 240
gtcaaagcgg ccgggtgttac cttcgcagtc tcagctctgg aacgcgtgat cgaagaacgt 300
gcacgcgggtg attcggctaa agcagctgaa attcgtggcg acctggaagc gaaagtgggc 360
tcaggtatcc gctcgggtggg tccgggcacg gcagaagcgg ccgcactgaa agctgcgctg 420
attgaagcgg gcatgtgggc acagtatctg gaagttgcga tcggcccgga tgccgaagtg 480
ttcaccagg caccgggtctt gtcggctatg ggctgggggtg cggaaattgg tatccgtagc 540
gattctgact ggaacaatcc ggaaccggaa gtcgtgctgg ttgtcgatcg caacggcgca 600
attaaagggtg ctacgctggg caacgatgtt aatctgcgtg actttgaagg tcgcagtgcc 660
ctgctgctgg gcaaagcaaa ggataacaat gcgtccaccg ccattgggtc gtttatccgt 720
ctgttcgatg acggctttac catggatgac gtgcgcagtg ccgtgggtga tctgacgatt 780
gacggcccgga aagggttatcg tctgtccggc accaacaaga tgagtgaat ctcccgatg 840
ccgaccgaac tgggttcgca gacgtgagc gaacatcaat acccggtgg ttttgctctg 900
ttcctgggca cgtgttcgc accggttcaa gatcgtgacc atccggggccg cggttttacc 960
cacaaccgg gtgatattgt ccgtatcagc acccogaagc tgggcacgct ggttaatcgc 1020
gtcaccacgt ctaaagccgc agctccgtgg accttcggta ttcgtgacct gatgcgtaat 1080
ctggcggtc gtggcctgct gtctcattcg tga 1113

<210> 94
<211> 1194
<212> ADN
<213> *Zobellia galactanivorans*

<220>
<221> CDS
<222> (1)..(1194)

<400> 94
atg aat tat atc gat tta aac cgc ctc ata ccg gaa tta gcg gaa aca 48
Met Asn Tyr Ile Asp Leu Asn Arg Leu Ile Pro Glu Leu Ala Glu Thr
1 5 10 15
ggc aca tgg ata ggc agg tgc atg gta ccg gca caa aaa gcg tac aat 96
Gly Thr Trp Ile Gly Arg Cys Met Val Pro Ala Gln Lys Ala Tyr Asn
20 25 30
ggg ata gct ggg ccc cac gtg gtg atg gcc cgc aaa gga aaa atc tac 144
Gly Ile Ala Gly Pro His Val Val Met Ala Arg Lys Gly Lys Ile Tyr
35 40 45
gac cta tcg gcc cac ttc aac tcc aca agt gag cta ttc aac tgt aag 192
Asp Leu Ser Ala His Phe Asn Ser Thr Ser Glu Leu Phe Asn Cys Lys
50 55 60
aac ccg gtt tcc cgc tta aaa gcc tta aac gat tta ccg gaa cta ggc 240

Asn	Pro	Val	Ser	Arg	Leu	Lys	Ala	Leu	Asn	Asp	Leu	Pro	Glu	Leu	Gly	
65					70				75						80	
agc	ctg	aaa	gat	gcc	ttg	aaa	aac	gcc	cta	tac	ttc	aat	cag	aat	cca	288
Ser	Leu	Lys	Asp	Ala	Leu	Lys	Asn	Ala	Leu	Tyr	Phe	Asn	Gln	Asn	Pro	
			85						90					95		
cta	ctc	ccc	tac	atc	ata	gcg	cca	aac	gat	ata	caa	gcc	gta	aaa	gcc	336
Leu	Leu	Pro	Tyr	Ile	Ile	Ala	Pro	Asn	Asp	Ile	Gln	Ala	Val	Lys	Ala	
			100					105					110			
tgt	ggt	gta	act	ttt	atc	aaa	agc	ttg	ctc	gag	agg	gtc	att	gaa	gaa	384
Cys	Gly	Val	Thr	Phe	Ile	Lys	Ser	Leu	Leu	Glu	Arg	Val	Ile	Glu	Glu	
		115					120					125				
aaa	gca	aaa	ggc	gat	gcc	cta	gta	gcg	aac	gac	ata	cgt	caa	acc	att	432
Lys	Ala	Lys	Gly	Asp	Ala	Leu	Val	Ala	Asn	Asp	Ile	Arg	Gln	Thr	Ile	
	130					135					140					
tac	gat	acc	tta	ggc	aac	gac	ttg	agc	aag	gtt	acc	ccc	ggt	tca	ccg	480
Tyr	Asp	Thr	Leu	Gly	Asn	Asp	Leu	Ser	Lys	Val	Thr	Pro	Gly	Ser	Pro	
145					150					155					160	
gaa	acc	gaa	aag	ctc	aag	gag	gaa	tta	cag	aaa	aaa	ggc	cta	tgg	tcg	528
Glu	Thr	Glu	Lys	Leu	Lys	Glu	Glu	Leu	Gln	Lys	Lys	Gly	Leu	Trp	Ser	
			165						170					175		
caa	tac	ctc	gaa	gtg	ggc	att	ggc	aag	gat	gcc	gag	gtt	ttc	acc	aag	576
Gln	Tyr	Leu	Glu	Val	Gly	Ile	Gly	Lys	Asp	Ala	Glu	Val	Phe	Thr	Lys	
			180					185					190			
gcc	caa	ccc	ttg	tct	gcc	gtt	ggt	ttc	gga	gcg	gag	ata	ggc	gtt	cta	624
Ala	Gln	Pro	Leu	Ser	Ala	Val	Gly	Phe	Gly	Ala	Glu	Ile	Gly	Val	Leu	
		195					200					205				
aag	agt	tcc	aaa	tgg	aac	aat	ccg	gaa	cct	gaa	atc	gtg	ctc	gcc	gta	672
Lys	Ser	Ser	Lys	Trp	Asn	Asn	Pro	Glu	Pro	Glu	Ile	Val	Leu	Ala	Val	
	210				215							220				
tcc	tcc	tcg	gga	aaa	atc	gtc	ggg	gcc	aca	ctc	ggc	aac	gac	gta	aac	720
Ser	Ser	Ser	Gly	Lys	Ile	Val	Gly	Ala	Thr	Leu	Gly	Asn	Asp	Val	Asn	
				230						235				240		
ctt	cgc	gat	tac	gaa	ggt	cgc	agc	gcc	cta	ttg	ctc	ggt	gag	gcg	aaa	768
Leu	Arg	Asp	Tyr	Glu	Gly	Arg	Ser	Ala	Leu	Leu	Leu	Gly	Glu	Ala	Lys	
			245						250					255		
gac	caa	aac	gga	tct	tgt	gcc	att	ggc	ccc	ttg	ttt	cgc	ttg	ttc	gat	816
Asp	Gln	Asn	Gly	Ser	Cys	Ala	Ile	Gly	Pro	Leu	Phe	Arg	Leu	Phe	Asp	
			260					265					270			
gag	act	ttt	tcc	cta	gat	gac	gta	aag	gat	tgc	gat	gta	atg	ttt	tct	864
Glu	Thr	Phe	Ser	Leu	Asp	Asp	Val	Lys	Asp	Cys	Asp	Val	Met	Phe	Ser	
		275					280					285				
atg	aag	ggg	aag	gac	aat	ttc	gcc	act	tcc	gga	agc	aat	aag	atg	aaa	912
Met	Lys	Gly	Lys	Asp	Asn	Phe	Ala	Thr	Ser	Gly	Ser	Asn	Lys	Met	Lys	
	290				295						300					
gag	att	agc	cgt	agt	cct	gaa	aat	tta	gtg	gcc	cag	gtc	ata	gga	aaa	960
Glu	Ile	Ser	Arg	Ser	Pro	Glu	Asn	Leu	Val	Ala	Gln	Val	Ile	Gly	Lys	
	305				310					315				320		
aac	cat	caa	tac	cca	gac	ggg	ctc	gtg	ctt	ttc	tta	ggt	act	atg	ttc	1008
Asn	His	Gln	Tyr	Pro	Asp	Gly	Leu	Val	Leu	Phe	Leu	Gly	Thr	Met	Phe	
			325						330					335		
gcc	cct	acg	gaa	gac	cgc	aat	gga	aaa	ggt	ttg	ggt	ttc	acc	cac	aaa	1056
Ala	Pro	Thr	Glu	Asp	Arg	Asn	Gly	Lys	Gly	Leu	Gly	Phe	Thr	His	Lys	
			340				345						350			
aag	gga	gat	caa	gtc	aac	atc	tct	tct	tcg	cat	ttg	ggc	aca	ctc	atc	1104
Lys	Gly	Asp	Gln	Val	Asn	Ile	Ser	Ser	Ser	His	Leu	Gly	Thr	Leu	Ile	
		355				360						365				
aat	tgg	gtg	aac	act	tgt	gac	caa	ata	cca	aag	tgg	gaa	ttc	ggc	att	1152

Asn Trp Val Asn Thr Cys Asp Gln Ile Pro Lys Trp Glu Phe Gly Ile
 370 375 380
 ggt gct ttt acg aac tat att gta aaa cgg aat tta aaa tag
 Gly Ala Phe Thr Asn Tyr Ile Val Lys Arg Asn Leu Lys
 385 390 395

1194

<210> 95
 <211> 397
 <212> PRT
 <213> Zobellia galactanivorans

<400> 95
 Met Asn Tyr Ile Asp Leu Asn Arg Leu Ile Pro Glu Leu Ala Glu Thr
 1 5 10 15
 Gly Thr Trp Ile Gly Arg Cys Met Val Pro Ala Gln Lys Ala Tyr Asn
 20 25 30
 Gly Ile Ala Gly Pro His Val Val Met Ala Arg Lys Gly Lys Ile Tyr
 35 40 45
 Asp Leu Ser Ala His Phe Asn Ser Thr Ser Glu Leu Phe Asn Cys Lys
 50 55 60
 Asn Pro Val Ser Arg Leu Lys Ala Leu Asn Asp Leu Pro Glu Leu Gly
 65 70 75 80
 Ser Leu Lys Asp Ala Leu Lys Asn Ala Leu Tyr Phe Asn Gln Asn Pro
 85 90 95
 Leu Leu Pro Tyr Ile Ile Ala Pro Asn Asp Ile Gln Ala Val Lys Ala
 100 105 110
 Cys Gly Val Thr Phe Ile Lys Ser Leu Leu Glu Arg Val Ile Glu Glu
 115 120 125
 Lys Ala Lys Gly Asp Ala Leu Val Ala Asn Asp Ile Arg Gln Thr Ile
 130 135 140
 Tyr Asp Thr Leu Gly Asn Asp Leu Ser Lys Val Thr Pro Gly Ser Pro
 145 150 155 160
 Glu Thr Glu Lys Leu Lys Glu Glu Leu Gln Lys Lys Gly Leu Trp Ser
 165 170 175
 Gln Tyr Leu Glu Val Gly Ile Gly Lys Asp Ala Glu Val Phe Thr Lys
 180 185 190
 Ala Gln Pro Leu Ser Ala Val Gly Phe Gly Ala Glu Ile Gly Val Leu
 195 200 205
 Lys Ser Ser Lys Trp Asn Asn Pro Glu Pro Glu Ile Val Leu Ala Val
 210 215 220
 Ser Ser Ser Gly Lys Ile Val Gly Ala Thr Leu Gly Asn Asp Val Asn
 225 230 235 240
 Leu Arg Asp Tyr Glu Gly Arg Ser Ala Leu Leu Leu Gly Glu Ala Lys
 245 250 255
 Asp Gln Asn Gly Ser Cys Ala Ile Gly Pro Leu Phe Arg Leu Phe Asp
 260 265 270
 Glu Thr Phe Ser Leu Asp Asp Val Lys Asp Cys Asp Val Met Phe Ser
 275 280 285
 Met Lys Gly Lys Asp Asn Phe Ala Thr Ser Gly Ser Asn Lys Met Lys
 290 295 300
 Glu Ile Ser Arg Ser Pro Glu Asn Leu Val Ala Gln Val Ile Gly Lys
 305 310 315 320
 Asn His Gln Tyr Pro Asp Gly Leu Val Leu Phe Leu Gly Thr Met Phe
 325 330 335
 Ala Pro Thr Glu Asp Arg Asn Gly Lys Gly Leu Gly Phe Thr His Lys
 340 345 350

Lys Gly Asp Gln Val Asn Ile Ser Ser Ser His Leu Gly Thr Leu Ile
 355 360 365
 Asn Trp Val Asn Thr Cys Asp Gln Ile Pro Lys Trp Glu Phe Gly Ile
 370 375 380
 Gly Ala Phe Thr Asn Tyr Ile Val Lys Arg Asn Leu Lys
 385 390 395

<210> 96
 <211> 1194
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 96
 atgaactata ttgatctgaa tcgtctgata ccggaactgg cggaaaccgg tacgtggatt 60
 ggccgctgca tgggtgcggc ccagaaagcc tataacggta tcgccggccc gcatgtggtt 120
 atggcacgta aaggcaaaat ttacgatctg agcgccact ttaatagtag cagcgaactg 180
 ttcaactgta aaaatccggg tagccgcctg aaagcactga acgatctgcc ggaactgggt 240
 tctctgaaag atgcgctgaa aaatgccctg tattttaacc agaatccgct gctgccgtac 300
 attatcgcg cgaacgatata tcaggcagtg aaagcgtgcg gcgttacctt catcaaaagc 360
 ctgctggaac gtgtgattga agaaaaagcc aaaggtgatg ccctggttgc aaacgatatt 420
 cgccagacca tctatgatac gctgggcaat gatctgagta aagtgacccc gggtagcccg 480
 gaaacggaaa aactgaaaga agaactgcag aaaaaaggcc tgtggtctca gtacctggaa 540
 gtgggcatcg gtaaagatgc cgaagttttt accaaagcac agccgctgag cgcggtgggt 600
 tttggtgcag aaattgggtg tctgaaaagc tctaaatgga acaatccgga accggaaatc 660
 gtgctggcgg ttagtagctc tggtaaaatt gtgggcgcca ccctgggtaa cgatgttaat 720
 ctgcgtgatt acgaaggccg cagtgcactg ctgctgggtg aagcgaaga tcagaatggc 780
 agctgcgcga ttggtccgct gtttcgtctg ttogatgaaa ctttttctct ggatgatgtg 840
 aaggattgtg atgttatgtt cagtatgaag ggcaaggata acttcgcaac gtctggtagt 900
 aacaagatga aggaaatcag ccgttctccg gaaaacctgg tggcgaggt tattggcaaa 960
 aatcatcagt atccgatgg cctggtgctg tttctgggca ccatgttcgc accgacggaa 1020
 gatcgcaacg gcaaaggtct gggttttacc cataaaaaag gcgatcaagt gaatatcagt 1080
 agctctcacc tgggcacccg gattaactgg gtttaatacgt gtgatcagat tccgaaatgg 1140
 gaatttggtg tcggcgcgtt cacgaactac attgttaaac gcaatctgaa atag 1194

<210> 97
 <211> 897
 <212> ADN
 <213> Thermobacillus composti

<220>
 <221> CDS
 <222> (1)..(897)

<400> 97
 atg cgc att att cgt tac atc ggg gac gac gga gct gcg cgc ctc gcg 48
 Met Arg Ile Ile Arg Tyr Ile Gly Asp Asp Gly Ala Ala Arg Leu Ala
 1 5 10 15
 gcg gtg acc gac gag gag cag gct ttc ccg ctc cgg tcg ccg gat ttc 96
 Ala Val Thr Asp Glu Glu Gln Ala Phe Pro Leu Arg Ser Pro Asp Phe
 20 25 30
 atg gcg ctg gtg cgg gaa gcg gac gag gcg ggg atc acg ccg ctg gaa 144
 Met Ala Leu Val Arg Glu Ala Asp Glu Ala Gly Ile Thr Pro Leu Glu

35	40	45	
gcg gtg cgc cgt cag atc gcg ggc gca cag ccg ctg ccc ggc gac tgg			192
Ala Val Arg Arg Gln Ile Ala Gly Ala Gln Pro Leu Pro Gly Asp Trp			
50	55	60	
cgc gag ttg aac ctg ctg acg ccg gtg gat gcg ccc gag gtg tgg gcg			240
Arg Glu Leu Asn Leu Leu Thr Pro Val Asp Ala Pro Glu Val Trp Ala			
65	70	75	80
gcc ggc gtc act tat gag cgc agc aaa gaa gcg cgc aac gag gag tcg			288
Ala Gly Val Thr Tyr Glu Arg Ser Lys Glu Ala Arg Asn Glu Glu Ser			
85	90	95	
aag ggc gcc gca acc ggc gac gaa acg ttc tac gac aag gtg tac cgg			336
Lys Gly Ala Ala Thr Gly Asp Glu Thr Phe Tyr Asp Lys Val Tyr Arg			
100	105	110	
gcc gag cgg ccg gag att ttc ttc aag tcg acg agc gcc cgc acg gcg			384
Ala Glu Arg Pro Glu Ile Phe Phe Lys Ser Thr Ser Ala Arg Thr Ala			
115	120	125	
cgg ccg ggc aca ccc gtc tgc atc cgc agc gac tcg gac tgg cag gtg			432
Arg Pro Gly Thr Pro Val Cys Ile Arg Ser Asp Ser Asp Trp Gln Val			
130	135	140	
ccc gag ccg gaa ctc ggc atc gtg ctg gac cgc ggc ggc cgc atc ctc			480
Pro Glu Pro Glu Leu Gly Ile Val Leu Asp Arg Gly Gly Arg Ile Leu			
145	150	155	160
ggc tac acg gtc ggc aac gac atg agc tgc cgc gac atc gag ggc gag			528
Gly Tyr Thr Val Gly Asn Asp Met Ser Cys Arg Asp Ile Glu Gly Glu			
165	170	175	
aac ccg ctc tat ctg ccg cag gcg aaa atc tgg cgc cgc tcc tgt tcc			576
Asn Pro Leu Tyr Leu Pro Gln Ala Lys Ile Trp Arg Arg Ser Cys Ser			
180	185	190	
atc ggt ccg gcg atc ccg ctg gcc gaa acc gtg ccg aat ccg tat gac			624
Ile Gly Pro Ala Ile Arg Leu Ala Glu Thr Val Pro Asn Pro Tyr Asp			
195	200	205	
ctg acg atc acc tgc cgc atc tat ccg gac ggt caa ttg gcg gtg aac			672
Leu Thr Ile Thr Cys Arg Ile Tyr Arg Asp Gly Gln Leu Ala Val Asn			
210	215	220	
gag acg gcc aat acg ggg cag ctc ccg ccg aag ctg gac gaa ctg gcg			720
Glu Thr Ala Asn Thr Gly Gln Leu Arg Arg Lys Leu Asp Glu Leu Ala			
225	230	235	240
tcc ttc ctc gtg ccg gac aac gtc gtg ttc gac ggc acc gtg ctg ctg			768
Ser Phe Leu Val Arg Asp Asn Val Val Phe Asp Gly Thr Val Leu Leu			
245	250	255	
acc ggc acc tgc atc gtg ccg ccc gac ccg ttc acg ctg cag ccc ggc			816
Thr Gly Thr Cys Ile Val Pro Pro Asp Arg Phe Thr Leu Gln Pro Gly			
260	265	270	
gac cgc atc gag atc gac att tcc ggc atc ggc acc ctg atc aac cct			864
Asp Arg Ile Glu Ile Asp Ile Ser Gly Ile Gly Thr Leu Ile Asn Pro			
275	280	285	
gtc gcg gcg gcg gac gcc gca atc cag gat taa			897
Val Ala Ala Ala Asp Ala Ala Ile Gln Asp			
290	295		

<210> 98

<211> 298

<212> PRT

<213> Thermobacillus composti

<400> 98

Met Arg Ile Ile Arg Tyr Ile Gly Asp Asp Gly Ala Ala Arg Leu Ala
 1 5 10 15
 Ala Val Thr Asp Glu Glu Gln Ala Phe Pro Leu Arg Ser Pro Asp Phe
 20 25 30
 Met Ala Leu Val Arg Glu Ala Asp Glu Ala Gly Ile Thr Pro Leu Glu
 35 40 45
 Ala Val Arg Arg Gln Ile Ala Gly Ala Gln Pro Leu Pro Gly Asp Trp
 50 55 60
 Arg Glu Leu Asn Leu Leu Thr Pro Val Asp Ala Pro Glu Val Trp Ala
 65 70 75 80
 Ala Gly Val Thr Tyr Glu Arg Ser Lys Glu Ala Arg Asn Glu Glu Ser
 85 90 95
 Lys Gly Ala Ala Thr Gly Asp Glu Thr Phe Tyr Asp Lys Val Tyr Arg
 100 105 110
 Ala Glu Arg Pro Glu Ile Phe Phe Lys Ser Thr Ser Ala Arg Thr Ala
 115 120 125
 Arg Pro Gly Thr Pro Val Cys Ile Arg Ser Asp Ser Asp Trp Gln Val
 130 135 140
 Pro Glu Pro Glu Leu Gly Ile Val Leu Asp Arg Gly Gly Arg Ile Leu
 145 150 155 160
 Gly Tyr Thr Val Gly Asn Asp Met Ser Cys Arg Asp Ile Glu Gly Glu
 165 170 175
 Asn Pro Leu Tyr Leu Pro Gln Ala Lys Ile Trp Arg Arg Ser Cys Ser
 180 185 190
 Ile Gly Pro Ala Ile Arg Leu Ala Glu Thr Val Pro Asn Pro Tyr Asp
 195 200 205
 Leu Thr Ile Thr Cys Arg Ile Tyr Arg Asp Gly Gln Leu Ala Val Asn
 210 215 220
 Glu Thr Ala Asn Thr Gly Gln Leu Arg Arg Lys Leu Asp Glu Leu Ala
 225 230 235 240
 Ser Phe Leu Val Arg Asp Asn Val Val Phe Asp Gly Thr Val Leu Leu
 245 250 255
 Thr Gly Thr Cys Ile Val Pro Pro Asp Arg Phe Thr Leu Gln Pro Gly
 260 265 270
 Asp Arg Ile Glu Ile Asp Ile Ser Gly Ile Gly Thr Leu Ile Asn Pro
 275 280 285
 Val Ala Ala Ala Asp Ala Ala Ile Gln Asp
 290 295

<210> 99

<211> 897

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> ADN tổng hợp

<400> 99

atgcgtatta tccgctatat tggatgat ggtgccgcac gtctggcagc cggtaccgat 60
 gaagaacagg catttccgct gcgttctccg gatttcatgg cgctgggttcg cgaagcggat 120
 gaagcgggta tcacgcgcgt ggaagcggtg cgtcgccaga ttgcgggtgc acagccgctg 180
 ccgggtgatt ggccggaact gaacctgctg accccgggtg atgcaccgga agtttgggca 240
 gcagggtgtg cgtacgaacg tagcaaagaa gcacgcaatg aagaatctaa aggcgcggcc 300
 accgggtgatg aaacctttta tgataaagtt taccgtgcgg aacgcccggga aatctttttc 360
 aaaagcacct ctgcacgtac cgcccgtccg ggcaccccgg tgtgcattcg tagtgatagc 420
 gattggcagg ttccggaacc ggaactgggt atcgtgctgg atcgtggcgg tcgcattctg 480

```

ggctataaccg tgggtaacga tatgagctgc cgtgatatcg aaggcgaaaa tccgctgtac 540
ctgccgcagg ccaaaatttg gcgtogctot tgtagtatcg gtccggccat tcgcctggca 600
gaaaccgttc cgaaccogta tgatctgacc atcacgtgcc gtatttaccg cgatggccag 660
ctggcggtga acgaaaccgc caatacgggt cagctgcgtc gcaaactgga tgaactggcg 720
agttttctgg ttccgcgataa cgtgggttttc gatggcaccg ttctgctgac cggtagctgc 780
atcgtgccgc cggatcgttt caccctgcag ccgggcgata gcattgaaat cgatattagc 840
ggcatcggca ccctgattaa tccggtggca gcggccgatg cagcgattca ggattaa 897

```

```

<210> 100
<211> 1227
<212> ADN
<213> Arthrobacter globiformis

```

```

<220>
<221> CDS
<222> (1)..(1227)

```

```

<400> 100
gtg acc gcg ccg ccg att ccc ggg agt tcc gtt cct ccg gta acc tct 48
Val Thr Ala Pro Pro Ile Pro Gly Ser Ser Val Pro Pro Val Thr Ser
1 5 10 15
gtc ctt ccg gat gac gct ggc cag gcg ctg ctg gtt ggc ccg atc tgg 96
Val Leu Pro Asp Asp Ala Gly Gln Ala Leu Leu Val Gly Arg Ile Trp
20 25 30
gac ccc gcc acc ggc ggt ccc ccg gtg gtg gct ctc agc ggg gac tcc 144
Asp Pro Ala Thr Gly Gly Pro Arg Val Val Ala Leu Ser Gly Asp Ser
35 40 45
gcc gtc gac ctg acg cgc ctg gcc gga acc gtc tcg gaa ctg ctc gaa 192
Ala Val Asp Leu Thr Arg Leu Ala Gly Thr Val Ser Glu Leu Leu Glu
50 55 60
ctg cct gat cca gcc gcg gca gtc cgt gcg gcg ctc ctt gac ccg gca 240
Leu Pro Asp Pro Ala Ala Val Arg Ala Ala Leu Leu Asp Pro Ala
65 70 75 80
ctg ggg gta cag ccg tgg gcg acg gcc gac gtc gtc gcc gcc tcg ctg 288
Leu Gly Val Gln Arg Trp Ala Thr Ala Asp Val Val Ala Ala Ser Leu
85 90 95
gcc ggt gac gca gca cgc ccg cac ctg ctc gca ccg gtt gac ctg cag 336
Ala Gly Asp Ala Ala Arg Pro His Leu Leu Ala Pro Val Asp Leu Gln
100 105 110
gtc atc aag gcc tgc ggc gtg acg ttt gtg gac agc atg atc gaa ccg 384
Val Ile Lys Ala Cys Gly Val Thr Phe Val Asp Ser Met Ile Glu Arg
115 120 125
gtg atc gag gaa agg tgc gcc ggc gac gcc gct ccg gcg gcc gag atg 432
Val Ile Glu Glu Arg Cys Ala Gly Asp Ala Ala Arg Ala Ala Glu Met
130 135 140
cgc gag ctt gtc ggg aag gcg ctc ggc gga agc atc gcc acg gtc ccg 480
Arg Glu Leu Val Gly Lys Ala Leu Gly Gly Ser Ile Ala Thr Val Arg
145 150 155 160
ccg ggg tcc ccg gag gcg gcc gag gcc aag cgc gtc ctg att gcc gag 528
Pro Gly Ser Pro Glu Ala Ala Glu Ala Lys Arg Val Leu Ile Ala Glu
165 170 175
ggg ctc tgg tcg cag tat ctc gag gtg ggc atc ggg ccg gac ccg gag 576
Gly Leu Trp Ser Gln Tyr Leu Glu Val Gly Ile Gly Pro Asp Pro Glu
180 185 190
gtg ttc acg aag gcg ccg gtg ctg tcc tcg gtg ggc ctg gcc gcg gcc 624
Val Phe Thr Lys Ala Pro Val Leu Ser Ser Val Gly Leu Gly Ala Gly

```

```

      195              200              205
atc ggg att ccg cgg ttc tcc tcc tgg aac aac ccg gag ccg gaa ctg      672
Ile Gly Ile Pro Arg Phe Ser Ser Trp Asn Asn Pro Glu Pro Glu Leu
      210              215              220
gtg ctg atc gtg acg tcc cgc ggc gag gtg gtg ggc gcc act ctg ggc      720
Val Leu Ile Val Thr Ser Arg Gly Glu Val Val Gly Ala Thr Leu Gly
      225              230              235              240
aac gac gtc aac ctg cgc gac gtt gag ggc cgc agc gcc ctg ctg ctg      768
Asn Asp Val Asn Leu Arg Asp Val Glu Gly Arg Ser Ala Leu Leu Leu
      245              250              255
ggc aag gcc aag gac aac aac gcg tcc agt gcc ctc gga cca ctg atc      816
Gly Lys Ala Lys Asp Asn Asn Ala Ser Ser Ala Leu Gly Pro Leu Ile
      260              265              270
agg ctc ttc gac ggc agc ttc acg gtg gac acc ctc cgt gag gag gag      864
Arg Leu Phe Asp Gly Ser Phe Thr Val Asp Thr Leu Arg Glu Glu Glu
      275              280              285
atc ctg ctg cgt gtc gaa ggc ctg gac ggc tac ctg ctg gag ggg cgg      912
Ile Leu Leu Arg Val Glu Gly Leu Asp Gly Tyr Leu Leu Glu Gly Arg
      290              295              300
aac acg ctg gcg agg atc agc agg ccc ttc gag gag ctc gtg gcg gcc      960
Asn Thr Leu Ala Arg Ile Ser Arg Pro Phe Glu Glu Leu Val Ala Ala
      305              310              315              320
acc cgc ggc agg cac cac cag tat ccg gac ggc ttc gca ctg ttc acc      1008
Thr Arg Gly Arg His His Gln Tyr Pro Asp Gly Phe Ala Leu Phe Thr
      325              330              335
ggc acg ctt ttc gca ccg acc cag gac cgc gac gag ccg ggg cag ggg      1056
Gly Thr Leu Phe Ala Pro Thr Gln Asp Arg Asp Glu Pro Gly Gln Gly
      340              345              350
ttc acg cac aag cac ggg gat gtt gtg acc atc cgg agc cgc cac ctc      1104
Phe Thr His Lys His Gly Asp Val Val Thr Ile Arg Ser Arg His Leu
      355              360              365
ggt gcc ctc atc aac cgg gtg ggc acg gca gag gag ctc ccg gag tgg      1152
Gly Ala Leu Ile Asn Arg Val Gly Thr Ala Glu Glu Leu Pro Glu Trp
      370              375              380
acc ttc ggc ctg cgg cag ctg ttc ggc tac ctc gcg gag cag cgg cag      1200
Thr Phe Gly Leu Arg Gln Leu Phe Gly Tyr Leu Ala Glu Gln Arg Gln
      385              390              395              400
gcg gag cta gct cag atg cag gag tag      1227
Ala Glu Leu Ala Gln Met Gln Glu
      405

```

<210> 101

<211> 408

<212> PRT

<213> Arthrobacter globiformis

<400> 101

```

Val Thr Ala Pro Pro Ile Pro Gly Ser Ser Val Pro Pro Val Thr Ser
1      5      10      15
Val Leu Pro Asp Asp Ala Gly Gln Ala Leu Leu Val Gly Arg Ile Trp
      20      25      30
Asp Pro Ala Thr Gly Gly Pro Arg Val Val Ala Leu Ser Gly Asp Ser
      35      40      45
Ala Val Asp Leu Thr Arg Leu Ala Gly Thr Val Ser Glu Leu Leu Glu
      50      55      60
Leu Pro Asp Pro Ala Ala Ala Val Arg Ala Ala Leu Leu Asp Pro Ala

```

65					70					75				80	
Leu	Gly	Val	Gln	Arg	Trp	Ala	Thr	Ala	Asp	Val	Val	Ala	Ala	Ser	Leu
				85					90					95	
Ala	Gly	Asp	Ala	Ala	Arg	Pro	His	Leu	Leu	Ala	Pro	Val	Asp	Leu	Gln
			100					105					110		
Val	Ile	Lys	Ala	Cys	Gly	Val	Thr	Phe	Val	Asp	Ser	Met	Ile	Glu	Arg
		115					120					125			
Val	Ile	Glu	Glu	Arg	Cys	Ala	Gly	Asp	Ala	Ala	Arg	Ala	Ala	Glu	Met
	130					135					140				
Arg	Glu	Leu	Val	Gly	Lys	Ala	Leu	Gly	Gly	Ser	Ile	Ala	Thr	Val	Arg
145					150					155					160
Pro	Gly	Ser	Pro	Glu	Ala	Ala	Glu	Ala	Lys	Arg	Val	Leu	Ile	Ala	Glu
				165					170					175	
Gly	Leu	Trp	Ser	Gln	Tyr	Leu	Glu	Val	Gly	Ile	Gly	Pro	Asp	Pro	Glu
			180					185					190		
Val	Phe	Thr	Lys	Ala	Pro	Val	Leu	Ser	Ser	Val	Gly	Leu	Gly	Ala	Gly
		195					200					205			
Ile	Gly	Ile	Pro	Arg	Phe	Ser	Ser	Trp	Asn	Asn	Pro	Glu	Pro	Glu	Leu
	210					215					220				
Val	Leu	Ile	Val	Thr	Ser	Arg	Gly	Glu	Val	Val	Gly	Ala	Thr	Leu	Gly
225					230					235					240
Asn	Asp	Val	Asn	Leu	Arg	Asp	Val	Glu	Gly	Arg	Ser	Ala	Leu	Leu	Leu
				245					250				255		
Gly	Lys	Ala	Lys	Asp	Asn	Asn	Ala	Ser	Ser	Ala	Leu	Gly	Pro	Leu	Ile
			260				265						270		
Arg	Leu	Phe	Asp	Gly	Ser	Phe	Thr	Val	Asp	Thr	Leu	Arg	Glu	Glu	Glu
		275					280					285			
Ile	Leu	Leu	Arg	Val	Glu	Gly	Leu	Asp	Gly	Tyr	Leu	Leu	Glu	Gly	Arg
	290					295					300				
Asn	Thr	Leu	Ala	Arg	Ile	Ser	Arg	Pro	Phe	Glu	Glu	Leu	Val	Ala	Ala
305					310					315					320
Thr	Arg	Gly	Arg	His	His	Gln	Tyr	Pro	Asp	Gly	Phe	Ala	Leu	Phe	Thr
				325					330					335	
Gly	Thr	Leu	Phe	Ala	Pro	Thr	Gln	Asp	Arg	Asp	Glu	Pro	Gly	Gln	Gly
			340					345					350		
Phe	Thr	His	Lys	His	Gly	Asp	Val	Val	Thr	Ile	Arg	Ser	Arg	His	Leu
		355					360					365			
Gly	Ala	Leu	Ile	Asn	Arg	Val	Gly	Thr	Ala	Glu	Glu	Leu	Pro	Glu	Trp
	370					375					380				
Thr	Phe	Gly	Leu	Arg	Gln	Leu	Phe	Gly	Tyr	Leu	Ala	Glu	Gln	Arg	Gln
385					390					395					400
Ala	Glu	Leu	Ala	Gln	Met	Gln	Glu								
				405											

<210> 102
 <211> 1227
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 102	
atgaccgccc cgccgatccc gggttcgtcc gtgccgccgg tgacctctgt tctgccggat	60
gatgctggcc aagctctgct ggttggtcgt atttgggacc cggcaaccgg cggtcgcgct	120
gtggttgccc tgagcgggta tagcgcggtg gacctgacct gtctggcagg cacggtgtct	180

```

gaactgctgg aactgccgga cccggcggcg gcggtgcgtg ctgcgctgct ggacccggcc 240
ctgggcgtcc agcgttgggc aaccgcagat gtctggcag caagcctggc aggtgacgct 300
gcacgtccgc atctgctggc accggtggat ctgcaagtta ttaaagcgtg cggcgtcacg 360
tttgtggact ctatgattga acgtgtgata gaagaacgtt gtgcaggtga tgcagcacgt 420
gctgcggaaa tgcgcgaact ggtgggcaaa gctctgggog gtagtatcgc aaccgttcgt 480
ccgggttccc cgggaagccgc agaagcaaaa cgcgttctga ttgcagaagg tctgtggagc 540
cagtatctgg aagtgggcat cggtcgggac ccggaagttt ttaccaaagc tccggtcctg 600
agctctgtgg gtctgggtgc aggtattggg atcccgcgtt tcagttcctg gaacaatccg 660
gaaccggaac tggttctgat tgtcacctca cgcggcgaag ttgtcgggtc aacgctgggc 720
aacgatgtta atctgcgtga cgtcgaaggt cgtcggctc tgctgctggg caaagcgaaa 780
gataacaatg cttcatcggc gctgggtccg ctgattcgtc tgtttgatgg cagtttcacc 840
gtggacacgc tgcgtgaaga agaaatcctg ctgcgcgttg aaggcctgga tggttatctg 900
ctggaaggtc gtaacaccct ggacgtatc tccgcgccgt ttgaagaact ggtggctgcg 960
acgcgtggtc gccatcacca ataccggat ggttttgccc tgttcaccgg cacgctgttt 1020
gcaccgaccc aggatcgtga cgaaccgggc caaggtttca occataaaca cggcgacgtg 1080
gttacgattc gtagtcgcca cctgggtgct ctgatcaatc gtgtgggcac cgcggaagaa 1140
ctgccggaat ggacgttcgg tctgcgccaa ctgtttggtt atctggcgga acaacgtcaa 1200
gcggaactgg ctcaaatgca ggaatga 1227

```

```

<210> 103
<211> 1446
<212> ADN
<213> Azospirillum brasilense

```

```

<220>
<221> CDS
<222> (1)..(1446)

```

```

<400> 103
atg gct aac gtg act tat acg gat acg caa ctg ctg atc gac ggc gag 48
Met Ala Asn Val Thr Tyr Thr Asp Thr Gln Leu Leu Ile Asp Gly Glu
1 5 10 15
tggtgc gac gcc gcg agc ggc aag acg atc gac gtc gtg aac ccg gcg 96
Trp Val Asp Ala Ala Ser Gly Lys Thr Ile Asp Val Val Asn Pro Ala
20 25 30
acc ggc aag ccg atc ggc agg gtg gcc cat gcg ggc atc gcc gat ctc 144
Thr Gly Lys Pro Ile Gly Arg Val Ala His Ala Gly Ile Ala Asp Leu
35 40 45
gac cgt gcg ctc gcc gcc gcg caa agc ggc ttc gag gca tgg cgc aag 192
Asp Arg Ala Leu Ala Ala Ala Gln Ser Gly Phe Glu Ala Trp Arg Lys
50 55 60
gtg ccc gcg cac gag cgc gcg gcg acg atg cgc aag gcg gcc gcg ctg 240
Val Pro Ala His Glu Arg Ala Ala Thr Met Arg Lys Ala Ala Ala Leu
65 70 75 80
gtg cgt gaa cgc gcc gac gcg atc gcg cag ctg atg acg cag gag cag 288
Val Arg Glu Arg Ala Asp Ala Ile Ala Gln Leu Met Thr Gln Glu Gln
85 90 95
ggc aag ccg ctc acc gaa gcg cgc gtc gaa gtg ctg tcg gcg gcg gac 336
Gly Lys Pro Leu Thr Glu Ala Arg Val Glu Val Leu Ser Ala Ala Asp
100 105 110
atc atc gaa tgg ttc gcg gac gaa ggc cgc cgc gtg tac ggc cgg atc 384
Ile Ile Glu Trp Phe Ala Asp Glu Gly Arg Arg Val Tyr Gly Arg Ile
115 120 125
gtg ccg ccg cgc aac ctc ggc gca cag cag acg gtc gtg aag gag ccg 432
Val Pro Pro Arg Asn Leu Gly Ala Gln Gln Thr Val Val Lys Glu Pro
130 135 140

```

gtc ggc ccg gtc gcc gcg ttc acg ccg tgg aat ttc ccg gtc aac cag	480
Val Gly Pro Val Ala Ala Phe Thr Pro Trp Asn Phe Pro Val Asn Gln	
145 150 155 160	
gtc gtg cgc aag ctg agc gcc gcg ctg gca acc ggc tgt tcg ttc ctc	528
Val Val Arg Lys Leu Ser Ala Ala Leu Ala Thr Gly Cys Ser Phe Leu	
165 170 175	
gtg aaa gcg ccg gaa gaa acc ccc gcg tcg ccg gcc gcg ctg ctg cgc	576
Val Lys Ala Pro Glu Glu Thr Pro Ala Ser Pro Ala Ala Leu Leu Arg	
180 185 190	
gcc ttc gtc gac gca ggc gtg ccg gcc ggc gtg atc ggc ctc gtg tac	624
Ala Phe Val Asp Ala Gly Val Pro Ala Gly Val Ile Gly Leu Val Tyr	
195 200 205	
ggc gat ccg gcc gaa atc tcg tcg tac ctg atc ccg cac ccg gtg atc	672
Gly Asp Pro Ala Glu Ile Ser Ser Tyr Leu Ile Pro His Pro Val Ile	
210 215 220	
cgc aag gtc acg ttc acg ggt tcg acg ccg gtc ggc aag cag ctc gcc	720
Arg Lys Val Thr Phe Thr Gly Ser Thr Pro Val Gly Lys Gln Leu Ala	
225 230 235 240	
tcg ctg gcg ggc ctg cac atg aag cgc gcg acg atg gag ctg ggc ggg	768
Ser Leu Ala Gly Leu His Met Lys Arg Ala Thr Met Glu Leu Gly Gly	
245 250 255	
cac gca ccg gtg atc gtg gcc gaa gac gcc gac gtt gcg ctc gcg gtg	816
His Ala Pro Val Ile Val Ala Glu Asp Ala Asp Val Ala Leu Ala Val	
260 265 270	
aaa gcg gcc ggc ggc gcg aag ttc cgc aac gcg ggc cag gtc tgc atc	864
Lys Ala Ala Gly Gly Ala Lys Phe Arg Asn Ala Gly Gln Val Cys Ile	
275 280 285	
tcg ccg acg cgc ttc ctc gtg cac aac agc atc cgc gac gaa ttc acg	912
Ser Pro Thr Arg Phe Leu Val His Asn Ser Ile Arg Asp Glu Phe Thr	
290 295 300	
cgc gcg ctg gtc aag cat gcc gaa ggc ctg aag gtc ggc aac ggc ctc	960
Arg Ala Leu Val Lys His Ala Glu Gly Leu Lys Val Gly Asn Gly Leu	
305 310 315 320	
gag gaa ggc acg acg ctc ggc gcg ctc gcg aac ccg cgc cgg ctg acc	1008
Glu Glu Gly Thr Thr Leu Gly Ala Leu Ala Asn Pro Arg Arg Leu Thr	
325 330 335	
gcg atg gcg tcg gtc atc gac aac gcg cgc aag gtc ggt gcg agc atc	1056
Ala Met Ala Ser Val Ile Asp Asn Ala Arg Lys Val Gly Ala Ser Ile	
340 345 350	
gaa acc ggc ggc gag cgg atc ggc tcg gaa ggc aac ttc ttc gcg ccg	1104
Glu Thr Gly Gly Glu Arg Ile Gly Ser Glu Gly Asn Phe Phe Ala Pro	
355 360 365	
acc gtg atc gcg aac gtg ccg ctc gat gcg gac gtg ttc aac aac gag	1152
Thr Val Ile Ala Asn Val Pro Leu Asp Ala Asp Val Phe Asn Asn Glu	
370 375 380	
ccg ttc ggc ccg gtc gcg gcg att cgc ggt ttc gac aag ctc gaa gag	1200
Pro Phe Gly Pro Val Ala Ala Ile Arg Gly Phe Asp Lys Leu Glu Glu	
385 390 395 400	
gcg atc gcg gaa gcg aac cgt ttg ccg ttc ggt ctt gcc ggc tac gcg	1248
Ala Ile Ala Glu Ala Asn Arg Leu Pro Phe Gly Leu Ala Gly Tyr Ala	
405 410 415	
ttc acg cgt tcg ttc gcg aac gtg cac ctg ctc acg cag cgc ctc gaa	1296
Phe Thr Arg Ser Phe Ala Asn Val His Leu Leu Thr Gln Arg Leu Glu	
420 425 430	
gtc ggg atg ctg tgg atc aac cag ccg gcg acg ccg tgg ccg gaa atg	1344
Val Gly Met Leu Trp Ile Asn Gln Pro Ala Thr Pro Trp Pro Glu Met	
435 440 445	

```

ccg ttc ggc ggc gtg aag gac tcg ggc tac ggt tcg gaa ggc ggc ccg      1392
Pro Phe Gly Gly Val Lys Asp Ser Gly Tyr Gly Ser Glu Gly Gly Pro
  450                      455                      460
gaa gcg ctc gag ccg tac ctg gtc acg aag tcg gtg acg gtg atg gcc      1440
Glu Ala Leu Glu Pro Tyr Leu Val Thr Lys Ser Val Thr Val Met Ala
  465                      470                      475                      480
gtc tga
Val

```

1446

```

<210> 104
<211> 481
<212> PRT
<213> Azospirillum brasilense.

```

```

<400> 104
Met Ala Asn Val Thr Tyr Thr Asp Thr Gln Leu Leu Ile Asp Gly Glu
  1          5          10          15
Trp Val Asp Ala Ala Ser Gly Lys Thr Ile Asp Val Val Asn Pro Ala
  20          25          30
Thr Gly Lys Pro Ile Gly Arg Val Ala His Ala Gly Ile Ala Asp Leu
  35          40          45
Asp Arg Ala Leu Ala Ala Ala Gln Ser Gly Phe Glu Ala Trp Arg Lys
  50          55          60
Val Pro Ala His Glu Arg Ala Ala Thr Met Arg Lys Ala Ala Ala Leu
  65          70          75          80
Val Arg Glu Arg Ala Asp Ala Ile Ala Gln Leu Met Thr Gln Glu Gln
  85          90          95
Gly Lys Pro Leu Thr Glu Ala Arg Val Glu Val Leu Ser Ala Ala Asp
  100         105         110
Ile Ile Glu Trp Phe Ala Asp Glu Gly Arg Arg Val Tyr Gly Arg Ile
  115         120         125
Val Pro Pro Arg Asn Leu Gly Ala Gln Gln Thr Val Val Lys Glu Pro
  130         135         140
Val Gly Pro Val Ala Ala Phe Thr Pro Trp Asn Phe Pro Val Asn Gln
  145         150         155         160
Val Val Arg Lys Leu Ser Ala Ala Leu Ala Thr Gly Cys Ser Phe Leu
  165         170         175
Val Lys Ala Pro Glu Glu Thr Pro Ala Ser Pro Ala Ala Leu Leu Arg
  180         185         190
Ala Phe Val Asp Ala Gly Val Pro Ala Gly Val Ile Gly Leu Val Tyr
  195         200         205
Gly Asp Pro Ala Glu Ile Ser Ser Tyr Leu Ile Pro His Pro Val Ile
  210         215         220
Arg Lys Val Thr Phe Thr Gly Ser Thr Pro Val Gly Lys Gln Leu Ala
  225         230         235         240
Ser Leu Ala Gly Leu His Met Lys Arg Ala Thr Met Glu Leu Gly Gly
  245         250         255
His Ala Pro Val Ile Val Ala Glu Asp Ala Asp Val Ala Leu Ala Val
  260         265         270
Lys Ala Ala Gly Gly Ala Lys Phe Arg Asn Ala Gly Gln Val Cys Ile
  275         280         285
Ser Pro Thr Arg Phe Leu Val His Asn Ser Ile Arg Asp Glu Phe Thr
  290         295         300
Arg Ala Leu Val Lys His Ala Glu Gly Leu Lys Val Gly Asn Gly Leu
  305         310         315         320
Glu Glu Gly Thr Thr Leu Gly Ala Leu Ala Asn Pro Arg Arg Leu Thr

```

325								330				335			
Ala	Met	Ala	Ser	Val	Ile	Asp	Asn	Ala	Arg	Lys	Val	Gly	Ala	Ser	Ile
340				345				350							
Glu	Thr	Gly	Gly	Glu	Arg	Ile	Gly	Ser	Glu	Gly	Asn	Phe	Phe	Ala	Pro
355				360				365							
Thr	Val	Ile	Ala	Asn	Val	Pro	Leu	Asp	Ala	Asp	Val	Phe	Asn	Asn	Glu
370				375				380							
Pro	Phe	Gly	Pro	Val	Ala	Ala	Ile	Arg	Gly	Phe	Asp	Lys	Leu	Glu	Glu
385				390				395				400			
Ala	Ile	Ala	Glu	Ala	Asn	Arg	Leu	Pro	Phe	Gly	Leu	Ala	Gly	Tyr	Ala
405				410				415							
Phe	Thr	Arg	Ser	Phe	Ala	Asn	Val	His	Leu	Leu	Thr	Gln	Arg	Leu	Glu
420				425				430							
Val	Gly	Met	Leu	Trp	Ile	Asn	Gln	Pro	Ala	Thr	Pro	Trp	Pro	Glu	Met
435				440				445							
Pro	Phe	Gly	Gly	Val	Lys	Asp	Ser	Gly	Tyr	Gly	Ser	Glu	Gly	Gly	Pro
450				455				460							
Glu	Ala	Leu	Glu	Pro	Tyr	Leu	Val	Thr	Lys	Ser	Val	Thr	Val	Met	Ala
465				470				475				480			
Val															

<210> 105
<211> 1446
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> ADN tổng hợp

<400>	105					
atggccaatg	tgacctatac	ggatacccag	ctgctgatcg	atggcgaatg	ggttgatgcg	60
gccagcggtg	aaacgattga	tgtggttaac	ccggcgaccg	gcaaaccgat	cggtcgtgtg	120
gcccattgcg	gcattgccga	tctggatcgc	gccctggcag	ccgcacagtc	tggttttgaa	180
gcctggcgta	aagtgcgggc	acacgaacgt	gcgccacca	tgcgtaaagc	ggcggcgctg	240
gtgctgaac	gtgccgatgc	catcgcacag	ctgatgacgc	aggaacaggg	caaaccgctg	300
accgaagccc	gcgtggaagt	tctgtctgcc	gcagatatta	tcgaatggtt	tgcggatgaa	360
ggcgcgtcgc	tgtacggtcg	tattgttccg	ccgcgcaatc	tgggtgccca	gcagaccgtg	420
gttaagaac	cggtgggtcc	ggttgccgca	tttaccocgt	ggaacttccc	ggtgaatcag	480
gtggttcgta	aactgagcgc	ggcgttggcc	accggttgca	gttttctggt	taaagcgccg	540
gaagaaaccc	cggaacgccc	ggcgcgactg	ctgcgcgcgt	ttgtggatgc	gggcgttccg	600
gcaggtgtg	ttggtctggt	ttatggtgat	ccggcgaaa	ttagctctta	cctgatcccg	660
catccggtga	ttcgtaaagt	taogtttaac	ggcagcacc	ctgtgggtaa	acagctggcg	720
tctctggccg	gtctgcata	gaaacgcgca	accatggaac	tgggcggtca	cgcgccggtg	780
attgttgca	aagatgcgga	tgtggcgctg	gcggttaag	ccgcggcg	tgcaaaattc	840
cgtaatgcg	gtcaggtgtg	catcagtcg	accgcctt	tggttcata	cagcattcgt	900
gatgaattta	ccgcgcct	ggtgaaacac	gcagaagtc	tgaaagttg	caacggtctg	960
gaagaaggca	ccacgctggg	tgcactggcc	aatccgcgtc	gcctgaccgc	aatggcaagc	1020
gtgatcgata	acgcgcgtaa	agttggcgcc	tctatcgaaa	cgggcggtga	acgcattggc	1080
agtgagggtg	actttttcgc	accgaccgtg	attgogaatg	ttccgctgga	tgcggatgtt	1140
tttaacaatg	aaccgttttg	cccggttgca	gcgatccgtg	gttttgata	actggaagaa	1200
gcgattgccg	aagcaaacgc	cctgccgttt	ggcctggccg	gttatgcatt	tacgcgtagt	1260
ttcgccaatg	tgcacctgct	gacccagcgc	ctggaagttg	gtatgctgtg	gatcaatcag	1320
ccggccaccc	cgtggccgga	aatgccgttt	ggcgggtgtg	aagatagtgg	ctatggtagc	1380
gaaggcggtc	cggaagccct	ggaaccgtac	ctgggtgacga	aaagcgtgac	cgttatggca	1440
qtttga						1446

<210> 106
 <211> 1440
 <212> ADN
 <213> Halomonas boliviensis

<220>
 <221> CDS
 <222> (1)..(1440)

<400> 106
 atg ggc gaa atg aac aac aat tta atc ggt ggc gta tgg cgc gag gcc 48
 Met Gly Glu Met Asn Asn Asn Leu Ile Gly Gly Val Trp Arg Glu Ala
 1 5 10 15
 gat gat gtc tca gaa aac act aac ccc tcc gat gtg acg gat ctg atc 96
 Asp Asp Val Ser Glu Asn Thr Asn Pro Ser Asp Val Thr Asp Leu Ile
 20 25 30
 ggg ctt tat gcc cgc ggc ggt gcc cag gat gtg acc gat gcc gcc gaa 144
 Gly Leu Tyr Ala Arg Gly Gly Ala Gln Asp Val Thr Asp Ala Ala Glu
 35 40 45
 gcg gca gag gtc gca atg ccc gcc tgg gcc ggc gcg acc cct cag ctg 192
 Ala Ala Glu Val Ala Met Pro Ala Trp Ala Gly Ala Thr Pro Gln Leu
 50 55 60
 cgc gct gat ctg ctg gac cgc gtg gcc gca gag atc acc cgc cgc gaa 240
 Arg Ala Asp Leu Leu Asp Arg Val Ala Ala Glu Ile Thr Arg Arg Glu
 65 70 75 80
 gat gag att gcg acg atg ctg gcc tgc gaa gag ggt aag gtt ctg tta 288
 Asp Glu Ile Ala Thr Met Leu Ala Cys Glu Gly Lys Val Leu Leu
 85 90 95
 gaa gcg ctg gga gaa gtt cgc cgc gcc gcg cat gtg ttt cgt ttt ttc 336
 Glu Ala Leu Gly Glu Val Arg Arg Ala Ala His Val Phe Arg Phe Phe
 100 105 110
 gcc gga gag gcg cta cgt ctg acg ggg gat gcc ctg gcc tcg gtg cga 384
 Ala Gly Glu Ala Leu Arg Leu Thr Gly Asp Ala Leu Ala Ser Val Arg
 115 120 125
 ccc ggc gtg gat gtg gag gtg acc cgc gag ccg tta ggc att gtc ggt 432
 Pro Gly Val Asp Val Glu Val Thr Arg Glu Pro Leu Gly Ile Val Gly
 130 135 140
 atc atc acc ccg tgg aac ttc ccc atc gcg att cct gcg tgg aaa att 480
 Ile Ile Thr Pro Trp Asn Phe Pro Ile Ala Ile Pro Ala Trp Lys Ile
 145 150 155 160
 gcc cct gcc tta gct tat ggc aac tgc gtg atc ttt aag cca gcg gaa 528
 Ala Pro Ala Leu Ala Tyr Gly Asn Cys Val Ile Phe Lys Pro Ala Glu
 165 170 175
 cag acg ccg ggc tca gcc cac ata ttg acc cag att att cat gag gcg 576
 Gln Thr Pro Gly Ser Ala His Ile Leu Thr Gln Ile Ile His Glu Ala
 180 185 190
 ggc tgc ccg gcg ggt gtt ttc aat ctc gtt atg ggg cgt ggc tcc gtc 624
 Gly Cys Pro Ala Gly Val Phe Asn Leu Val Met Gly Arg Gly Ser Val
 195 200 205
 gtc ggt gag gcg atg acc acg gat ccg cgc atc tcg gga att tcc ttt 672
 Val Gly Glu Ala Met Thr Thr Asp Pro Arg Ile Ser Gly Ile Ser Phe
 210 215 220
 acc ggg tcg gtg cag gta ggg cgg cag ctg gct ggc gcc tgc gcg gcg 720
 Thr Gly Ser Val Gln Val Gly Arg Gln Leu Ala Gly Ala Cys Ala Ala
 225 230 235 240
 aac atg aaa aaa ctt cag cta gag atg ggc ggt aag aac ccg gtg atc 768

Asn	Met	Lys	Lys	Leu	Gln	Leu	Glu	Met	Gly	Gly	Lys	Asn	Pro	Val	Ile	
				245					250					255		
gtg	atg	gac	gac	gcc	gat	ctg	gaa	acc	gct	gtt	tct	gtc	tgt	ctc	aac	816
Val	Met	Asp	Asp	Ala	Asp	Leu	Glu	Thr	Ala	Val	Ser	Val	Cys	Leu	Asn	
				260					265					270		
ggg	gcc	ttt	tac	cag	acg	ggg	cag	cgc	tgt	acc	gca	tcg	tcg	cgt	cta	864
Gly	Ala	Phe	Tyr	Gln	Thr	Gly	Gln	Arg	Cys	Thr	Ala	Ser	Ser	Arg	Leu	
				275					280					285		
atc	gta	cag	tcg	ggc	atc	cat	gat	gcc	ttt	atc	gcg	gaa	ctc	agt	cgc	912
Ile	Val	Gln	Ser	Gly	Ile	His	Asp	Ala	Phe	Ile	Ala	Glu	Leu	Ser	Arg	
				290					295					300		
cag	atg	cag	gcg	ctc	aag	gtg	ggg	cat	gcg	ctc	gcc	gag	ggg	atg	cag	960
Gln	Met	Gln	Ala	Leu	Lys	Val	Gly	His	Ala	Leu	Ala	Glu	Gly	Met	Gln	
				305											320	
att	ggc	ccc	gtc	gcc	gcc	cgg	agt	cag	ctc	gaa	agc	aat	ctt	tac	tac	1008
Ile	Gly	Pro	Val	Ala	Ala	Arg	Ser	Gln	Leu	Glu	Ser	Asn	Leu	Tyr	Tyr	
				325											335	
gtg	gcg	ctc	gcc	gct	gaa	gag	ggc	tgc	gat	gta	ttg	ggg	gga	gag	cag	1056
Val	Ala	Leu	Ala	Ala	Glu	Glu	Gly	Cys	Asp	Val	Leu	Gly	Gly	Glu	Gln	
				340											350	
ctc	aca	cgc	gac	act	gaa	ggg	tat	ttc	cag	gca	ccc	gca	ctt	ttc	ctt	1104
Leu	Thr	Arg	Asp	Thr	Glu	Gly	Tyr	Phe	Gln	Ala	Pro	Ala	Leu	Phe	Leu	
				355											365	
ggg	gcg	agt	aat	gcc	atg	cgt	agc	gcc	cgt	gaa	gag	atc	ttc	ggc	ccc	1152
Gly	Ala	Ser	Asn	Ala	Met	Arg	Ser	Ala	Arg	Glu	Glu	Ile	Phe	Gly	Pro	
				370											380	
tgt	gcc	agc	gtg	att	cgt	gtc	gat	gac	ttt	gaa	gaa	gca	gta	gcg	ctg	1200
Cys	Ala	Ser	Val	Ile	Arg	Val	Asp	Asp	Phe	Glu	Glu	Ala	Val	Ala	Leu	
				385											400	
gcc	aat	gac	acc	cag	ttc	ggg	ctg	tcg	tcg	ggg	att	tgt	acc	acc	aat	1248
Ala	Asn	Asp	Thr	Gln	Phe	Gly	Leu	Ser	Ser	Gly	Ile	Cys	Thr	Thr	Asn	
				405											415	
cta	cgc	tat	gcg	cgg	gag	ttc	aag	cgc	cgt	tct	gcc	gct	ggc	atg	gtg	1296
Leu	Arg	Tyr	Ala	Arg	Glu	Phe	Lys	Arg	Arg	Ser	Ala	Ala	Gly	Met	Val	
				420											430	
atg	ctc	aac	ttg	ccg	acc	gcg	ggc	gtg	gat	tat	cac	gtg	ccc	ttt	ggc	1344
Met	Leu	Asn	Leu	Pro	Thr	Ala	Gly	Val	Asp	Tyr	His	Val	Pro	Phe	Gly	
				435											445	
ggg	cgt	aaa	gcg	tcg	agc	ttc	ggg	tca	cgt	gag	cag	ggg	agc	tat	gca	1392
Gly	Arg	Lys	Ala	Ser	Ser	Phe	Gly	Ser	Arg	Glu	Gln	Gly	Ser	Tyr	Ala	
				450											460	
gcc	gaa	ttc	tat	aca	tcg	gtg	aaa	acc	gcc	tat	acg	tat	gcc	ggg	taa	1440
Ala	Glu	Phe	Tyr	Thr	Ser	Val	Lys	Thr	Ala	Tyr	Thr	Tyr	Ala	Gly		
				465											475	

<210> 107

<211> 479

<212> PRT

<213> Halomonas boliviensis

<400> 107

Met	Gly	Glu	Met	Asn	Asn	Asn	Leu	Ile	Gly	Gly	Val	Trp	Arg	Glu	Ala
1				5					10					15	
Asp	Asp	Val	Ser	Glu	Asn	Thr	Asn	Pro	Ser	Asp	Val	Thr	Asp	Leu	Ile
				20				25					30		
Gly	Leu	Tyr	Ala	Arg	Gly	Gly	Ala	Gln	Asp	Val	Thr	Asp	Ala	Ala	Glu

Ala	Ala	Glu	Val	Ala	Met	Pro	Ala	Trp	Ala	Gly	Ala	Thr	Pro	Gln	Leu
50						55					60				
Arg	Ala	Asp	Leu	Leu	Asp	Arg	Val	Ala	Ala	Glu	Ile	Thr	Arg	Arg	Glu
65					70					75					80
Asp	Glu	Ile	Ala	Thr	Met	Leu	Ala	Cys	Glu	Glu	Gly	Lys	Val	Leu	Leu
				85					90					95	
Glu	Ala	Leu	Gly	Glu	Val	Arg	Arg	Ala	Ala	His	Val	Phe	Arg	Phe	Phe
			100					105					110		
Ala	Gly	Glu	Ala	Leu	Arg	Leu	Thr	Gly	Asp	Ala	Leu	Ala	Ser	Val	Arg
	115						120					125			
Pro	Gly	Val	Asp	Val	Glu	Val	Thr	Arg	Glu	Pro	Leu	Gly	Ile	Val	Gly
	130					135					140				
Ile	Ile	Thr	Pro	Trp	Asn	Phe	Pro	Ile	Ala	Ile	Pro	Ala	Trp	Lys	Ile
145					150					155					160
Ala	Pro	Ala	Leu	Ala	Tyr	Gly	Asn	Cys	Val	Ile	Phe	Lys	Pro	Ala	Glu
				165					170					175	
Gln	Thr	Pro	Gly	Ser	Ala	His	Ile	Leu	Thr	Gln	Ile	Ile	His	Glu	Ala
			180					185					190		
Gly	Cys	Pro	Ala	Gly	Val	Phe	Asn	Leu	Val	Met	Gly	Arg	Gly	Ser	Val
	195						200					205			
Val	Gly	Glu	Ala	Met	Thr	Thr	Asp	Pro	Arg	Ile	Ser	Gly	Ile	Ser	Phe
	210					215					220				
Thr	Gly	Ser	Val	Gln	Val	Gly	Arg	Gln	Leu	Ala	Gly	Ala	Cys	Ala	Ala
225					230					235					240
Asn	Met	Lys	Lys	Leu	Gln	Leu	Glu	Met	Gly	Gly	Lys	Asn	Pro	Val	Ile
				245					250					255	
Val	Met	Asp	Asp	Ala	Asp	Leu	Glu	Thr	Ala	Val	Ser	Val	Cys	Leu	Asn
		260						265					270		
Gly	Ala	Phe	Tyr	Gln	Thr	Gly	Gln	Arg	Cys	Thr	Ala	Ser	Ser	Arg	Leu
	275						280					285			
Ile	Val	Gln	Ser	Gly	Ile	His	Asp	Ala	Phe	Ile	Ala	Glu	Leu	Ser	Arg
	290					295					300				
Gln	Met	Gln	Ala	Leu	Lys	Val	Gly	His	Ala	Leu	Ala	Glu	Gly	Met	Gln
305					310					315					320
Ile	Gly	Pro	Val	Ala	Ala	Arg	Ser	Gln	Leu	Glu	Ser	Asn	Leu	Tyr	Tyr
				325					330					335	
Val	Ala	Leu	Ala	Ala	Glu	Glu	Gly	Cys	Asp	Val	Leu	Gly	Gly	Glu	Gln
		340						345					350		
Leu	Thr	Arg	Asp	Thr	Glu	Gly	Tyr	Phe	Gln	Ala	Pro	Ala	Leu	Phe	Leu
	355						360					365			
Gly	Ala	Ser	Asn	Ala	Met	Arg	Ser	Ala	Arg	Glu	Glu	Ile	Phe	Gly	Pro
	370					375					380				
Cys	Ala	Ser	Val	Ile	Arg	Val	Asp	Asp	Phe	Glu	Glu	Ala	Val	Ala	Leu
385					390					395					400
Ala	Asn	Asp	Thr	Gln	Phe	Gly	Leu	Ser	Ser	Gly	Ile	Cys	Thr	Thr	Asn
				405					410					415	
Leu	Arg	Tyr	Ala	Arg	Glu	Phe	Lys	Arg	Arg	Ser	Ala	Ala	Gly	Met	Val
			420					425					430		
Met	Leu	Asn	Leu	Pro	Thr	Ala	Gly	Val	Asp	Tyr	His	Val	Pro	Phe	Gly
	435						440					445			
Gly	Arg	Lys	Ala	Ser	Ser	Phe	Gly	Ser	Arg	Glu	Gln	Gly	Ser	Tyr	Ala
	450					455					460				
Ala	Glu	Phe	Tyr	Thr	Ser	Val	Lys	Thr	Ala	Tyr	Thr	Tyr	Ala	Gly	
465					470					475					

<210> 108
 <211> 1440
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp

<400> 108
 atgggcgaaa tgaacaataa cctgatcggc ggtgtttggc gtgaagccga tgatgtgagt 60
 gaaaaatacca acccgagcga tgtgacggat ctgattggcc tgtatgcccg cggcggtgca 120
 caggatgtta ccgatgcggc cgaagcggcc gaagtggcaa tgccggcctg ggcaggtgca 180
 acgccgcagc tgcgtgcgga tctgctggat cgcgttgccg cagaaatcac ccgtcgcgaa 240
 gatgaaaattg cgacgatgct ggcctgcgaa gaaggcaaaag tgctgctgga agccctgggt 300
 gaagtctctc gcgcggccca tgtttttcgt ttctttgcag gtgaagccct gcgtctgacc 360
 ggtgatgccc tggcatctgt tccgtccggc gtggatgttg aagtgaccg cgaaccgctg 420
 ggcacgtgg gtattatcac gccgtggaat ttcccgattg ccatcccggc atggaaaatc 480
 gcaccggcgc tggcctatgg caactgcgtt atttttaaac cggcagaaca gaccccggtg 540
 agcgcgcata ttctgacgca gattatccac gaagcaggct gtccggcggg tgttttcaat 600
 ctggtgatgg gccgtggtag cgtggttggg gaagcgatga ccacggatcc gcgcattagc 660
 ggcacatctt ttaccgggtt tgttcaagtg ggccgtcagc tggcaggtgc ctgcgcagcg 720
 aatatgaaaa aactgcagct ggaaatgggc ggtaaaaacc cggttatcgt gatggatgat 780
 gcggatctgg aaaccgccgt tagcgtgtgc ctgaacggcg cgttctatca gaccggtcag 840
 cgttgtaagg ccagctctcg cctgattgtg cagtctggca tccatgatgc gtttattgcc 900
 gaactgagtc gtcagatgca ggccctgaaa gttggtcacg ccctggcaga aggcattgag 960
 attggtccgg tggccgcacg cagtcagctg gaaagcaatc tgtattacgt tgcgctggcg 1020
 gccgaagaag gctgtgatgt gctgggcggg gaacagctga cccgtgatac ggaaggctac 1080
 ttccaggcgc cggccctgtt cctgggtgca tctaacgca tgcgtagtgc ccggaagaa 1140
 atctttggtc cgtgcgcaag tgttattcgc gtggatgatt tcgaagaagc agttgcgctg 1200
 gccaatgata cccagtttgg cctgagtagc ggtatttgta ccacgaacct gcgttatgcg 1260
 cgcgaattta aacgtcgag cgcagcgggc atggtgatgc tgaatctgcc gaccgccggt 1320
 gttgattacc acgtgccgtt tggcggctgt aaagcgtcta gtttcggcag ccgcgaacag 1380
 ggttcttatg ccgcagaatt ttacaccagc gtgaaaaccg cctatacgta cgcagggttaa 1440

<210> 109
 <211> 1467
 <212> ADN
 <213> Bacillus subtilis

<220>
 <221> CDS
 <222> (1)..(1467)

<400> 109
 atg tct gtg atc acg gaa caa aac acg tac ctc aac ttt att aac gga 48
 Met Ser Val Ile Thr Glu Gln Asn Thr Tyr Leu Asn Phe Ile Asn Gly
 1 5 10 15
 gag tgg gtt aag tct caa tca ggc gat atg gtc aaa gtc gaa aac cct 96
 Glu Trp Val Lys Ser Gln Ser Gly Asp Met Val Lys Val Glu Asn Pro
 20 25 30
 gcc gat gtg aat gat att gtc gga tat gta cag aat tca acg gct gaa 144
 Ala Asp Val Asn Asp Ile Val Gly Tyr Val Gln Asn Ser Thr Ala Glu
 35 40 45
 gat gtg gaa cgt gcc gtc acc gcc gcc aat gaa gcc aaa acg gct tgg 192
 Asp Val Glu Arg Ala Val Thr Ala Ala Asn Glu Ala Lys Thr Ala Trp
 50 55 60

aga aag ctg acg ggt gcc gag cgc ggc caa tac tta tac aaa aca gcg	240
Arg Lys Leu Thr Gly Ala Glu Arg Gly Gln Tyr Leu Tyr Lys Thr Ala	
65 70 75 80	
gat atc atg gag cag cgc ttg gag gaa atc gcc gcc tgt gca acg cgt	288
Asp Ile Met Glu Gln Arg Leu Glu Glu Ile Ala Ala Cys Ala Thr Arg	
85 90 95	
gaa atg ggt aaa aca ttg ccg gaa gcg aag gga gaa aca gcc cgg ggg	336
Glu Met Gly Lys Thr Leu Pro Glu Ala Lys Gly Glu Thr Ala Arg Gly	
100 105 110	
att gcc att ctg cgc tat tac gcc gga gag ggc atg cga aaa acg ggt	384
Ile Ala Ile Leu Arg Tyr Tyr Ala Gly Glu Gly Met Arg Lys Thr Gly	
115 120 125	
gac gtc att ccg tct act gac aaa gac gcg ctc atg ttt acc acc cgt	432
Asp Val Ile Pro Ser Thr Asp Lys Asp Ala Leu Met Phe Thr Thr Arg	
130 135 140	
gtt ccg ctc ggt gtg gtc ggt gtg att tct ccg tgg aac ttc cca gtg	480
Val Pro Leu Gly Val Val Gly Val Ile Ser Pro Trp Asn Phe Pro Val	
145 150 155 160	
gcg att ccg att tgg aaa atg gcg ccg gca ttg gta tac ggc aat acc	528
Ala Ile Pro Ile Trp Lys Met Ala Pro Ala Leu Val Tyr Gly Asn Thr	
165 170 175	
gtt gtc atc aaa ccg gcg aca gaa aca gct gtg aca tgc gcg aag atc	576
Val Val Ile Lys Pro Ala Thr Glu Thr Ala Val Thr Cys Ala Lys Ile	
180 185 190	
att gcc tgc ttt gag gaa gcg ggg ctc ccg gca ggg gtc atc aat ttg	624
Ile Ala Cys Phe Glu Glu Ala Gly Leu Pro Ala Gly Val Ile Asn Leu	
195 200 205	
gtg aca ggc ccg ggt tct gtt gtc ggg cag ggg ctt gct gag cat gac	672
Val Thr Gly Pro Gly Ser Val Val Gly Gln Gly Leu Ala Glu His Asp	
210 215 220	
ggt gta aac gcc gtt acg ttt acc ggt tca aat caa gtc gga aaa atc	720
Gly Val Asn Ala Val Thr Phe Thr Gly Ser Asn Gln Val Gly Lys Ile	
225 230 235 240	
atc ggg caa gcc gct tta gcg agg gga gcc aaa tat cag ctt gag atg	768
Ile Gly Gln Ala Ala Leu Ala Arg Gly Ala Lys Tyr Gln Leu Glu Met	
245 250 255	
ggc ggc aaa aac cct gtc atc gta gct gat gac gct gac ctt gaa gct	816
Gly Gly Lys Asn Pro Val Ile Val Ala Asp Asp Ala Asp Leu Glu Ala	
260 265 270	
gcg gca gaa gct gtc ata acg ggg gcc ttc cgt tca acc ggc cag aaa	864
Ala Ala Glu Ala Val Ile Thr Gly Ala Phe Arg Ser Thr Gly Gln Lys	
275 280 285	
tgc acc gcg aca agc cgt gtc atc gta caa agc gga att tac gag cgc	912
Cys Thr Ala Thr Ser Arg Val Ile Val Gln Ser Gly Ile Tyr Glu Arg	
290 295 300	
ttt aaa gaa aaa ctg ctc cag cgc aca aaa gat att aca atc gga gac	960
Phe Lys Glu Lys Leu Leu Gln Arg Thr Lys Asp Ile Thr Ile Gly Asp	
305 310 315 320	
agc tta aaa gag gat gtc tgg atg gga ccg ata gcc agc aag aat cag	1008
Ser Leu Lys Glu Asp Val Trp Met Gly Pro Ile Ala Ser Lys Asn Gln	
325 330 335	
ctt gat aac tgc ctg tca tac att gag aaa ggc aaa cag gag ggc gct	1056
Leu Asp Asn Cys Leu Ser Tyr Ile Glu Lys Gly Lys Gln Glu Gly Ala	
340 345 350	
tcc ctt tta ata gga gga gaa aag ctg gag aac gga aag tat caa aac	1104
Ser Leu Leu Ile Gly Gly Glu Lys Leu Glu Asn Gly Lys Tyr Gln Asn	
355 360 365	

```

ggc tat tat gtt cag cct gcc atc ttt gac aat gtg aca tct gag atg      1152
Gly Tyr Tyr Val Gln Pro Ala Ile Phe Asp Asn Val Thr Ser Glu Met
370 375 380
aca att gcc cag gag gaa att ttc ggt ccg gtg atc gcc ttg atc aag      1200
Thr Ile Ala Gln Glu Glu Ile Phe Gly Pro Val Ile Ala Leu Ile Lys
385 390 395 400
gtg gac tcg ata gag gag gcg ctg aac atc gcc aat gat gtg aag ttc      1248
Val Asp Ser Ile Glu Glu Ala Leu Asn Ile Ala Asn Asp Val Lys Phe
405 410 415
ggg tta agt gca tcc atc ttc acg gaa aac atc ggc cga atg ctt tct      1296
Gly Leu Ser Ala Ser Ile Phe Thr Glu Asn Ile Gly Arg Met Leu Ser
420 425 430
ttc att gat gaa atc gat gcc ggg ctg gtt cgg atc aat gca gaa agc      1344
Phe Ile Asp Glu Ile Asp Ala Gly Leu Val Arg Ile Asn Ala Glu Ser
435 440 445
gca ggt gtt gag ctg cag gcg cct ttt ggc ggc atg aag cag tcg agc      1392
Ala Gly Val Glu Leu Gln Ala Pro Phe Gly Gly Met Lys Gln Ser Ser
450 455 460
tcc cac tcc cga gaa cag ggt gag gca gcg aag gac ttt ttc aca gcg      1440
Ser His Ser Arg Glu Gln Gly Glu Ala Ala Lys Asp Phe Phe Thr Ala
465 470 475 480
atc aaa act gtt ttt gtg aag ccg taa      1467
Ile Lys Thr Val Phe Val Lys Pro
485

```

```

<210> 110
<211> 488
<212> PRT
<213> Bacillus subtilis

```

```

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

```

195	200	205
Val Thr Gly Pro Gly Ser Val Val Gly Gln Gly Leu Ala Glu His Asp		
210	215	220
Gly Val Asn Ala Val Thr Phe Thr Gly Ser Asn Gln Val Gly Lys Ile		
225	230	235
Ile Gly Gln Ala Ala Leu Ala Arg Gly Ala Lys Tyr Gln Leu Glu Met		240
	245	250
		255
Gly Gly Lys Asn Pro Val Ile Val Ala Asp Asp Ala Asp Leu Glu Ala		
260	265	270
Ala Ala Glu Ala Val Ile Thr Gly Ala Phe Arg Ser Thr Gly Gln Lys		
275	280	285
Cys Thr Ala Thr Ser Arg Val Ile Val Gln Ser Gly Ile Tyr Glu Arg		
290	295	300
Phe Lys Glu Lys Leu Leu Gln Arg Thr Lys Asp Ile Thr Ile Gly Asp		
305	310	315
Ser Leu Lys Glu Asp Val Trp Met Gly Pro Ile Ala Ser Lys Asn Gln		320
	325	330
		335
Leu Asp Asn Cys Leu Ser Tyr Ile Glu Lys Gly Lys Gln Glu Gly Ala		
340	345	350
Ser Leu Leu Ile Gly Gly Glu Lys Leu Glu Asn Gly Lys Tyr Gln Asn		
355	360	365
Gly Tyr Tyr Val Gln Pro Ala Ile Phe Asp Asn Val Thr Ser Glu Met		
370	375	380
Thr Ile Ala Gln Glu Glu Ile Phe Gly Pro Val Ile Ala Leu Ile Lys		
385	390	395
Val Asp Ser Ile Glu Glu Ala Leu Asn Ile Ala Asn Asp Val Lys Phe		400
	405	410
		415
Gly Leu Ser Ala Ser Ile Phe Thr Glu Asn Ile Gly Arg Met Leu Ser		
420	425	430
Phe Ile Asp Glu Ile Asp Ala Gly Leu Val Arg Ile Asn Ala Glu Ser		
435	440	445
Ala Gly Val Glu Leu Gln Ala Pro Phe Gly Gly Met Lys Gln Ser Ser		
450	455	460
Ser His Ser Arg Glu Gln Gly Glu Ala Ala Lys Asp Phe Phe Thr Ala		
465	470	475
		480
Ile Lys Thr Val Phe Val Lys Pro		
485		

<210> 111
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> đoạn môi Ptac_xylXABC_F

<400> 111
 tgcgaagggc cttcggagcg atccggaggc

30

<210> 112
 <211> 36
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> đoạn mỗi Ptac_xylXABC_R
 <400> 112
 cgaaggccct tcgcattaga catggcggac ctcattg 36
 <210> 113
 <211> 36
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi Ptac_xylABCD_F
 <400> 113
 caatttcaca caggaggatg accgacaccc tgcgcc 36
 <210> 114
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi Ptac_xylABCD_R
 <400> 114
 cctcctgtgt gaaattgtta tccgctcacg 30
 <210> 115
 <211> 35
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi Ptac_xylXBCD_F
 <400> 115
 cgaaagcttc ccgcgatgtc ctcagccatc tatcc 35
 <210> 116
 <211> 30
 <212> ADN
 <213> Trình tự nhân tạo
 <220>
 <223> đoạn mỗi Ptac_xylXBCD_R
 <400> 116
 cgcggaagc tttcgattag aggaggccgc 30
 <210> 117
 <211> 37

<212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi yjhG_F

 <400> 117
 cacaaggaga ctcccatgtc tggtcgcaat atttttg 37

 <210> 118
 <211> 36
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi yjhG_R

 <400> 118
 gaactggcgg ctccctcagt ttttattcat aaaatc 36

 <210> 119
 <211> 36
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi yagF_F

 <400> 119
 cacaaggaga ctcccatgac cattgagaaa attttc 36

 <210> 120
 <211> 38
 <212> ADN
 <213> Trình tự nhân tạo

 <220>
 <223> đoạn mỗi yagF_R

 <400> 120
 gaactggcgg ctcccttaaa ttccgagcgc ttttttac 38