(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) **1-0048541**

(51)[7] **C12P 13/12**; C12R 1/19; C12N 15/70 (13) **B**

(21) 1-2015-00931 (22) 31/07/2013
(86) PCT/EP2013/066066 31/07/2013 (87) WO2014/029592 27/02/2014
(30) 12181028.7 20/08/2012 EP
(45) 25/07/2025 448 (43) 25/11/2015 332A
(73) Evonik Operations GmbH (DE)
Rellinghauser Strasse 1-11, 45128 Essen, Germany
(72) BATHE, Brigitte (DE); MOLCK, Stella (DE); PRIEFERT, Horst (DE).
(74) Công ty Luật TNHH Phạm và Liên danh (PHAM & ASSOCIATES)

(54) QUY TRÌNH SẢN XUẤT L-AXIT AMIN HOẶC CHẤT PHỤ GIA THỨC ĂN GIA SÚC BẰNG CÁCH LÊN MEN VI SINH VẬT THUỘC HỌ ENTEROBACTERIACEAE VÀ VI SINH VẬT THUỘC HỌ ENTEROBACTERIACEAE

(21) 1-2015-00931

(57) Sáng chế đề cập đến quy trình sản xuất L-axit amin hoặc chất phụ gia thức ăn gia súc bằng phương pháp lên men bằng cách sử dụng các vi sinh vật thuộc họ *Enterobacteriaceae*, mang gen *proP* được làm suy giảm, các vi sinh vật thích hợp cho quá trình sản xuất này, các polynucleotit mã hoá các biến thể của chất vận chuyển ProP, vi sinh vật tái tổ hợp và quy trình nhận diện vi sinh vật.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến quy trình sản xuất L-axit amin bằng phương pháp lên men bằng cách sử dụng các vi sinh vật thuộc họ *Enterobacteriaceae*, mang gen *proP* được làm suy giảm, đến các vi sinh vật thích hợp cho quá trình sản xuất nêu trên và các polynucleotit mã hoá các biến thể của chất vận chuyển ProP.

Tình trạng kỹ thuật của sáng chế

Các hợp chất hóa hữu cơ, cụ thể hơn L-axit amin chứa lưu huỳnh, là có ý nghĩa kinh tế quan trọng. L-xystein được sử dụng làm chất bổ sung vào thực phẩm, làm nguyên liệu điều chế dược chất (ví dụ N-axetylxystein) và dùng làm mỹ phẩm. Axit amin L-metionin đóng vai trò chủ đạo làm thức ăn cho động vật và là một trong số axit amin thiết yếu không thể tạo ra được bằng con đường sinh tổng hợp trong quá trình trao đổi chất của động vật có xương sống. Khi nuôi động vật phải đảm bảo đủ lượng axit amin nhất định trong thức ăn. Tuy nhiên, vì L-metionin, chẳng hạn, thường có ở các loại cây làm thức ăn thông thường (như cây đậu tương hoặc cây ngũ cốc) với lượng quá thấp để đảm bảo làm nguồn dinh dưỡng cho động vật tối ưu, đặc biệt là đối với lợn và gia cầm, sẽ có lợi nếu phối trộn metionin như là chất phụ gia của thức ăn động vật. Động vật có xương sống có thể chuyển hoá D-metionin thành L-metionin có hoạt tính sinh học. Do đó, raxemat của D- và L-metionin thường được bổ sung vào thức ăn động vật. Động vật có thể chuyển hoá L-homoxystein thành L-metionin bằng cách chuyển hóa metyl, và do đó L-homoxystein có thể thay thế L-metionin.

Các hợp chất hữu cơ được đề cập sau đây dùng để chỉ một hoặc nhiều hợp chất được chọn từ nhóm gồm các L-axit amin, tốt hơn nếu L-axit amin chứa lưu huỳnh, đặc biệt là L-metionin, L-xystein, L-xystin, L-homoxystein và L-homoxystin. L-metionin là được ưu tiên.

Trong tình trạng kỹ thuật, các axit amin như metionin được điều chế bằng cách tổng hợp hóa học. Trước tiên phương pháp điều chế này bao gồm việc cho acrolein phản ứng với metyl mercaptan để tạo ra 3-(metylthio)propionalđehyt mà chất này lại phản ứng với xyanua, amoniac và cacbon monoxit để tạo ra hyđantoin. Cuối cùng, hyđantoin có thể được thủy phân thành raxemat, hỗn hợp đẳng mol của hai chất đồng phân lập thể, D- và L-metionin. Vì dạng có hoạt tính sinh học của phân tử này chỉ là dạng L, dạng D có mặt trong thức ăn gia súc trước hết phải được chuyển hóa bằng cách trao đổi chất bằng quá trình khử và chuyển hóa amin thành dạng L hoạt tính.

Ngược lại với metionin, phần lớn các axit amin có bản chất protein tự nhiên khác như L-threonin, chẳng hạn, chủ yếu được tạo ra bằng cách lên men vi sinh vật. Quá trình dựa trên điều là các vi sinh vật có các quá trình sinh tổng hợp thích hợp để tổng hợp nên các axit amin tự nhiên. Hơn nữa, nhiều quy trình lên men như vậy cho chi phí sản xuất rất kinh tế do sử dụng các nguyên liệu rẻ tiền như glucoza và các muối vô cơ, và còn chuyển vận dạng L có hoạt tính sinh học của axit amin nhất định.

Đã biết rằng nhiều hợp chất có thể được tạo ra bằng cách lên men chủng *Enterobacteriaceae*, đặc biệt là *Escherichia coli* (*E. coli*) và *Serratia marcescens*. Do các quy trình như vậy có ý nghĩa quan trọng, nên người ta luôn cố gắng cải thiện các quy trình này. Các phương pháp cải thiện quy trình có thể bao gồm các biện pháp liên quan đến công nghệ lên men, ví dụ việc khuấy trộn và cung cấp oxy, hoặc thành phần của môi trường dinh dưµng, như, ví dụ, chọn lọc đường được sử dụng hoặc nồng độ đường trong quá trình lên men, hoặc cho đến bước hoàn thiện dạng sản phẩm, ví dụ bằng cách sắc ký trao đổi ion, hoặc các đặc tính biểu hiện nội tại của chính vi sinh vật.

Các quá trình sinh tổng hợp axit amin ở các chủng kiểu dại chịu sự kiểm soát trao đổi chất nghiêm ngặt để các axit amin được tạo ra chỉ để đáp ứng nhu cầu nội tại của tế bào. Do đó, một yêu cầu quan trọng để có quy trình sản xuất hiệu quả là có các vi sinh vật thích hợp mà khác với các sinh vật kiểu dại có sản lượng sản xuất được tăng lên đáng kể, vượt quá nhu cầu nội tại (sản xuất quá mức), để sản xuất axit amin mong muốn.

Các vi sinh vật sản xuất quá mức axit amin như vậy có thể được tạo ra bằng các quá trình gây đột biến/chọn lọc kinh điển và/hoặc bằng các kỹ thuật tái tổ hợp đặc hiệu hiện đại ("kỹ thuật điều khiển trao đổi chất"). Các kỹ thuật tái tổ hợp trước hết bao gồm việc nhận diện gen hoặc alen gây ra sự sản xuất axit amin quá mức do sự cải biến, hoạt hoá hoặc làm bất hoạt. Sau đó, các gen/alen này được đưa vào chủng vi sinh vật hoặc được làm bất hoạt bằng cách sử dụng các kỹ thuật sinh học phân tử để đạt được sự sản xuất quá mức tối ưu. Tuy nhiên, thông thường chỉ bằng cách kết hợp nhiều biện pháp khác nhau mới mang lại quá trình sản xuất thật sự hiệu quả.

L-metionin, cùng với lysin và threonin, là thu được từ aspartat. Lưu huỳnh được đưa vào ở dạng L-xystein (thông qua xystathionin là hợp chất trung gian) được chuyển thành L-metionin thông qua quá trình chuyển lưu huỳnh. Nhóm CH_3 của L-metionin có nguồn gốc từ quá trình trao đổi chất C1 và được chuyển thành L-homoxystein bằng các MetE hoặc MetH metionin synthaza (xem: Greene RC (1996) trong Neidhardt FC et al. (eds.) "Escherichia coli and Salmonella", 2nd edition, các trang 542-560). Các chủng và quy trình sản xuất bằng phương pháp lên men L-metionin đã được mô tả cho *E. coli* trong ví dụ WO2006/001616 hoặc WO2009/043803.

Hondrop et al. mô tả sự tăng trưởng của *E. coli* trong môi trường chứa metionin (*Protein Engineering for Therapeutics*, Part B, Vol. 290, No. 11, p. 59). Đơn yêu cầu cấp patent châu Âu số EP2479279A1 mô tả quá trình lên men của các axit amin chứa lưu huỳnh, trong đó các vi sinh vật được sử dụng thể hiện các hoạt tính tăng hoặc giảm của các enzym liên quan và/hoặc các chất vận chuyển và/hoặc chất xuất metionin.

ProP là tải đồng vận hòa tan proton/tương thích của *E. coli*, và do đó là một trong các chất vận chuyển ở *E. coli* có hoạt tính trong điều kiện tăng thẩm thấu (Grothe et al., Journal of Bacteriology 166, 253-259 (1986); Racher et al., Biochemistry 38, 1676-1684 (1999); Wood, Methods in Enzymology 428, 77-107 (2007)). Khi có sự tăng nồng độ thẩm thấu của môi trường bao quanh tế bào, thế năng nước của môi trường giảm đi và các phân tử nước khuếch tán theo građient thẩm thấu

ra ngoài tế bào. Áp suất sức trương của tế bào giảm đi, và dẫn đến, protein tế bào chất bị mất vỏ hyđrat hóa liên quan chức năng của nó. Do quá trình mất nước này, quá trình trao đổi chất của tế bào và sự phân chia tế bào bị dừng lại. Để ngăn ngừa quá trình này, các vi sinh vật đã phát triển các kiểu khác nhau trong quá trìn tiến hóa, ví dụ bằng cách tổng hợp và/hoặc hấp thu chất được gọi là các chất tan tương thích. Nếu các chất bảo vệ có trong tự nhiên này, ví dụ prolin hoặc glyxin betain, là có sẵn ở môi trường bên ngoài, thì sự hấp thu chúng, là nhanh hơn và thuận lợi hơn về mặt năng lượng, được ưu tiên hơn so với sự tổng hợp của vi sinh vật (Wood, Microbiology and Molecular Biology Reviews 63, 230-262 (1999)). Tải đồng vận ProP thuộc họ MFS (siêu họ chất tạo thuận lợi chủ yếu) và làm xúc tác sự hấp thu prolin, glyxin betain, prolin betain, ectoin và các cơ chất tương tự về mặt cấu trúc đồng vận chuyển với proton vào tế bào trong điều kiện tăng thẩm thấu. ProP là chất vận chuyển trước tiên mà các đặc tính cảm ứng thẩm thấu và điều hoà thẩm thấu được phát hiện trong hệ thống được hoàn nguyên. Miền đầu tận C có khả năng tạo ra motif cuộn xoắn được cuộn xoắn. Sự tạo thành miền này đóng vai trò quan trọng cho sự điều hoà thẩm thấu và có thể cho sự cảm ứng kích thích thẩm thấu (Culham et al., Journal of Molecular Recognition 13, 309-322 (2000)). ProP được hoạt hoá bằng sự tăng nội tại nồng độ ion kali và bằng sự dồn lại đại phân tử giống như liposom gắn protein (proteoliposome) bằng cách sử dụng polyetylen glycol (PEG) có chiều dài mạch khác nhau (Racher et al., Biochemistry 40, 7324-7333 (2001); Culham et al., Biochemistry 42, 410-420 (2003)). Mặc dù ProP có khả năng cảm ứng tình trạng stress thẩm thấu mà không cần các yếu tố bổ sung, tuy nhiên hoạt tính đầy đủ của chất mang *in vivo* đòi hỏi sự có mặt của protein ProQ tế bào chất (Kunte et al., Journal of Bacteriology 181, 1537-43 (1999)).

Mellies et al. mô tả việc loại bỏ gen proP ở *E. coli* với mục đích điều tra chức năng của proP trong bối cảnh bị sốc thẩm thấu (Journal of Bacteriology, Vol. 177, No. 1, p. 144-151).

Do đó, trước đây đã biết chất vận chuyển ProP của *E. coli* về chức năng của nó trong việc điều hoà thẩm thấu.

Các tác giả sáng chế nay đã bất ngờ phát hiện ra rằng theo sáng chế sự sản xuất L-axit amin bởi các vi sinh vật thuộc họ *Enterobacteriaceae* có thể được tăng lên bằng cách làm suy giảm gen *proP*.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là tạo ra quy trình và vi sinh vật, có khả năng sản xuất quá mức axit amin chứa lưu huỳnh, cụ thể hơn là sản xuất gia tăng L-metionin.

Đối tượng thứ nhất của sáng chế là quy trình sản xuất L-axit amin hoặc chất phụ gia thức ăn gia súc chứa L-axit amin bằng cách lên men vi sinh vật thuộc họ *Enterobacteriaceae*, khác biệt ở chỗ, sử dụng vi sinh vật có gen *proP* đã được làm suy giảm, trong đó suy giảm có nghĩa là hoạt tính hay nồng độ của protein ProP được giảm xuống còn từ 0 đến 75% hoạt tính hay nồng độ của protein trong vi sinh vật mà ko phải là tái tổ hợp đối với protein tương ứng. Vi sinh vật nêu trên sản xuất L-axit amin và tốt hơn nếu tiết L-axit amin vào môi trường xung quanh.

Ngoài ra, vi sinh vật làm cho L-axit amin tích lũy tốt hơn nếu trong môi trường và/hoặc bên trong tế bào (sự tích lũy), với ưu tiên đặc biệt là sự tích lũy trong môi trường.

Một đối tượng nữa của sáng chế là vi sinh vật thuộc họ *Enterobacteriaceae*, mang gen *proP* được làm suy giảm, khác biệt ở chỗ, vi sinh vật này sản xuất quá mức L-metionin và tốt hơn nếu tiết chất này vào môi trường này, tốt hơn nếu tạo ra sự tích lũy L-axit amin trong môi trường và/hoặc bên trong tế bào, tốt hơn nếu trong môi trường, khác biệt ở chỗ, vi sinh vật có một hoặc nhiều dấu hiệu được chọn từ nhóm sau đây:

a) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống vận chuyển CysPUWA thiosulphat/sulphat (EC 3.6.3.25),

b) polynucleotit được biểu hiện quá mức mã hoá CysH 3'-phosphoađenosin 5'-phosphosulphat ređuctaza (EC 1.8.4.8),

c) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của CysJI sulphit ređuctaza (EC 1.8.1.2),

d) polynucleotit được biểu hiện quá mức mã hoá CysK xystein synthaza A (EC 2.5.1.47),

e) polynucleotit được biểu hiện quá mức mã hoá CysM xystein synthaza B (EC 2.5.1.47),

f) polynucleotit được biểu hiện quá mức mã hoá CysE serin axetyltransferaza (EC 2.3.1.30),

g) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống phân cắt GcvTHP-Lpd glyxin (EC 2.1.2.10, EC 1.4.4.2, EC 1.8.1.4),

h) polynucleotit được biểu hiện quá mức mã hoá LipA lipoyl synthaza (EC 2.8.1.8),

i) polynucleotit được biểu hiện quá mức mã hoá LipB lipoyl-protein ligaza (EC 2.3.1.181),

j) polynucleotit được biểu hiện quá mức mã hoá SerA phosphoglyxerat đehyđrogenaza (EC 1.1.1.95),

k) polynucleotit được biểu hiện quá mức mã hoá SerB 3-phosphoserin phosphataza (EC 3.1.3.3),

l) polynucleotit được biểu hiện quá mức mã hoá SerC 3-phosphoserin/phosphohyđroxythreonin aminotransferaza (EC 2.6.1.52),

m) polynucleotit được biểu hiện quá mức mã hoá GlyA serin hyđroxymetyltransferaza (EC 2.1.2.1),

n) polynucleotit được biểu hiện quá mức mã hoá ThrA aspartokinaza I và homoserin đehyđrogenaza I (EC 2.7.2.4, EC 1.1.1.3),

o) polynucleotit được biểu hiện quá mức mã hoá LysC aspartat kinaza (EC 2.7.2.4),

p) polynucleotit được biểu hiện quá mức mã hoá Hom homoserin đehyđrogenaza (EC 1.1.1.3),

q) polynucleotit được biểu hiện quá mức mã hoá MetX homoserin O-axetyltransferaza (EC 2.3.1.31),

r) polynucleotit được biểu hiện quá mức mã hoá MetA homoserin O-succinyltransferaza (EC 2.3.1.46),

s) polynucleotit được biểu hiện quá mức mã hoá MetB xystathionin gama synthaza (EC 2.5.1.48),

t) polynucleotit được biểu hiện quá mức mã hoá AecD β-C-S-lyaza (EC 4.4.1.8, còn được gọi là beta-lyaza),

u) polynucleotit được biểu hiện quá mức mã hoá MetC xystathionin beta-lyaza (EC 4.4.1.8),

v) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza độc lập với MetE B12 (EC 2.1.1.14),

w) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza phụ thuộc MetH B12 (EC 2.1.1.13),

x) polynucleotit được biểu hiện quá mức mã hoá MetF metylen tetrahyđrofolat ređuctaza (EC 1.5.1.20),

y) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất BrnFE L-metionin ra khỏi *Corynebacterium glutamicum*,

z) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất YgaZH valin *Escherichia coli* (b2682, b2683),

aa) polynucleotit được biểu hiện quá mức mã hoá chất vận chuyển *Escherichia coli* YjeH giả định (b4141),

bb) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của PntAB pyriđin nucleotit transhyđrogenaza (EC 1.6.1.2),

cc) polynucleotit được biểu hiện quá mức mã hoá MetZ O-succinylhomoserin sulphhyđrylaza (EC 2.5.1.48),

dd) polynucleotit được biểu hiện quá mức mã hoá Pyc phosphoenolpyruvat carboxylaza (EC 4.1.1.31),

ee) polynucleotit được biểu hiện quá mức mã hoá RDL2 thiosulphat sulphurtransferaza (EC 2.8.1.1),

ff) polynucleotit được biểu hiện quá mức mã hoá thiosulphat—thiol sulphurtransferaza (EC 2.8.1.3),

gg) polynucleotit được biểu hiện quá mức mã hoá thiosulphat—đithiol sulphurtransferaza (EC 2.8.1.5),

hh) hoạt tính gia tăng của MetA homoserin O-succinyltransferaza (EC 2.3.1.46),

ii) hoạt tính gia tăng của CysE serin axetyltransferaza (EC 2.3.1.30),

jj) làm suy giảm gen *metJ* mã hoá chất điều hoà phiên mã sinh tổng hợp L-metionin (MetJ) (b3938, ECK3930),

kk) làm suy giảm gen pgi mã hoá glucoza-6-phosphat isomeraza (Pgi, EC No. 5.3.1.9) (b4025, ECK4017),

ll) làm suy giảm gen thrB mã hoá homoserin kinaza (ThrB, EC 2.7.1.39) (b0003, ECK0003),

mm) làm suy giảm gen metK mã hoá S-ađenosylmetionin synthaza (MetK, EC No. 2.5.1.6) (b2942, ECK2937),

nn) làm suy giảm gen dapA mã hoá đihyđrođipicolinat synthaza (DapA, EC No. 4.2.1.52) (b2478, ECK2474),

oo) làm suy giảm gen pck mã hoá phosphoenolpyruvat carboxykinaza (Pck, EC No. 4.1.1.49) (b3403, ECK3390),

pp) làm suy giảm gen purU mã hoá formyltetrahyđrofolat hydrolaza (PurU, EC No. 3.5.1.10) (b1232, ECK1227),

qq) làm suy giảm gen pykA mã hoá pyruvat kinaza II (PykA, EC No. 2.7.1.40) (b1854, ECK1855)

rr) làm suy giảm gen pykF mã hoá pyruvat kinaza I (PykF, EC 2.7.1.40) (b1676, ECK1672),

ss) làm suy giảm gen metQ mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0197, ECK0197),

tt) làm suy giảm gen metI mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0198, ECK0198),

uu) làm suy giảm gen mã hoá metN cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0199, ECK0199),

vv) làm suy giảm gen dcd mã hoá đeoxyxytiđin-5'-triphosphat đeaminaza (Dcd, EC No. 3.5.4.13) (b2065, ECK2059),

ww) làm suy giảm gen yncA mã hoá N-axyltransferaza giả định (YncA, Khám phá trao đổi chất: Metabolic Explorer WO2010/020681) (b1448, ECK1442),

xx) làm suy giảm gen rpoS mã hoá yếu tố RpoS sigma (b2741, ECK2736),

yy) làm suy giảm chất điều hòa protein MetJ.

zz) làm suy giảm S-adenosylmetionin synthaza MetK,

trong đó "làm suy giảm" có nghĩa là hoạt tính hay nồng độ của protein tương ứng được giảm xuống còn từ 0 đến 75% hoạt tính hay nồng độ của protein trong vi sinh vật mà không phải là tái tổ hợp đối với protein tương ứng và trong đó "biểu hiện quá mức", "tăng cường" hoặc "hoạt tính gia tăng" có nghĩa là hoạt tính của protein tương ứng được tăng lên so với chủng loại hoang dại với hệ số ít nhất 2 lần.

Tốt hơn nếu, vi sinh vật cho mức sản xuất tăng lên và tốt hơn nếu tiết L-axit amin mong muốn trong quá trình lên men quy trình, so với chủng ban đầu hoặc chủng cha mẹ được sử dụng mà không mang gen *proP* được làm suy giảm.

"Được làm suy giảm" theo sáng chế dùng để chỉ gen *proP* hoặc được biểu hiện ở mức thấp hoặc đã được tắt đi hoàn toàn.

Trong bản mô tả này, thuật ngữ "sự suy giảm" theo sáng chế dùng để chỉ việc làm giảm hoặc hạn chế hoạt tính hoặc nồng độ nội bào của một hoặc nhiều enzym

hoặc protein, cụ thể hơn trong bản mô tả này chất vận chuyển ProP, ở vi sinh vật, mà được mã hoá bởi ADN tương ứng, cụ thể hơn trong bản mô tả này là gen *proP*, bằng cách sử dụng ví dụ trình tự khởi đầu yếu hơn so với ở vi sinh vật hoặc chủng cha mẹ mà không phải là tái tổ hợp về enzym hoặc protein tương ứng, hoặc bằng cách sử dụng gen hoặc alen mã hoá enzym hoặc protein có hoạt tính thấp tương ứng, hoặc bất hoạt enzym hoặc protein hoặc khung đọc mở hoặc gen tương ứng, và, nếu thích hợp, kết hợp các biện pháp này.

Các biện pháp làm suy giảm thông thường làm giảm hoạt tính hoặc nồng độ của protein tương ứng từ 0 đến 75%, 0 đến 50%, 0 đến 25%, 0 đến 10%, hoặc 0 đến 5%, hoạt tính hoặc nồng độ của protein kiểu dại, hoặc hoạt tính hoặc nồng độ của protein ở vi sinh vật hoặc chủng cha mẹ mà không phải tái tổ hợp đối với enzym hoặc protein tương ứng. Vi sinh vật hoặc chủng cha mẹ không tái tổ hợp được hiểu là dùng để chỉ vi sinh vật mà được xử lý làm suy giảm hoặc hạn chế theo sáng chế.

Có thể đạt được sự làm suy giảm, ví dụ, bằng cách làm giảm hoặc làm hạn chế hoặc biểu hiện của gen hoặc khung đọc mở hoặc đặc tính xúc tác của protein enzym. Cả hai biện pháp có thể được dùng kết hợp, nếu thích hợp.

Sự biểu hiện gen có thể được làm giảm bằng quy trình nuôi cấy thích hợp, bằng cách cải biến di truyền (đột biến) cấu trúc tín hiệu của biểu hiện gen, hoặc là bằng công nghệ ARN đối nghĩa. Ví dụ về cấu trúc tín hiệu của biểu hiện gen là gen kìm hãm, gen hoạt hoá, gen điều khiển, trình tự khởi đầu, gen làm suy giảm, vị trí gắn kết ribosom, cođon bắt đầu và trình tự kết thúc. Thông tin về vấn đề này có thể được người có hiểu biết trung bình trong lĩnh vực kỹ thuật này tìm thấy, ví dụ, trong Jensen and Hammer (Biotechnology and Bioengineering 58: 191-195 (1998)), trong Carrier and Keasling (Biotechnology Progress 15: 58-64 (1999)), Franch and Gerdes (Current Opinion in Microbiology 3: 159-164 (2000)), Kawano et al. (Nucleic Acids Research 33(19), 6268-6276 (2005)) và trong các giáo trình đã biết thuộc lĩnh vực di truyền và sinh học phân tử, ví dụ giáo trình của Knippers ("Molekulare Genetik", 6th edition, Georg Thieme Verlag, Stuttgart, Germany, 1995) hoặc của Winnacker ("Gene und Klone", VCH Verlagsgesellschaft, Weinheim, Germany, 1990).

Các đột biến mà gây ra sự thay đổi hoặc sự giảm đặc tính xúc tác của protein enzym đã được bộc lộ trong tình trạng kỹ thuật; Ví dụ mà có thể được đề cập đến là các nghiên cứu của Qiu and Goodman (Journal of Biological Chemistry 272: 8611-8617 (1997)), Yano et al. (Proceedings of the National Academy of Sciences of the United States of America 95: 5511-5515 (1998)), Wente and Schachmann (Journal of Biological Chemistry 266: 20833-20839 (1991)). Có thể xem tổng quan trong các giáo trình đã biết của ngành di truyền và sinh học phân tử, ví dụ giáo trình của Hagemann ("Allgemeine Genetik", Gustav Fischer Verlag, Stuttgart, 1986).

Các đột biến mà có thể được xét đến là đồng hoán, dị hoán, cài xen và làm khuyết ít nhất một (1) cặp bazơ hoặc nucleotit. Tuỳ thuộc vào ảnh hưởng của thay thế axit amin gây ra do đột biến đến hoạt tính enzym, có thể kể đến đột biến sai nghĩa hoặc đột biến vô nghĩa. Đột biến sai nghĩa dẫn đến thay thế axit amin đã biết trong protein bằng một axit amin khác, cụ thể hơn là thay thế axit amin không bảo toàn. Quá trình này làm suy giảm chức năng hoặc hoạt tính của protein nêu trên, mà được giảm từ 0 đến 75%, 0 đến 50%, 0 đến 25%, 0 đến 10%, hoặc 0 đến 5%. Đột biến vô nghĩa gây ra codon dừng trong vùng mã hoá của gen và do đó làm cho quá trình dịch mã bị kết thúc sớm. Việc cài xen hoặc làm khuyết ít nhất một cặp bazơ trong gen dẫn đến đột biến dịch khung do một axit amin không đúng được kết hợp vào hoặc quá trình dịch mã bị kết thúc sớm. Nếu do quá trình đột biến tạo ra codon dừng trong vùng mã hoá, thì tương tự vậy quá trình dịch mã sẽ bị kết thúc sớm. Tương tự, việc làm khuyết ít nhất một (1) hoặc nhiều codon thường làm mất hoàn toàn hoạt tính của enzym. WO 03/074719 đề cập đến việc giảm biểu hiện gen bằng cách ức chế đột biến codon dừng trong vùng mã hoá bằng chất ức chế t-ARN thích hợp.

Các chỉ dẫn để tạo ra các đột biến như vậy là đã biết trong tình trạng kỹ thuật và có thể tham khảo trong các giáo trình di truyền và sinh học phân tử đã biết, ví dụ giáo trình của Knippers ("Molekulare Genetik", 6th edition, Georg Thieme Verlag, Stuttgart, Germany, 1995), của Winnacker ("Gene und Klone", VCH Verlagsgesellschaft, Weinheim, Germany, 1990) hoặc của Hagemann ("Allgemeine Genetik", Gustav Fischer Verlag, Stuttgart, 1986).

"Được làm suy giảm" và "biểu hiện ở mức thấp" dùng để chỉ đặc biệt là với gen *proP* rằng hoạt tính và/hoặc nồng độ của chất vận chuyển *ProP* được giảm đi bằng các biện pháp được thể hiện trên đây từ 0 đến 75%, 0 đến 50%, 0 đến 25%, 0 đến 10%, hoặc 0 đến 5%, hoạt tính và/hoặc nồng độ của protein kiểu dại, hoặc hoạt tính và/hoặc nồng độ của protein ở vi sinh vật hoặc chủng cha mẹ mà không tái tổ hợp chất vận chuyển ProP.

Theo sáng chế, tốt hơn nếu L-axit amin được sản xuất quá mức bởi vi sinh vật. "Sản xuất quá mức" theo sáng chế dùng để chỉ khả năng sản xuất L-axit amin được tăng lên đáng kể vượt hơn nhiều nhu cầu nội tại của vi sinh vật.

Tốt hơn nếu L-axit amin theo sáng chế là axit amin chứa lưu huỳnh, đặc biệt là L-metionin, L-xystein, L-xystin, L-homoxystein hoặc L-homoxystin, đặc biệt tốt hơn nếu L-metionin.

Vi sinh vật theo sáng chế và được sử dụng trong các quy trình theo sáng chế tốt hơn nếu có khác biệt là có khả năng chống chịu metionin tăng lên so với các vi sinh vật không có gen *proP* được làm suy giảm. Tốt hơn nếu, chúng có khả năng phát triển ngay cả ở nồng độ L-metionin là 50 hoặc 60 gam trên một lít, đặc biệt tốt hơn nếu ngay cả ở nồng độ L-metionin của 70, 75, 80, 90, hoặc 100 gam trên một lít, vì chúng có thể chống chịu nồng độ metionin cao như vậy. Số liệu về mức dung nạp liên quan đến L-metionin tốt hơn nếu được đo trên trên môi trường thạch tối thiểu có nồng độ L-metionin tương ứng.

Các chủng có kháng metionin như vậy có thể được phân lập bằng cách chọn lọc trên môi trường thạch tối thiểu chứa metionin, bắt đầu từ chủng *E. coli* đã sản xuất L-metionin.

Vi khuẩn kháng metionin theo sáng chế được tạo ra bằng cách sử dụng tốt hơn là các phương pháp chọn lọc như đã được mô tả trong tình trạng kỹ thuật, có thể tham khảo các phương pháp này ví dụ trong Miller (A Short Course in Bacterial Genetics, A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria (Cold Spring Harbor Laboratory Press, 1992)) hoặc trong "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D.C., USA).

Để thực hiện cụ thể đột biến ở gen *proP*, có thể áp dụng quy trình gây đột biến trực tiếp tại vị trí bằng cách sử dụng oligonucleotit đột biến (T. A. Brown: Gentechnologie fỹr Einsteiger [tiêu đề gốc: Gene Cloning and DNA Analysis – An Introduction], Spektrum Akademischer Verlag, Heidelberg, Germany, 1993) hoặc phản ứng chuỗi polymeaza (PCR), như mô tả trong sổ tay của Gait: Oligonucleotide synthesis: A Practical Approach (IRL Press, Oxford, UK, 1984) hoặc của Newton và Graham (PCR, Spektrum Akademischer Verlag, Heidelberg, 1994). Để điều khiển đột biến, ví dụ có thể áp dụng kit gây đột biến trực tiếp tại vị trí Quick Change Site-Directed Mutagenesis của Stratagene (Amsterdam, Netherlands). Bằng cách áp dụng các phương pháp này bao gồm khuếch đại gen *proP* như được mô tả trong tình trạng kỹ thuật nhờ trợ giúp của phản ứng chuỗi polymeaza (PCR) bắt đầu từ tổng AND được phân lập của chủng kiểu dại, nhân dòng gen nêu trên vào vật truyền plasmit thích hợp, và sau đó đưa ADN đến quy trình gây đột biến. Nhờ công nghệ "GeneSOEing" (cắt ghép gen bằng cách kéo dài đoạn lặp Gene Splicing by Overlap Extension, Horton, Molecular Biotechnology 3: 93-98 (1995)), thậm chí có thể thu được đột biến điểm bằng PCR. Cũng có thể áp dụng quá trình tổng hợp gen *de novo* (ví dụ bằng GENEART AG, Regensburg, Germany) của trình tự nucleotit để tạo ra đột biến ở gen *proP*. Các đột biến tạo ra có thể được xác định và kiểm tra bằng cách xác định trình tự ADN, ví dụ bằng phương pháp của Sanger et al. (Proceedings of the National Academy of Science USA 74 (12): 5463-5467, 1977).

Các alen tạo ra này có thể được kết hợp vào nhiễm sắc thể của chủng thích hợp, ví dụ bằng cách biến nạp và phương pháp thay thế gen hoặc alen.

Một phương pháp thông thường, được mô tả bởi Hamilton et al. (Journal of Bacteriology 171, 4617 – 4622 (1989)), là phương pháp thay thế gen với sự trợ giúp của dẫn xuất pMAK705 sao chép một cách có điều kiện pSC101 hoặc của pKO3 (Link et al., Journal of Bacteriology 179: 6228-6237). Tương tự, có thể áp dụng các phương pháp khác như phương pháp được mô tả trong tình trạng kỹ thuật, ví dụ phương pháp của Martinez-Morales et al. (Journal of Bacteriology 1999, 7143-7148 (1999)) hoặc phương pháp của Boyd et al. (Journal of Bacteriology 182, 842-847 (2000)).

Các phương pháp thông thường khác bao gồm việc kết hợp thông qua các trình tự nằm chặn tương đồng ngắn là mảnh ADN được tạo ra bằng PCR hoặc tổng hợp gen trực tiếp vào nhiễm sắc thể nhờ trợ giúp của Lambda Red recombinaza, hoặc tiến hành thay thế (Proc. Natl. Acad. Sci. USA 97(12), 6640-6645 (2000); Nature Genetics 20, 123 – 128, 1998).

Tương tự có thể chuyển, bằng cách liên hợp hoặc tải nạp, các alen tạo ra này vào các chủng khác nhau.

Các trình tự axit nucleic *proP* có thể được tìm thấy trong dữ liệu của National Center for Biotechnology Information (NCBI), the National Library of Medicine (Bethesda, MD, USA), dữ liệu trình tự nucleotit của European Molecular Biologies Laboratories (EMBL, Heidelberg, Germany và Cambridge, UK) hoặc nguồn dữ liệu ADN của Nhật (DDBJ, Mishima, Japan).

Cho mục đích minh họa, trình tự đã biết của gen *proP* của *Escherichia coli* được liệt kê trong SEQ ID NO: 1, và trình tự đã biết của gen *proP* của *Salmonella enterica* và *Shigella sonnei* tương tự thuộc vào họ *Enterobacteriaceae*, được liệt kê trong SEQ ID NO: 3 và SEQ ID NO: 5. Các protein được mã hoá bởi các khung đọc này được liệt kê bằng SEQ ID NO: 2, SEQ ID NO: 4 và SEQ ID NO: 6. Các trình tự nucleotit nữa cho gen *proP* được tìm thấy, ví dụ, ở các loài *Enterobacteriaceae* sau: *Shigella boydii* (Số truy cập: NC_007613); *Shigella flexneri* (Số truy cập: NC_004741); *Shigella dysenteriae* (Số truy cập: NC_007606); *Citrobacter rodentium* (Số truy cập: NC_013716); *Erwinia pyrifoliae* (Số truy cập: NC_012214); *Klebsiella pneumoniae* (Số truy cập: NC_011283).

Protein được mã hoá bởi gen *proP* của *Escherichia coli* K12 còn được gọi là chất vận chuyển MFS nhạy cảm thẩm thấu ProP hoặc là tải đồng vận hòa tan proton/tương thích (Số truy cập: 11612 (Vùng: 4328525-4330027); các tên gen khác: *b4111*, *ECK4104*); Grothe et al., Journal of Bacteriology 166, 253-259 (1986); Racher et al., Biochemistry 38, 1676-1684 (1999); Wood, Methods in Enzymology 428, 77-107 (2007)).

Mô tả vắn tắt các hình vẽ

Fig.1 thể hiện bản đồ plasmit pCC3.

Fig.2 thể hiện bản đồ plasmit pME-RDL2a.

Fig.3 thể hiện bản đồ plasmit pMAK_proP-M8.

Mô tả chi tiết sáng chế

Theo một phương án được ưu tiên theo sáng chế, gen *proP* được làm suy giảm là gen có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, dựa trên tốt hơn nếu trình tự polynucleotit đầy đủ của SEQ ID NO: 1, 3, 5 hoặc 7, đặc biệt tốt hơn nếu dựa trên trình tự polynucleotit đầy đủ của SEQ ID NO: 1.

Các tài liệu tham khảo trong lĩn vực liên quan đến các gen *proP* và các khung đọc mở của các gen *proP* nêu trên đã được chỉ ra trên đây. Các gen này có thể được sử dụng theo sáng chế theo cách thích hợp để làm suy giảm gen *proP*. Ngoài ra, các alen của gen hoặc khung đọc mở mà tạo ra từ sự suy biến mã di truyền hoặc tính đến đột biến có nghĩa trung tính về mặt chức năng cũng có thể được sử dụng để tạo ra gen *proP* được làm suy giảm. Việc sử dụng gen nội sinh hoặc khung đọc mở nội sinh là được ưu tiên.

Các alen của gen *proP* chứa đột biến có nghĩa trung tính về mặt chức năng bao gồm, ngoài các alen khác, các alen tạo ra không nhiều hơn 40 hoặc không nhiều hơn 30 hoặc không nhiều hơn 20, tốt hơn nếu không nhiều hơn 10 hoặc không nhiều hơn 5, rất đặc biệt tốt hơn nếu không nhiều hơn 3 hoặc không nhiều hơn 2, hoặc chính xác một thay thế axit amin bảo toàn trong protein được mã hoá bởi chúng.

Trong trường hợp axit amin thơm, thay thế bảo toàn là các thay thế trong đó phenylalanin, tryptophan và tyrosin được thế cho nhau. Trong trường hợp axit amin kỵ nước, các thay thế bảo toàn là các thay thế trong đó leuxin, isoleuxin và valin được thế cho nhau. Trong trường hợp axit amin có cực, thay thế bảo toàn là các thay thế trong đó glutamin và asparagin được thay thế cho nhau. Trong trường hợp axit amin kiềm, thay thế bảo toàn là các thay thế trong đó arginin, lysin và histiđin được thế cho nhau. Trong trường hợp axit amin có tính axit, thay thế bảo toàn là các thay thế trong đó axit aspartic và axit glutamic được thay thế cho nhau. Trong trường hợp axit amin chứa các

nhóm hyđroxyl, thay thế bảo toàn là các thay thế trong đó serin và threonin được thay thế cho nhau.

Tương tự, cũng có thể sử dụng các trình tự nucleotit mã hóa các biến thể của các protein được đề cập, mà ngoài ra còn chứa đoạn kéo dài hoặc cắt cụt bởi ít nhất một (1) axit amin ở đầu tận N hoặc C. Đoạn kéo dài hoặc cắt cụt này tạo thành không nhiều hơn 10, 5, 3 hoặc 2 axit amin hoặc gốc axit amin.

Các biến thể thích hợp cũng bao gồm các biến thể mã hoá protein trong đó ít nhất một (1) axit amin đã được cài xen (cài xen) hoặc được loại bỏ (khuyết đoạn). Số lượng tối đa các các cải biến như vậy được gọi là indel có thể ảnh hưởng 2, 3, 5, nhưng không có trường hợp nào nhiều hơn nhiều hơn 10 axit amin.

Các biến thể thích hợp cũng bao gồm các biến thể thu được bằng cách lai, đặc biệt là trong điều kiện nghiêm ngặt, bằng cách sử dụng SEQ ID NO: 1, SEQ ID NO: 3 hoặc SEQ ID NO: 5 hoặc một phần của chúng, đặc biệt là vùng mã hoá và/hoặc trình tự bổ trợ với chúng.

Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này có thể tìm thấy các chỉ dẫn liên quan đến việc nhận diện trình tự ADN bằng cách lai trong các tài liệu trong số đó có sổ tay "The DIG System User's Guide for Filter Hybridization" từ Boehringer Mannheim GmbH (Mannheim, Germany, 1993) và trong Liebl et al. (International Journal of Systematic Bacteriology 41: 255-260 (1991)). Quá trình lai diễn ra trong điều kiện nghiêm ngặt, tức là có thể nói tạo ra chỉ các thể lai trong đó mẫu dò và trình tự đích, tức là các polynucleotit được xử lý bằng mẫu dò này, là ít nhất 70% đồng nhất. Độ nghiêm ngặt của điều kiện lai, bao gồm các bước rửa, đã được biết đến là chịu ảnh hưởng hoặc được xác định bằng cách thay đổi thành phần dung dịch đệm, nhiệt độ và nồng độ muối. Phản ứng lai thường được thực hiện ở điều kiện nghiêm ngặt tương đối thấp so với các bước rửa (Hybaid Hybridisation Guide, Hybaid Limited, Teddington, UK, 1996).

Ví dụ, có thể sử dụng dung dịch đệm 5× SSC ở nhiệt độ vào khoảng 50°C - 68°C cho phản ứng lai. Ở đây, các mẫu dò cũng có thể lai với các polynucleotit có độ đồng nhất ít hơn 70% so với trình tự của mẫu dò. Các thể lai như vậy là kém ổn định

và được loại bỏ bằng cách rửa trong điều kiện nghiêm ngặt. Điều này có thể đạt được, ví dụ, bằng cách giảm nồng độ muối xuống 2× SSC và, nếu thích hợp, sau đó 0,5× SSC (The DIG System User's Guide for Filter Hybridisation, Boehringer Mannheim, Mannheim, Germany, 1995), với nhiệt độ được thiết lập vào khoảng 50°C – 68°C, khoảng 52°C – 68°C, khoảng 54°C – 68°C, khoảng 56°C – 68°C, khoảng 58°C – 68°C, khoảng 60°C – 68°C, khoảng 62°C – 68°C, khoảng 64°C – 68°C, khoảng 66°C – 68°C. Khoảng nhiệt độ được ưu tiên vào khoảng 64°C – 68°C hoặc khoảng 66°C – 68°C. Tuỳ chọn có thể giảm nồng độ muối đến nồng độ tương ứng với 0,2× SSC hoặc 0,1× SSC. Bằng cách tăng dần nhiệt độ lai ở các bước vào khoảng 1 – 2°C từ 50°C đến 68°C, có thể phân lập mảnh polynucleotit mà là, ví dụ, ít nhất 70% hoặc ít nhất 80% hoặc ít nhất 90%, ít nhất 91%, ít nhất 92%, ít nhất 93%, ít nhất 94%, ít nhất 95%, ít nhất 96%, ít nhất 97%, ít nhất 98%, hoặc ít nhất 99%, đồng nhất với trình tự của mẫu dò được sử dụng hoặc trình tự nucleotit được thể hiện trong SEQ ID NO: 1, SEQ ID NO: 3 hoặc SEQ ID NO: 5. Có thể xem các chỉ dẫn bổ sung liên quan đến điều kiện lai trên thị trường ở dạng "các bộ kit" (ví dụ DIG Easy Hyb from Roche Diagnostics GmbH, Mannheim, Germany, Catalogue No. 1603558).

Gen *proP* được làm suy giảm mã hoá protein chất vận chuyển ProP tốt hơn nếu có trình tự axit amin ít nhất 85%, đặc biệt là ít nhất 90%, tốt hơn nếu ít nhất 95%, đặc biệt tốt hơn nếu ít nhất 98%, ít nhất 99%, hoặc 100%, đồng nhất với trình tự axit amin của SEQ ID NO:2, SEQ ID NO: 4, SEQ ID NO: 6 hoặc SEQ ID NO: 8, tốt hơn nếu với trình tự axit amin của SEQ ID NO: 2, tốt hơn nếu độ đồng nhất nêu trên là trên toàn bộ chiều dài của (các) trình tự được nêu ra. Tốt hơn nếu protein chất vận chuyển ProP chứa hoặc về cơ bản có chiều dài 500 axit amin, với chiều dài 500 axit amin là được ưu tiên. Protein ProP được ưu tiên rất đặc biệt nếu chứa hoặc có trình tự axit amin của SEQ ID NO: 2, mà trình tự này có thể tuỳ ý bao gồm không nhiều hơn 40, không nhiều hơn 30, tốt hơn nếu không nhiều hơn 20, không nhiều hơn 10, không nhiều hơn 5, không nhiều hơn 3, không nhiều hơn 2, đặc biệt tốt hơn nếu không nhiều hơn một, thay thế axit amin bảo toàn. Các thay thế axit amin bảo toàn về cơ bản không làm biến đổi hoạt tính của chất vận chuyển ProP, tức là theo sáng chế tốt hơn nếu hoạt

tính nêu trên được biến đổi bởi thay thế axit amin bảo toàn nêu trên không nhiều hơn 10%, tính theo trình tự bắt đầu.

Về điểm này, thuật ngữ "về cơ bản có chiều dài 500 axit amin" tính đến trường hợp là có sự cài xen hoặc làm khuyết một (1) hoặc nhiều, không nhiều hơn 10, 9, 8, 7, 6, 5, 4, 3 hoặc 2, axit amin trong polypeptit hoặc ở đầu tận N hoặc C của polypeptit nêu trên dẫn đến sự thay đổi nhẹ chiều dài của polypeptit được mã hoá trong các loài hoặc chủng khác nhau thuộc họ *Enterobacteriaceae*. Một ví dụ của polypeptit này là protein ProP của *Erwinia pyrifoliae*. Trong trường hợp này, chiều dài của polypeptit này (xem SEQ ID No:8) là 501 axit amin.

Theo các phương án được ưu tiên theo sáng chế, sự làm suy giảm của gen *proP* của SEQ ID NO: 1 đạt được bởi gen có đột biến bất kỳ trong số các đột biến sau:

a) thay thế nucleobazơ guanin ở vị trí 971 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ thymin;

b) thay thế nucleobazơ thymin ở vị trí 1399 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ xytosin;

c) thay thế nucleobazơ guanin ở vị trí 1234 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ thymin;

d) làm khuyết nucleobazơ ađenin ở vị trí 854 của trình tự polynucleotit của SEQ ID NO: 1;

e) làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của trình tự polynucleotit của SEQ ID NO:1;

f) cài xen nucleobazơ xytosin ở vị trí 842 của trình tự polynucleotit của SEQ ID NO:1;

g) cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của trình tự polynucleotit của SEQ ID NO:1;

h) cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của trình tự polynucleotit của SEQ ID NO:1.

Do đó, theo các phương án được ưu tiên theo sáng chế, gen *proP* được làm suy giảm là polynucleotit có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt tốt hơn nếu ít nhất 98%, 99% hoặc 100%, tính theo polynucleotit của SEQ ID NO: 1, tốt hơn nếu tính theo trình tự đầy đủ của polynucleotit của SEQ ID NO: 1, và mà ngoài ra nhất thiết có một hoặc nhiều, tốt hơn nếu chính xác một, đột biến sau so với polynucleotit của SEQ ID NO: 1:

a) thay thế bộ ba mã hoá L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b) thay thế bộ ba mã hoá L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c) thay thế bộ ba mã hoá L-axit glutamic ở vị trí 412 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá cođon dừng;

d) làm khuyết nucleobazơ ađenin ở vị trí 854 của gen *proP* của SEQ ID NO:1;

e) làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của gen *proP* của SEQ ID NO:1;

f) cài xen nucleobazơ xytosin ở vị trí 842 của gen *proP* của SEQ ID NO:1;

g) cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của gen *proP* của SEQ ID NO:1;

a) cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của gen *proP* của SEQ ID NO:1.

Việc nhận diện được nêu là dựa trên polynucleotit của SEQ ID NO: 1 ở bản mô tả này trong mỗi trường hợp dùng để chỉ trình tự polynucleotit mà không có một hoặc nhiều đột biến bắt buộc được đề cập trên đây.

"Vị trí tương đương" được nhận diện một cách dễ dàng bằng cách so sánh trình tự axit amin bằng cách cách sắp xếp, ví dụ nhờ chương trình Clustal W (Thompson et al., Nucleic Acids Research 22, 4637-4680 (1994)) hoặc nhờ chương trình MAFFT (Katoh et al., Genome Information 2005; 16(1), 22-33).

Tốt hơn nếu vi sinh vật theo sáng chế và được sử dụng trong các quy trình theo sáng chế thuộc họ *Enterobacteriaceae* là vi khuẩn thuộc giống *Escherichia*, *Erwinia*, *Providencia* hoặc *Serratia*, đặc biệt là giống *Escherichia*. Đực ưu tiên đặc biệt là *Escherichia coli*.

Trong các quy trình theo sáng chế dùng để sản xuất L-axit amin, đặc biệt là L-axit amin chứa lưu huỳnh, tốt hơn nếu quá trình lên men được thực hiện trong môi trường chứa nguồn lưu huỳnh vô cơ. Nguồn lưu huỳnh có thể được sử dụng là muối của axit đithiosulfuric (thiosulphat), nếu thích hợp cùng với các nguồn lưu huỳnh khác như sulphat, sulphit hoặc đithionit, ví dụ (xem ví dụ đơn EP 11151526.8).

Tốt hơn nếu L-axit amin được làm tích lũy trong môi trường lên men lỏng thu được và sau đó, khi thích hợp, được phân lập, được gom và/hoặc được tinh chế.

L-axit amin cũng có thể được phân lập hoặc được gom cùng với các thành phần từ môi trường lên men lỏng và/hoặc sinh khối.

Lượng vi khuẩn được phân lập hoặc của quá trình lên men sử dụng nó xét về một hoặc nhiều thông số được chọn từ nhóm gồm nồng độ sản phẩm (sản phẩm trên thể tích), hiệu suất sản phẩm (sản phẩm tạo ra trên mỗi nguồn cacbon tiêu thụ) và sự tạo thành sản phẩm (sản phẩm tạo ra trên mỗi đơn vị thể tích và thời gian), hoặc các thông số quy trình khác nữa và kết hợp của các thông số này, tốt hơn nếu được cải thiện ít nhất 0,5%, ít nhất 1%, ít nhất 1,5% hoặc ít nhất 2%, tính theo chủng ban đầu hoặc chủng cha mẹ hoặc quy trình lên men bằng cách sử dụng các chủng này.

Trong quy trình theo sáng chế, để sản xuất L-axit amin vi khuẩn có thể được nuôi cấy một cách liên tục – như mô tả, ví dụ, trong PCT/EP2004/008882 - hoặc một cách gián đoạn trong quy trình gián đoạn (nuôi cấy theo mẻ) hoặc trong quy trình cấp gián đoạn hoặc quy trình cấp gián đoạn lặp lại. Tóm tắt về phần bản chất chung các phương pháp nuôi cấy đã biết có trong giáo trình của Chmiel (Bioprozesstechnik 1.

Einfỹhrung in die Bioverfahrenstechnik (Gustav Fischer Verlag, Stuttgart, 1991)) hoặc trong giáo trình của Storhas (Bioreaktoren und periphere Einrichtungen (Vieweg Verlag, Brunswick/Wiesbaden, 1994)).

Môi trường nuôi cấy hoặc môi trường lên men được sử dụng phải đáp ứng thích hợp nhu cầu của chủng cụ thể. Mô tả về môi trường nuôi cấy các vi sinh vật khác nhau được nêu trong Manual of Methods for General Bacteriology of the American Society for Bacteriology (Washington D.C., USA, 1981). Các thuật ngữ môi trường nuôi cấy và môi trường lên men hoặc môi trường là dùng thay được cho nhau.

Nguồn cacbon mà có thể được sử dụng là các đường và hyđrat cacbon như, ví dụ, glucoza, sucroza, lactoza, fructoza, maltoza, mật đường, các dung dịch chứa sucroza từ các quy trình sản xuất đường từ cây củ cải đường hoặc đường mía, tinh bột, dịch thuỷ phân tinh bột và xenluloza, dầu và các chất béo như, ví dụ, dầu đậu nành, dầu hướng dương, dầu từ loài thân củ mọc dưới đất và dầu dừa, các axit béo như, ví dụ, axit palmitic, axit stearic và axit linoleic, các rượu như, ví dụ, glyxerol, metanol và etanol, và các axit hữu cơ như, ví dụ, axit axetic. Các chất này có thể được sử dụng đơn lẻ hoặc dưới dạng hỗn hợp.

Nguồn nitơ có thể được sử dụng là các hợp chất hữu cơ chứa nitơ như pepton, dịch chiết nấm men, nước thịt, dịch chiết malt, dịch chiết ngô đặc, bột đậu nành và ure, hoặc các hợp chất vô cơ như amoni sulphat, amoni clorua, amoni phosphat, amoni cacbonat và amoni nitrat. Nguồn nitơ có thể được sử dụng đơn lẻ hoặc dưới dạng hỗn hợp.

Nguồn phospho mà có thể được sử dụng là axit phosphoric, kali đihyđro phosphat hoặc đikali hyđro phosphat hoặc các muối chứa natri tương ứng.

Tốt hơn nữa nếu môi trường nuôi cấy chứa các muối, ví dụ ở dạng clorua, của kim loại như, ví dụ, natri, kali, magie, canxi và sắt, như, ví dụ, magie sulphat hoặc sắt sulphat, là các chất cần thiết cho sự phát triển. Cuối cùng, các chất sinh trưởng thiết yếu như axit amin, ví dụ homoserin, và các vitamin, ví dụ cobalamin, thiamin, biotin hoặc axit pantothenic, có thể được sử dụng thêm ngoài các chất được đề cập trên đây.

Hơn nữa, các tiền chất thích hợp của axit amin cụ thể có thể được cho vào môi trường nuôi cấy.

Các nguyên liệu nêu trên có thể được cho vào môi trường nuôi cấy làm một lần hoặc được cấp theo cách thích hợp trong quá trình nuôi cấy.

Độ pH của môi trường nuôi cấy được khống chế bằng cách sử dụng các hợp chất kiềm theo cách thích hợp như natri hyđroxit, kali hyđroxit, amoniac hoặc trong nước amoniac, hoặc các hợp chất axit như axit phosphoric hoặc axit sulfuric. Độ pH thường được điều chỉnh đến giá trị nằm trong khoảng từ 6,0 đến 9,0, tốt hơn nếu nằm trong khoảng từ 6,5 đến 8. Để khống chế quá trình tạo bọt, có thể sử dụng các chất chống tạo bọt như, ví dụ, este của axit béo và polyglycol. Để duy trì độ ổn định của plasmit, có thể đưa vào môi trường các chất có tác động chọn lọc thích hợp như, ví dụ, các chất kháng sinh. Để duy trì điều kiện ưa khí, đưa oxy hoặc các hỗn hợp khí chứa oxy như, ví dụ, không khí vào môi trường nuôi cấy. Tương tự có thể sử dụng các chất lỏng được làm giàu hyđro peroxit. Quá trình lên men được thực hiện, nếu thích hợp, ở áp suất nâng cao, ví dụ ở áp suất nằm trong khoảng từ 0,03 đến 0,2 MPa. Nhiệt độ môi trường nuôi cấy thông thường nằm trong khoảng từ 20°C đến 45°C và tốt hơn nếu nằm trong khoảng từ 25°C đến 40°C. Trong các quy trình gián đoạn, quá trình nuôi cấy được tiếp tục cho đến khi tạo ra lượng axit amin mong muốn tối đa. Thông thường đạt được mục đích này trong thời gian từ 10 giờ đến 160 giờ. Trong các quy trình liên tục, có thể mất thời gian nuôi cấy lâu hơn.

Môi trường lên men thích hợp được mô tả trong các tài liệu ví dụ như US 6,221,636, US 5,840,551, US 5,770,409, US 5,605,818, US 5,275,940 và US 4,224,409.

Phương pháp xác định L-axit amin đã được bộc lộ trong tình trạng kỹ thuật. Ví dụ, có thể thực hiện phép phân tích, như được mô tả trong Spackman et al. (Analytical Chemistry, 30, (1958), 1190), bằng cách sắc ký trao đổi ion với quá trình tạo dẫn xuất ninhyđrin sau đó, hoặc có thể được thực hiện bằng HPLC đảo pha, như mô tả trong Lindroth et al. (Analytical Chemistry (1979) 51: 1167-1174).

Sau đó tốt hơn nếu môi trường lên men lỏng được tạo ra theo cách này được xử lý thành sản phẩm rắn hoặc lỏng.

Môi trường lên men lỏng dùng để chỉ môi trường lên men trong đó vi sinh vật đã được nuôi cấy trong một thời gian nhất định và ở nhiệt độ nhất định. Tốt hơn nếu môi trường lên men hoặc môi trường được sử dụng trong quá trình lên men bao gồm chất hay hợp phần bất kỳ đảm bảo quá trinh nhân lên của vi sinh vật nêu trên và tạo ra axit amin mong muốn.

Khi hoàn thành quá trình lên men, môi trường lên men lỏng thu được theo đó chứa a) sinh khối của vi sinh vật, được tạo ra do sự nhân tế bào vi sinh vật nêu trên, b) axit amin mong muốn được tạo ra trong quá trình lên men, c) các sản phẩm phụ hữu cơ được tạo ra trong quá trình lên men, và d) các chất cấu thành các môi trường lên men/môi trường lên men được sử dụng hoặc của các nguyên liệu như, ví dụ, các vitamin như biotin, axit amin như homoserin hoặc các muối như magie sulphat, mà không được tiêu thụ trong quá trình lên men nêu trên.

Các sản phẩm phụ hữu cơ bao gồm các chất được tạo ra bởi các vi sinh vật được sử dụng trong quá trình lên men tuỳ ý ngoài L-axit amin cụ thể mong muốn và mà chất này có thể được tiết ra. Các chất bao gồm L-axit amin chiếm ít hơn 30%, 20% hoặc 10%, so với axit amin mong muốn. Ngoài ra các chất này bao gồm các axit hữu cơ mang từ một đến ba nhóm carboxyl, ví dụ như axit axetic, axit lactic, axit xitric, axit malic hoặc axit fumaric. Cuối cùng, các chất này cũng bao gồm các đường ví dụ như trehaloza.

Các môi trường lên men lỏng điển hình mà thích hợp cho mục đích công nghiệp và được ưu tiên theo sáng chế có hàm lượng axit amin nằm trong khoảng từ 40 g/kg đến 180 g/kg hoặc 50 g/kg đến 150 g/kg. Hàm lượng sinh khối (sinh khối khô) thông thường nằm trong khoảng từ 20 đến 50 g/kg.

Theo đó, sáng chế còn đề cập đến các polynucleotit có mức đồng nhất ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98, 99 hoặc 100%, tính theo trình tự của SEQ ID NO: 1, tốt hơn nếu mức đồng nhất được nêu ra liên quan đến trình tự đầy đủ của SEQ ID NO: 1, khác biệt ở chỗ polynucleotit này

nhất thiết có một hoặc nhiều đột biến so với polynucleotit của SEQ ID NO: 1, được chọn từ:

a) thay thế bộ ba mã hoá L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b) thay thế bộ ba mã hoá L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c) thay thế bộ ba mã hoá L-axit glutamic ở vị trí 412 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá cođon dừng;

d) làm khuyết nucleobazơ ađenin ở vị trí 854 của gen *proP* của SEQ ID NO:1;

e) làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của gen *proP* của SEQ ID NO:1;

f) cài xen nucleobazơ xytosin ở vị trí 842 của gen *proP* của SEQ ID NO:1;

g) cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của gen *proP* của SEQ ID NO:1;

h) cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của gen *proP* của SEQ ID NO:1.

Các nhận diện nêu trên dựa trên polynucleotit của SEQ ID NO: 1 trong mỗi trường hợp ở đây liên quan đến trình tự polynucleotit mà không chứa một hoặc nhiều đột biến bắt buộc được đề cập trên đây.

Ưu tiên đặc biệt là một hoặc nhiều đột biến bắt buộc được chọn từ:

a) thay thế nucleobazơ guanin ở vị trí 971 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ thymin;

b)thay thế nucleobazơ thymin ở vị trí 1399 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ xytosin;

c)thay thế nucleobazơ guanin ở vị trí 1234 hoặc vị trí tương đương của trình tự polynucleotit của SEQ ID NO:1 bằng nucleobazơ thymin;

d)làm khuyết nucleobazơ ađenin ở vị trí 854 của trình tự polynucleotit của SEQ ID NO: 1;

e)làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của trình tự polynucleotit của SEQ ID NO:1;

f) cài xen nucleobazơ xytosin ở vị trí 842 của trình tự polynucleotit của SEQ ID NO:1;

g)cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của trình tự polynucleotit của SEQ ID NO:1;

h)cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của trình tự polynucleotit của SEQ ID NO:1.

Theo sáng chế, các polynucleotit được đề cập trên đây theo sáng chế tốt hơn nếu được khác biệt bằng cách lai với một hoặc nhiều polynucleotit bổ trợ với SEQ ID NO: 1, 3 hoặc 5, tốt hơn nếu bổ trợ với SEQ ID NO: 1, tốt hơn nếu trong điều kiện lai nghiêm ngặt, tốt hơn nếu điều kiện nghiêm ngặt này đạt được bằng bước rửa trong đó khoảng nhiệt độ nằm trong khoảng từ 64°C đến 68°C và nồng độ muối của dung dịch đệm nằm trong khoảng từ 2xSSC đến 0,1xSSC.

Đặc biệt được ưu tiên nếu các polynucleotit theo sáng chế có mức đồng nhất trình tự ít nhất 90%, tốt hơn nếu ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với các polynucleotit của SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21 hoặc SEQ ID NO: 23.

Rất đặc biệt được ưu tiên nếu các polynucleotit theo sáng chế là và/hoặc chứa các polynucleotit của SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21 và SEQ ID NO: 23.

Sáng chế còn đề xuất các vật truyền chứa các polynucleotit theo sáng chế.

Theo đó, sáng chế còn đề xuất các polypeptit có mức đồng nhất ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự của SEQ ID NO: 2, khác biệt ở chỗ, polypeptit này nhất thiết có một hoặc nhiều đột biến so với polypeptit của SEQ ID NO: 2, được chọn từ:

a. thay thế L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b. thay thế L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c. làm khuyết 17 axit amin từ vị trí 392 đến vị trí 408 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin;

d. làm khuyết lên đến 420 axit amin ở đầu tận C, tính theo trình tự của SEQ ID NO: 2, tốt hơn nếu làm khuyết 88, 163, 202, 212 hoặc 420 axit amin ở đầu tận C của trình tự axit amin của SEQ ID NO: 2.

Các nhận diện được nêu ra ở đây dùng để chỉ các đột biến của a và b so với trình tự đầy đủ của SEQ ID NO: 2, đối với các đột biến c và d trong mỗi trường hợp so với vùng nhỏ của trình tự của SEQ ID NO: 2 mà không có vùng khuyết đoạn được chỉ ra.

Các polypeptit được ưu tiên theo sáng chế là các polypeptit có mức đồng nhất trình tự ít nhất 90%, tốt hơn nếu ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với các polypeptit của SEQ ID NO: 10, 12, 14, 16, 18, 20, 22 hoặc 24.

Các polypeptit đặc biệt được ưu tiên theo sáng chế là và/hoặc chứa các polypeptit của SEQ ID NO: 10, 12, 14, 16, 18, 20, 22 hoặc 24.

Sáng chế còn đề cập đến các vi sinh vật tái tổ hợp mang các polynucleotit và/hoặc các vật truyền và/hoặc các polypeptit theo sáng chế.

Tốt hơn nếu các vi sinh vật theo sáng chế và các vi sinh vật được sử dụng trong các quy trình theo sáng chế có hoạt tính enzym aspartat kinaza (EC 2.7.2.4) tăng lên, với ưu tiên cho các alen tính kháng cơ chế ức chế ngược. *E. coli* có ba aspartat kinaza khác nhau được mã hoá bởi gen *thrA*, *metL* hoặc *lysC*. Theo sáng chế, ưu tiên đặc biệt là có hoạt tính ThrA aspartat kinaza tăng lên.

Còn tốt hơn nữa nếu các vi sinh vật theo sáng chế và các vi sinh vật được sử dụng trong các quy trình theo sáng chế được khác biệt ở chỗ có hoạt tính MetA homoserin O-succinyltransferaza (EC 2.3.1.46) và/hoặc CysE serin axetyltransferaza (EC 2.3.1.30) được tăng lên.

Ngoài ra có thể làm tăng sinh tổng hợp L-metionin bằng cách làm suy giảm hoặc làm khuyết protein điều hoà MetJ được mã hoá bởi gen *metJ*. MetJ là chất ức chế chủ yếu quá trình sinh tổng hợp L-metionin ở *E. coli*. Theo đó, ngoài ra theo sáng chế còn ưu tiên là gen *metJ* được làm suy giảm.

Các vi sinh vật theo sáng chế và các vi sinh vật được sử dụng trong các quy trình theo sáng chế còn tốt hơn nữa nếu được khác biệt ở chỗ có hoạt tính MetK S-ađenosylmetionin synthaza (EC 2.5.1.6) giảm.

Ngoài ra có thể tốt hơn nếu quá trình sản xuất axit amin chứa lưu huỳnh bằng cách sử dụng vi khuẩn thuộc họ *Enterobacteriaceae* để tăng cường một hoặc nhiều enzym trong các quá trình sinh tổng hợp axit amin đã biết hoặc enzym của quá trình trao đổi chất bổ sung (anaplerotic) hoặc enzym để sản xuất nicotinamit ađenin đinucleotit phosphat giảm hoặc enzym của quá trình đường phân hoặc enzym PTS hoặc enzym của quá trình trao đổi chất lưu huỳnh hoặc để tăng hoạt tính của chúng.

Theo các phương án được ưu tiên hơn nữa, vi khuẩn sản xuất L-metionin có một hoặc nhiều đặc điểm được chọn từ nhóm bao gồm:

a) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống vận chuyển CysPUWA thiosulphat/sulphat (EC 3.6.3.25),

b) polynucleotit được biểu hiện quá mức mã hoá CysH 3'-phosphoađenosin 5'-phosphosulphat ređuctaza (EC 1.8.4.8),

c) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của CysJI sulphit ređuctaza (EC 1.8.1.2),

d) polynucleotit được biểu hiện quá mức mã hoá CysK xystein synthaza A (EC 2.5.1.47),

e) polynucleotit được biểu hiện quá mức mã hoá CysM xystein synthaza B (EC 2.5.1.47),

f) polynucleotit được biểu hiện quá mức mã hoá CysE serin axetyltransferaza (EC 2.3.1.30),

g) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống phân cắt GcvTHP-Lpd glyxin (EC 2.1.2.10, EC 1.4.4.2, EC 1.8.1.4),

h) polynucleotit được biểu hiện quá mức mã hoá LipA lipoyl synthaza (EC 2.8.1.8),

i) polynucleotit được biểu hiện quá mức mã hoá LipB lipoyl-protein ligaza (EC 2.3.1.181),

j) polynucleotit được biểu hiện quá mức mã hoá SerA phosphoglyxerat đehyđrogenaza (EC 1.1.1.95),

k) polynucleotit được biểu hiện quá mức mã hoá SerB 3-phosphoserin phosphataza (EC 3.1.3.3),

l) polynucleotit được biểu hiện quá mức mã hoá SerC 3-phosphoserin/phosphohyđroxythreonin aminotransferaza (EC 2.6.1.52),

m) polynucleotit được biểu hiện quá mức mã hoá GlyA serin hyđroxymetyltransferaza (EC 2.1.2.1),

n) polynucleotit được biểu hiện quá mức mã hoá ThrA aspartokinaza I và homoserin đehyđrogenaza I (EC 2.7.2.4, EC 1.1.1.3),

o) polynucleotit được biểu hiện quá mức mã hoá LysC aspartat kinaza (EC 2.7.2.4),

p) polynucleotit được biểu hiện quá mức mã hoá Hom homoserin đehyđrogenaza (EC 1.1.1.3),

q) polynucleotit được biểu hiện quá mức mã hoá MetX homoserin O-axetyltransferaza (EC 2.3.1.31),

r) polynucleotit được biểu hiện quá mức mã hoá MetA homoserin O-succinyltransferaza (EC 2.3.1.46),

s) polynucleotit được biểu hiện quá mức mã hoá MetB xystathionin gama synthaza (EC 2.5.1.48),

t) polynucleotit được biểu hiện quá mức mã hoá AecD β-C-S-lyaza (EC 4.4.1.8, còn được gọi là beta-lyaza),

u) polynucleotit được biểu hiện quá mức mã hoá MetC xystathionin beta-lyaza (EC 4.4.1.8),

v) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza độc lập với MetE B12 (EC 2.1.1.14),

w) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza phụ thuộc MetH B12 (EC 2.1.1.13),

x) polynucleotit được biểu hiện quá mức mã hoá MetF metylen tetrahyđrofolat ređuctaza (EC 1.5.1.20),

y) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất BrnFE L-metionin ra khỏi *Corynebacterium glutamicum*,

z) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất YgaZH valin *Escherichia coli* (b2682, b2683),

aa) polynucleotit được biểu hiện quá mức mã hoá chất vận chuyển *Escherichia coli* YjeH giả định (b4141),

bb) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của PntAB pyriđin nucleotit transhyđrogenaza (EC 1.6.1.2),

cc) polynucleotit được biểu hiện quá mức mã hoá MetZ O-succinylhomoserin sulphhyđrylaza (EC 2.5.1.48),

dd) polynucleotit được biểu hiện quá mức mã hoá Pyc phosphoenolpyruvat carboxylaza (EC 4.1.1.31),

ee) polynucleotit được biểu hiện quá mức mã hoá RDL2 thiosulphat sulphurtransferaza (EC 2.8.1.1),

ff) polynucleotit được biểu hiện quá mức mã hoá thiosulphat—thiol sulphurtransferaza (EC 2.8.1.3),

gg) polynucleotit được biểu hiện quá mức mã hoá thiosulphat—đithiol sulphurtransferaza (EC 2.8.1.5).

Các đặc điểm được ưu tiên là một hoặc nhiều đặc điểm được chọn từ nhóm bao gồm:

a) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống vận chuyển CysPUWA thiosulphat/sulphat (EC 3.6.3.25),

b) polynucleotit được biểu hiện quá mức mã hoá CysH 3'-phosphoađenosin 5'-phosphosulphat ređuctaza (EC 1.8.4.8),

c) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của CysJI sulphit ređuctaza (EC 1.8.1.2),

d) polynucleotit được biểu hiện quá mức mã hoá CysK xystein synthaza A (EC 2.5.1.47),

e) polynucleotit được biểu hiện quá mức mã hoá CysM xystein synthaza B (EC 2.5.1.47),

f) polynucleotit được biểu hiện quá mức mã hoá CysE serin axetyltransferaza (EC 2.3.1.30),

g) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống phân cắt GcvTHP-Lpd glyxin (EC 2.1.2.10, EC 1.4.4.2, EC 1.8.1.4),

h) polynucleotit được biểu hiện quá mức mã hoá LipA lipoyl synthaza (EC 2.8.1.8),

i) polynucleotit được biểu hiện quá mức mã hoá LipB lipoyl-protein ligaza (EC 2.3.1.181),

j) polynucleotit được biểu hiện quá mức mã hoá SerA phosphoglyxerat đehyđrogenaza (EC 1.1.1.95),

k) polynucleotit được biểu hiện quá mức mã hoá SerB 3-phosphoserin phosphataza (EC 3.1.3.3),

l) polynucleotit được biểu hiện quá mức mã hoá SerC 3-phosphoserin/phosphohyđroxythreonin aminotransferaza (EC 2.6.1.52),

m) polynucleotit được biểu hiện quá mức mã hoá GlyA serin hyđroxymetyltransferaza (EC 2.1.2.1),

n) polynucleotit được biểu hiện quá mức mã hoá ThrA aspartokinaza I và homoserin đehyđrogenaza I (EC 2.7.2.4, EC 1.1.1.3),

o) polynucleotit được biểu hiện quá mức mã hoá LysC aspartat kinaza (EC 2.7.2.4),

p) polynucleotit được biểu hiện quá mức mã hoá Hom homoserin đehyđrogenaza (EC 1.1.1.3),

q) polynucleotit được biểu hiện quá mức mã hoá MetX homoserin axetyltransferaza (EC 2.3.1.31),

r) polynucleotit được biểu hiện quá mức mã hoá MetA homoserin O-transsucxinylaza (EC 2.3.1.46),

s) polynucleotit được biểu hiện quá mức mã hoá MetB xystathionin gama synthaza (EC 2.5.1.48),

t) polynucleotit được biểu hiện quá mức mã hoá AecD β-C-S-lyaza (EC 4.4.1.8, còn được gọi là beta-lyaza),

u) polynucleotit được biểu hiện quá mức mã hoá MetC xystathionin beta-lyaza (EC 4.4.1.8),

v) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza độc lập với MetE B12 (EC 2.1.1.14),

w) polynucleotit được biểu hiện quá mức mã hoá B12 homoxystein S-metyltransferaza phụ thuộc MetH (EC 2.1.1.13),

x) polynucleotit được biểu hiện quá mức mã hoá MetF metylen tetrahyđrofolat ređuctaza (EC 1.5.1.20),

y) polynucleotit được biểu hiện quá mức mã hoá RDL2 thiosulphat sulphurtransferaza (EC 2.8.1.1).

Các đặc điểm được ưu tiên rất đặc biệt ở bản mô tả này được chọn từ nhóm bao gồm:

a) polynucleotit được biểu hiện quá mức mã hoá ThrA aspartokinaza I và homoserin đehyđrogenaza I (EC 2.7.2.4, EC 1.1.1.3),

b) polynucleotit được biểu hiện quá mức mã hoá CysE serin axetyltransferaza (EC 2.3.1.30),

c) polynucleotit được biểu hiện quá mức mã hoá LysC aspartat kinaza (EC 2.7.2.4),

d) polynucleotit được biểu hiện quá mức mã hoá Hom homoserin đehyđrogenaza (EC 1.1.1.3),

e) polynucleotit được biểu hiện quá mức mã hoá MetX homoserin axetyltransferaza (EC 2.3.1.31),

f) polynucleotit được biểu hiện quá mức mã hoá MetA homoserin O-transscucxinylaza (EC 2.3.1.46),

g) polynucleotit được biểu hiện quá mức mã hoá MetB xystathionin gama synthaza (EC 2.5.1.48),

h) polynucleotit được biểu hiện quá mức mã hoá AecD β-C-S-lyaza (EC 4.4.1.8, còn được gọi là beta-lyaza),

i) polynucleotit được biểu hiện quá mức mã hoá MetC xystathionin beta-lyaza (EC 4.4.1.8),

j) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza độc lập với MetE B12 (EC 2.1.1.14),

k) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza phụ thuộc MetH B12 (EC 2.1.1.13),

l) polynucleotit được biểu hiện quá mức mã hoá MetF metylen tetrahyđrofolat ređuctaza (EC 1.5.1.20),

m) polynucleotit được biểu hiện quá mức mã hoá RDL2 thiosulphat sulphurtransferaza (EC 2.8.1.1).

Theo sáng chế, thuật ngữ "biểu hiện quá mức", "tăng cường" hoặc "hoạt tính được tăng lên" đề cập đến sự gia tăng hoạt tính enzym nội bào của một hoặc nhiều enzym ở vi sinh vật, được mã hoá bởi ADN tương ứng.

Về mặt nguyên tắc hoạt tính enzym có thể được tăng lên, ví dụ, bằng cách tăng số lượng bản sao của trình tự gen hoặc các trình tự gen mã hoá enzym này, bằng cách sử dụng trình tự khởi đầu mạnh hoặc bằng cách sử dụng gen hoặc alen mã hoá enzym tương ứng có hoạt tính được tăng lên, và bằng cách kết hợp các biện pháp này, nếu thích hợp. Ví dụ, các tế bào được cải biến di truyền theo sáng chế được tạo ra bằng cách biến nạp, tải nạp, liên hợp, hoặc kết hợp của các phương pháp này, với vật truyền chứa gen mong muốn, alen của gen này hoặc một phần của chúng, và vật truyền có khả năng làm cho gen nêu trên được biểu hiện. Đặc biệt là đạt được sự biểu hiện khác loài bằng cách hợp nhất gen hoặc alen này vào nhiễm sắc thể cảu tế bào hoặc vật truyền sao chép ngoài nhiễm sắc thể.

Tổng quan về khả năng làm tăng hoạt tính enzym trong tế bào với ví dụ pyruvat carboxylaza được nêu trong DE-A-100 31 999 tài liệu này được kết hợp vào bản mô tả này theo cách viện dẫn và nội dung bộc lộ của tài liệu này về khả năng làm tăng hoạt tính enzym trong tế bào tạo thành một phần của mô tả sáng chế.

Sự gia tăng hoạt tính enzym có thể đạt được, ví dụ, bằng cách tăng số lượng bản sao của các polynucleotit tương ứng ở nhiễm sắc thể hoặc ngoài nhiễm sắc thể lên ít nhất một bản sao.

Một phương pháp được sử dụng rộng rãi để làm tăng số lượng bản sao bao gồm việc kết hợp polynucleotit tương ứng vào vật truyền, tốt hơn nếu là plasmit, mà được sao chép bởi vi khuẩn.

Ví dụ về các vật truyền plasmit thích hợp cho *Enterobacteriaceae* là các vật truyền tách dòng thu được từ pACYC184 (Bartolomộ et al.; Gene 102: 75-78 (1991)), pTrc99A (Amann et al.; Gene 69: 301-315 (1988)) hoặc các dẫn xuất pSC101 (Vocke and Bastia; Proceedings of the National Academy of Sciences USA 80(21): 6557-6561 (1983)) có thể được sử dụng. Các plasmit thu được từ pCL1920 (Lerner, C.G. and Inouye, M., Nucl. Acids Res. (1990) 18:4631 [PMID: 2201955]) cũng đặc biệt thích hợp. Các vật truyền plasmit thu được từ nhiễm sắc thể nhân tạo của vi khuẩn (bacterial artificial chromosome: BAC), như ví dụ pCC1BAC (EPICENTRE Biotechnologies, Madison, USA), cũng thích hợp để làm tăng số lượng bản sao của các polynucleotit tương ứng ở *E. coli*.

Ngoài ra, các gen nhảy, yếu tố cài xen (yếu tố IS: insertion element) hoặc thể thực khuẩn có thể được sử dụng làm vật truyền. Các hệ thống như vậy được mô tả, ví dụ, trong các bản mô tả patent US 4,822,738, US 5,804,414 và US 5,804,414. Yếu tố IS ISaB1 như được mô tả trong WO 92/02627 hoặc gen nhảy Tn 45 của plasmit pXZ10142 (được nêu trong "Handbook of *Corynebacterium glutamicum*" (nhà xuất bản: L. Eggeling and M. Bott)) có thể được sử dụng theo cách tương tự.

Một phương pháp được dùng rộng rãi khác để đạt được sự biểu hiện quá mức là quá trình khuếch đại gen ở nhiễm sắc thể. Trong phương pháp này, ít nhất một bản sao bổ sung của polynucleotit quan tâm được xen vào nhiễm sắc thể của vi khuẩn. Các phương pháp khuếch đại như vậy được mô tả, ví dụ, trong WO 03/014330 hoặc WO 03/040373.

Một phương pháp khác để đạt được sự biểu hiện quá mức bao gồm việc liên kết gen hoặc alen tương ứng theo cách chức năng (được liên kết có điều khiển) với trình tự khởi đầu hoặc caxet biểu hiện. Các ví dụ đã biết về các trình tự khởi đầu thích hợp cho *E. coli* là các trình tự khởi đầu T3, T7, SP6, M13, lac, tac và trc được mô tả bởi Amann et al. (Gene 69(2), 301-315 (1988)) và Amann và Brosius (Gene 40(2-3), 183-190 (1985)). Trình tự khởi đầu như vậy có thể được cài xen, ví dụ, nằm trước của gen quan tâm, thông thường ở khoảng cách khoảng 1 - 500 nucleobazơ từ cođon bắt đầu. US 5,939,307 báo cáo rằng việc kết hợp các caxet biểu hiện hoặc trình tự khởi đầu như, ví dụ, trình tự khởi đầu tac, trình tự khởi đầu trp, trình tự khởi đầu lpp hoặc thể thực khuẩn ở PL và trình tự khởi đầu PR, ví dụ nằm trước operon threonin của nhiễm sắc thể, có thể đạt được sự tăng biểu hiện. Trình tự khởi đầu của thể thực khuẩn T7, trình tự khởi đầu gear-box, trình tự khởi đầu nar hoặc trình tự khởi đầu của gen rrsG, rnpB, csrA, csrB, ompA, fusA, pepQ, rplX hoặc rpsG có thể được sử dụng theo cùng cách như vậy. Các caxet biểu hiện hoặc trình tự khởi đầu như vậy cũng có thể được sử dụng cho biểu hiện quá mức gen được liên kết với plasmit, như mô tả trong EP 0 593 792. Bằng cách sử dụng alen lacIQ lại có thể khống chế biểu hiện của gen được liên kết với plasmit (Glascock and Weickert, Gene 223, 221-231 (1998)). Ngoài ra, có thể khống chế làm hoạt tính của trình tự khởi đầu được tăng lên do cải biến trình tự của chúng bằng một hoặc nhiều thay thế, cài xen và/hoặc làm khuyết nucleotit.

Nếu như sự gia tăng hoạt tính enzym được thực hiện bằng cách gây đột biến gen nội sinh, thì các đột biến như vậy có thể được tạo ra bằng phương pháp thông thường theo cách không hướng đích, ví dụ bằng cách chiếu xạ bằng UV hoặc bằng hóa chất gây đột biến, hoặc theo cách hướng đích bằng các phương pháp kỹ thuật di truyền như làm khuyết), cài xen và/hoặc thay thế nucleotit. Các đột biến nêu trên tạo ra các tế bào được cải biến di truyền. Các enzym đột biến đặc biệt được ưu tiên đặc biệt là các enzym không bị ức chế bởi cơ chế ngược, hoặc ít nhất là có cơ chế ức chế ngược giảm đi so với enzym kiểu dại.

Nếu sự gia tăng hoạt tính enzym được thực hiện bằng cách tăng biểu hiện của enzym, thì ví dụ, số lượng bản sao của gen tương ứng được tăng lên, hoặc trình tự khởi

đầu và vùng điều hoà hoặc vị trí gắn kết ribosom nằm trước gen cấu trúc được gây đột biến. Caxet biểu hiện được kết hợp nằm trước gen cấu trúc cũng có tác dụngnhư vậy . Trình tự khởi đầu cảm ứng có thể giúp biểu hiện tăng thêm ở thời điểm bất kỳ. Ngoài ra, tuy nhiên, gen được gọi là "gen tăng cường" cũng có thể quy là gen như là trình tự điều hoà mà tương tự cũng làm tăng biểu hiện của gen bằng tương tác được cải thiện giữa ARN polymeaza và ADN. Biện pháp kéo dài sự tồn tại của ARN thông tin tương tự cùng cải thiện sự biểu hiện. Ngoài ra, hoạt tính enzym cũng tăng lên bằng cách ngăn sự thoái biến của protein enzym. Trong bản mô tả này, gen hoặc các cấu trúc gen hoặc là có mặt ở plasmit có số lượng bản sao khác nhau hoặc được hợp nhất vào nhiễm sắc thể và được khuếch đại. Theo cách khác, ngoài ra có thể đạt được biểu hiện quá mức của gen quan tâm bằng cách thay đổi thành phần của môi trường và quy trình nuôi cấy. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này có thể tìm các chỉ dẫn trong số các tài liệu Martin et al. (Bio/Technology 5, 137-146 (1987)), trong Guerrero et al. (Gene 138, 35-41 (1994)), Tsuchiya and Morinaga (Bio/Technology 6, 428-430 (1988)), trong Eikmanns et al. (Gene 102, 93-98 (1991)), trong EP-A-0 472 869, trong US 4,601,893, trong Schwarzer and Pÿhler (Bio/Technology 9, 84-87 (1991)), trong Reinscheid et al. (Applied and Environmental Microbiology 60, 126-132 (1994)), trong LaBarre et al. (Journal of Bacteriology 175, 1001-1007 (1993)), trong WO-A-96/15246, trong Malumbres et al. (Gene 134, 15-24 (1993)), trong JP-A-10-229891, trong Jensen and Hammer (Biotechnology and Bioengineering 58, 191-195 (1998)) và trong các giáo trình đã biết về di truyền và sinh học phân tử. Các biện pháp đã nêu trên, giống như đột biến, tạo ra các tế bào được cải biến di truyền.

Ngoài ra, các vật truyền plasmit này cũng thích hợp mà nhờ chúng có thể áp dụng quy trình khuếch đại gen bằng cách hợp nhất vào nhiễm sắc thể. Sau khi tái tổ hợp tương đồng bằng cách lai chéo, các chủng thu được chứa ít nhất hai bản sao của gen quan tâm. Quy trình cho *E. coli* được mô tả, ví dụ, trong Link, A.J., Phillips, D. và Church, G.M. (1997), J. Bacteriology 179: 6228-6237.

Cũng có thể sử dụng các quy trình qua trung gian recombinaza, ví dụ các quy trình được mô tả bởi Datsenko KA, Wanner BL., 2000, Proc Natl Acad Sci USA., 97(12):6640-5, để xen đoạn hoặc làm khuyết ADN trong nhiễm sắc thể.

Cụm từ "hoạt tính enzym được tăng lên so với chủng kiểu dại hoặc chủng bắt đầu của nó" được sử dụng trên đây và trong phần giải thích sau đây tốt hơn nếu luôn dùng để chỉ hoạt tính của một enzym cụ thể đã được tăng lên với hệ số ít nhất 2, đặc biệt tốt hơn nếu ít nhất 10, ngoài ra tốt hơn nếu ít nhất 100, ngoài ra còn tốt hơn nữa nếu ít nhất 1000 và tốt nhất nếu ít nhất 10000. Tế bào theo sáng chế có "hoạt tính enzym được tăng lên so với chủng kiểu dại hoặc chủng bắt đầu" ngoài ra còn bao gồm đặc biệt là tế bào mà chủng kiểu dại hoặc chủng bắt đầu không có hoạt tính hoặc ít nhất có hoạt tính không thể phát hiện được của enzym nêu trên và chỉ có thể phát hiện được hoạt tính của enzym này sau khi làm tăng hoạt tính enzym, ví dụ bằng biểu hiện quá mức. Về điểm này, thuật ngữ "biểu hiện quá mức" hoặc cụm từ "tăng biểu hiện" được sử dụng trong phần giải thích sau đâu cũng bao gồm trường hợp trong đó tế bào bắt đầu, ví dụ tế bào kiểu dại, không có biểu hiện hoặc ít nhất có biểu hiện không thể phát hiện và chỉ nhờ các quy trình tái tổ hợp mới có thể phát hiện được biểu hiện của enzym này.

Các phương pháp xác định hoạt tính enzym của các enzym khác nhau có thể xem trong tài liệu.

Biểu hiện của các enzym hoặc gen được đề cập trên đây có thể được phát hiện bằng phân đoạn gel protein một và hai chiều và sau đó nhận diện quang học nồng độ của protein trong gel này, bằng cách sử dụng phần mềm đánh giá thích hợp . Nếu như sự tăng hoạt tính enzym chỉ dựa trên sự tăng biểu hiện của gen tương ứng, thì sự tăng hoạt tính enzym có thể được định lượng theo cách đơn giản bằng cách so sánh phân đoạn protein một và hai chiều giữa tế bào kiểu dại và được cải biến di truyền. Phương pháp thông thường dùng để điều chế gel protein trong trường hợp vi khuẩn và nhận diện protein là quy trình được mô tả của Hermann et al. (Electrophoresis, 22: 1712.23 (2001)). Tương tự, có thể phân tích nồng độ protein bằng phương pháp lai thấm Tây với kháng thể đặc hiệu cho protein sẽ được phát hiện (Sambrook et al., Molecular

Cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY USA, 1989) và đánh giá bằng quang học sau đó bằng cách sử dụng phần mềm thích hợp để đo nồng độ (Lohaus and Meyer (1989) Biospektrum, 5: 32-39; Lottspeich (1999), Angewandte Chemie 111: 2630-2647). Hoạt tính của protein gắn kết với ADN có thể được đo bằng thử nghiệm dịch chuyển dải ADN (còn được gọi là trì hoãn gel: gel retardation) (Wilson et al. (2001) Journal of Bacteriology, 183: 2151-2155). Tác động của protein gắn kết với ADN đến biểu hiện của các gen khác có thể được phát hiện bằng phương pháp thử nghiệm dùng gen thông báo khác nhau được mô tả rõ (Sambrook et al., Molecular Cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY USA, 1989). Hoạt tính enzym nội bào có thể được xác định bằng các phương pháp khác nhau đã được mô tả (Donahue et al. (2000) Journal of Bacteriology 182 (19): 5624-5627; Ray et al. (2000) Journal of Bacteriology 182 (8): 2277-2284; Freedberg et al. (1973) Journal of Bacteriology 115 (3): 816-823). Nếu như trong phần giải thích sau không chỉ ra các phương pháp cụ thể để xác định hoạt tính của enzym nhất định, thì tốt hơn nếu sự tăng hoạt tính enzym và cũng như sự giảm hoạt tính enzym được xác định bằng phương pháp như được mô tả trong Hermann et al., Electophoresis, 22: 1712-23 (2001), Lohaus et al., Biospektrum 5 32-39 (1998), Lottspeich, Angewandte Chemie 111: 2630-2647 (1999) và Wilson et al., Journal of Bacteriology 183: 2151-2155 (2001).

Ngoài ra, để sản xuất L-axit amin, đặc biệt là L-metionin có thể tốt hơn nếu hạn chế được các phản ứng phụ không mong muốn (Nakayama: "Breeding of Amino Acid Producing Microorganisms", trong: Overproduction of Microbial Products, Krumphanzl, Sikyta, Vanek (eds.), Academic Press, London, UK, 1982).

Do đó, để cải thiện sự sản xuất L-metionin trong *E.coli*, có thể có lợi nếu được làm suy giảm, tuỳ ý tắt đi, một hoặc nhiều gen, hoặc làm giảm biểu hiện, được chọn từ nhóm bao gồm:

a) gen *metJ* mã hoá chất điều hoà phiên mã sinh tổng hợp L-metionin (MetJ) (b3938, ECK3930),

b) gen pgi mã hoá glucoza-6-phosphat isomeraza (Pgi, EC No. 5.3.1.9) (b4025, ECK4017),

c) gen thrB mã hoá homoserin kinaza (ThrB, EC 2.7.1.39) (b0003, ECK0003),

d) gen metK mã hoá S-ađenosylmetionin synthaza (MetK, EC No. 2.5.1.6) (b2942, ECK2937),

e) gen dapA mã hoá đihyđrođipicolinat synthaza (DapA, EC No. 4.2.1.52) (b2478, ECK2474),

f) gen pck mã hoá phosphoenolpyruvat carboxykinaza (Pck, EC No. 4.1.1.49) (b3403, ECK3390),

g) gen purU mã hoá formyltetrahyđrofolat hydrolaza (PurU, EC No. 3.5.1.10) (b1232, ECK1227),

h) gen pykA mã hoá pyruvat kinaza II (PykA, EC No. 2.7.1.40) (b1854, ECK1855)

i) gen pykF mã hoá pyruvat kinaza I (PykF, EC 2.7.1.40) (b1676, ECK1672),

j) gen metQ mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0197, ECK0197),

k) gen metI mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0198, ECK0198),

l) gen mã hoá metN cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0199, ECK0199),

m) gen dcd mã hoá đeoxyxytiđin-5'-triphosphat đeaminaza (Dcd, EC No. 3.5.4.13) (b2065, ECK2059),

n) gen yncA mã hoá N-axyltransferaza giả định (YncA, Khám phá trao đổi chất: Metabolic Explorer WO2010/020681) (b1448, ECK1442),

o) gen fnrS mã hoá gen điều hoà FnrS sRNA (b4699, ECK4511),

p) gen rpoS mã hoá yếu tố RpoS sigma (b2741, ECK2736).

Các đặc điểm được ưu tiên đặc biệt được chọn nhóm bao gồm:

a) gen *metJ* mã hoá chất điều hoà phiên mã sinh tổng hợp L-metionin (MetJ) (b3938, ECK3930),

b) gen metK mã hoá S-ađenosylmetionin synthaza (MetK, EC No. 2.5.1.6) (b2942, ECK2937),

c) gen pck mã hoá phosphoenolpyruvat carboxykinaza (Pck, EC No. 4.1.1.49) (b3403, ECK3390),

d) gen purU mã hoá formyltetrahyđrofolat hydrolaza (PurU, EC No. 3.5.1.10) (b1232, ECK1227),

e) gen pykA mã hoá pyruvat kinaza II (PykA, EC No. 2.7.1.40) (b1854, ECK1855)

f) gen pykF mã hoá pyruvat kinaza I (PykF, EC 2.7.1.40) (b1676, ECK1672),

g) gen metQ mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0197, ECK0197),

h) gen metI mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0198, ECK0198),

i) gen metN mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0199, ECK0199),

j) gen yncA mã hoá N-axyltransferaza giả định (YncA, Khám phá trao đổi chất: Metabolic Explorer WO2010/020681) (b1448, ECK1442),

k) gen rpoS mã hoá yếu tố RpoS sigma (b2741, ECK2736).

Các đặc điểm được ưu tiên rất đặc biệt ở bản mô tả này được chọn từ nhóm bao gồm:

a) gen *metJ* mã hoá chất điều hoà phiên mã sinh tổng hợp L-metionin (MetJ) (b3938, ECK3930),

b) gen metK mã hoá S-ađenosylmetionin synthaza (MetK, EC No. 2.5.1.6) (b2942, ECK2937),

c) gen metQ mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0197, ECK0197),

d) gen metI mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0198, ECK0198),

e) gen metN mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0199, ECK0199),

f) gen yncA mã hoá N-axyltransferaza giả định (YncA, Khám phá trao đổi chất: Metabolic Explorer WO2010/020681) (b1448, ECK1442),

g) gen rpoS mã hoá yếu tố RpoS sigma (b2741, ECK2736).

Nếu thích hợp, tiến hành các biện pháp làm suy giảm bổ sung cho hoặc được kết hợp thích hợp với các biện pháp nêu ra để làm tăng gen nhằm làm tăng sản xuất metionin.

Theo sáng chế, các vi sinh vật sản xuất L-axit amin trước khi thực hiện đo lường theo sáng chế không bao gồm chủng kiểu dại và thông thường sử dụng các chủng dùng trong phòng thí nghiệm như, ví dụ các chủng DH5ỏ, DH5ỏmcr, W3110, MG1655, MC4100, Y1089, H560, BL21 và MM152.

Tuy nhiên, các vi sinh vật sản xuất L-axit amin có thể thu được từ các chủng kiểu dại nêu trên và chủng mà phòng thí nghiệm thường được sử dụng.

Do đó, tốt hơn nếu chủng vi sinh vật ban đầu thu được từ nhóm bao gồm *Escherichia coli* MG1655, *Escherichia coli* W3110, *Escherichia coli* DH5α, *Escherichia coli* DH10B, *Escherichia coli* BW2952, *Escherichia coli* REL606.

Một ví dụ về chủng sản xuất hoặc tiết L-metionin mà được ưu tiên theo sáng chế là chủng sản xuất *E. coli* MG1655 ΔmetJ metA*11 Ptrc-metH Ptrc-metF PtrcF-cysPUWAM PtrcF-cysJIH ΔpykF ΔpykA Ptrc09-gcvTHP ΔpurU Ptrc36-ARNmst17-metF, mang plasmit sản xuất pME101-thrA*1-cysE-Pgap-metA*11 và pCC1BAC-serB-glyA-serA-serC (WO2009/043803).

Một ví dụ nữa về chủng sản xuất hoặc tiết L-metionin mà được ưu tiên theo sáng chế là chủng sản xuất *E. coli* MG1655 ΔmetJ Ptrc-metH Ptrc-metF PtrcF-

cysPUWAM PtrcF-cysJIH Ptrc09-gcvTHP, mang plasmit sản xuất pME101-thrA*1-cysE-Pgap-metA*11.

Chủng sản xuất *E. coli* MG1655ΔmetJ Ptrc-metH Ptrc-metF PtrcF-cysPUWAM PtrcF-cysJIH Ptrc09-gcvTHP có thể được tách dòng bằng hàng loạt tải nạp và xử lý P1, như được mô tả trong đơn yêu cầu cấp patent WO2009/043803. Chủng này dựa trên cơ sở chủng kiểu dại *E. coli* K12 MG1655. Các cải biến sau được đưa vào hệ gen của chủng này:

- Gen *metJ* cho chất ức chế sinh tổng hợp L-metionin được làm khuyết.
- trình tự khởi đầu trc mạnh được cài vào phía trước gen *metH* (mã hoá metionin synthaza phụ thuộc cobalamin).
- trình tự khởi đầu trc mạnh được cài xen ở phía trước gen *metF* gen (mã hoá 5,10-metylentetrahyđrofolat ređuctaza).
- trình tự khởi đầu trcF mạnh được cài xen ở phía trước operon *cysPUWAM*. *cysPUWA* mã hoá chất vận chuyển hấp thu sulphat/thiosulphat. *cysM* mã hoá xystein synthaza B.
- trình tự khởi đầu trcF mạnh được cài xen ở phía trước operon *cysJIH*. *cysJI* mã hoá sulphit ređuctaza và *cysH* mã hoá 3'-phosphoadenylyl-sulphat ređuctaza.
- trình tự khởi đầu trc09 mạnh được cài xen ở phía trước operon *gcvTHP*. *gcvT, gcvH* và *gcvP* mã hóa ba hợp phần của hệ thống phân cắt glyxin.

Việc nhân dòng plasmit sản xuất *E. coli* pME101-thrA*1-cysE-Pgap-metA*11 được mô tả trong các đơn yêu cầu cấp patent WO2007/077041 và WO2009/043803. Plasmit này là plasmit có số bản sao thấp (plasmit bản sao thấp) trên cơ sở vật truyền pCL1920 (Lerner, C.G. and Inouye, M., Nucl. Acids Res. (1990) 18:4631 [PMID: 2201955]). Plasmit trống, pME101, có gen $lacI^q$ gen mã hoá alen được biểu hiện mạnh của chất ức chế lac. Gen *thrA*1* được tách dòng nằm sau trình tự khởi đầu trc mạnh mà có thể được ức chế bởi chất ức chế Lac. Gen này mã hoá biến thể có tính kháng cơ chế ức chế ngược của *E. coli* ThrA aspartat kinaza/homoserin đehyđrogenaza. Nằm phía sau, theo cùng một hướng, là gen *cysE* cùng với trình tự khởi đầu tự nhiên của nó.

Gen này mã hoá serin axetyltransferaza ở *E. coli*. *cysE* được theo sau bởi trình tự khởi đầu *E. coli gapA* mạnh mà không chế biểu hiện của gen *metA*11*. *metA*11* mã hoá biến thể tính kháng cơ chế ức chế ngược của homoserin O-succinyltransferaza *E. coli*.

Ví dụ về các vi sinh vật tiết hoặc sản xuất L-metionin mà được ưu tiên theo sáng chế có thể được đề cập là các chủng sau:

- *E. coli* TF4076BJF metA#10 + metYX(Lm) (WO2008/127240; trang 46);

- *E. coli* W3110ΔJ/pKP451 (EP 1 445 310 B1, trang 7, ví dụ 4);

- *E. coli* WΔthrBCΔmetJmetK32 pMWPthrmetA4Δ5Δ9 (Yoshihiro Usuda and Osamu Kurahashi, 2005, Applied and Environmental Microbiology, Vol. 71, NO: 6, các trang 3228-3234);

- W3110/pHC34 (WO01/27307 trang 13, ví dụ 3);

- *E. coli* ECM2 (EP2205754A2, EP2326724A1 và EP12156052,8).

Các ví dụ tiếp theo về các vi sinh vật thích hợp khác nhau được mô tả trong Gomes et al. (Enzyme and Microbial Technology 37 (2005), 3-18).

Trình tự nucleotit của gen hoặc khung đọc mở (ORF) của *Escherichia coli* là đã biết trong tình trạng kỹ thuật và có thể xem trong trình tự hệ gen của *Escherichia coli* được công bố bởi Blattner et al. (Science 277: 1453–1462 (1997)). Enzym nội sinh đối với cơ thể chủ (metionin aminopeptiđaza) đã được biết đến là có thể tách loại axit amin metionin ở đầu tận N.

Các giải thích chi tiết hơn về các thuật ngữ của sinh học di truyền và phân tử có thể xem trong các giáo trình đã biết của ngành sinh học di truyền và phân tử như, ví dụ, giáo trình của Birge (Bacterial and Bacteriophage Genetics, 4th ed., Springer Verlag, New York (USA), 2000) hoặc giáo trình của Berg, Tymoczko and Stryer (Biochemistry, 5th ed., Freeman and Company, New York (USA), 2002) hoặc sổ tay của Sambrook et al. (Molecular Cloning, A Laboratory Manual, (3-volume set), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (USA), 2001).

Thuật ngữ "gen" ở bản mô tả này dùng để chỉ phần axit đeoxyribonucleic (ADN), bao gồm các thông tin sản xuất (hiên mã) axit ribonucleic (ARN) thứ nhất mà

bao gồm thông tin sản xuất (dịch mã) protein (polypeptit), trong trường hợp polypeptit có hoạt tính (p)ppGpp synthetaza II. Nếu như gen hoặc polynucleotit bao gồm thông tin sản xuất protein thì còn được gọi là mã hoá protein hoặc polypeptit bởi gen hoặc bởi ARN. Các gen và polynucleotit nội sinh dùng để chỉ khung đọc mở (ORF), gen hoặc alen và các polynucleotit của chúng tương ứng có trong quần thể của một loài. Các thuật ngữ "gen" và "open reading frame: ORF" (khung đọc mở) được sử dụng theo sáng chế là đồng nghĩa.

Thuật ngữ "polynucleotit" thông thường dùng để chỉ polyribonucleotit và polyđeoxyribonucleotit, mà có thể là ARN hoặc ADN không cải biến hoặc ARN hoặc AND cải biến.

Thuật ngữ "polypeptit" dùng để chỉ peptit hoặc protein chứa hai axit amin hoặc nhiều hơn được nối với nhau thông qua liên kết peptit. Các thuật ngữ polypeptit và protein được sử dụng đồng nghĩa. Protein là nằm trong khối cấu trúc cơ bản của tất cả các tế bào. Protein không chỉ tạo nên cấu trúc tế bào mà còn là "cỗ máy" phân tử vận chuyển các chất, xúc tác các phản ứng hoá học và nhận biết các chất tín hiệu.

"Axit amin chuẩn (proteinogenic)" dùng để chỉ các axit amin được tìm thấy trong các protein tự nhiên, tức là trong protein của các vi sinh vật, thực vật, động vật và người. Đặc biệt hơn, các axit amin bao gồm L-axit amin được chọn từ nhóm bao gồm axit L-aspartic, L-asparagin, L-threonin, L-serin, axit L-glutamic, L-glutamin, glyxin, L-alanin, L-xystein, L-valin, L-metionin, L-isoleuxin, L-leuxin, L-tyrosin, L-phenylalanin, L-histiđin, L-lysin, L-tryptophan, L-prolin và L-arginin, và cả selenoxystein. Các axit amin chuẩn luôn là các ỏ-axit amin. Nguyên tử cacbon ỏ là không đối xứng cho tất cả các axit amin chuẩn (chúng là các phân tử không đối xứng), chỉ trừ axit amin glyxin: có hai chất đồng phân đối ảnh ở mỗi trong số các axit amin này. Chỉ một trong số hai chất đồng phân đối ảnh là chuẩn, đó là L-axit amin: là cơ cấu cần thiết tổng hợp protein – ribosom, tRNA, aminoaxyl-tRNA synthetaza (mà tích tRNA bằng axit amin) và các axit amin khác có bản chất không đối xứng và có thể nhận biết chỉ biến thể L.

Thuật ngữ "biểu hiện gen" (gọi tắt là "biểu hiện") thông thường dùng để chỉ biểu hiện của thông tin di truyền bằng kiểu hình. Theo nghĩa hẹp, biểu hiện gen dùng để chỉ sự phiên mã một gen thành ARN, sự dịch mã sau đó ARN thành polypeptit mà có thể có hoạt tính enzym.

"Chủng bắt đầu" (chủng cha mẹ) dùng để chỉ chủng vi sinh vật mà được áp dụng biện pháp xử lý để làm tăng sản xuất một hoặc nhiều axit amin, peptit hoặc protein, hoặc biện pháp để làm tăng hoạt tính của một hoặc nhiều enzym (ví dụ biện pháp dẫn đến biểu hiện quá mức). Chủng bắt đầu có thể là chủng kiểu dại hoặc chủng đã được cải biến trước đây (ví dụ vi sinh vật sản xuất L-axit amin (chủng sản xuất).

"Kiểu dại" của tế bào thường dùng để chỉ tế bào mà hệ gen của nó ở trạng thái như được phát triển trong tự nhiên trong quá trình tiến hóa. Thuật ngữ này được sử dụng để chỉ cả tế bào toàn và các gen đơn lẻ. Do đó, đặc biệt thuật ngữ "kiểu dại" không bao gồm các tế bào hoặc gen, mà trình tự gen chúng đã được cải biến, ít nhất một phần, bởi con người bằng các quá trình tái tổ hợp.

Sáng chế sẽ được mô tả chi tiết hơn sau đây dựa trên các phương án được lấy làm ví dụ.

Môi trường tối thiểu (M9) và đầy đủ (LB) được sử dụng cho *Escherichia coli* đã được mô tả bởi J.H. Miller (A Short Course in Bacterial Genetics (1992), Cold Spring Harbor Laboratory Press). Trừ khi được mô tả khác đi, quá trình phân lập ADN plasmit từ *Escherichia coli* và tất cả các kỹ thuật liên quan đến phân cắt giới hạn, ghép nối, xử lý bằng Klenow và xử lý bằng phosphataza kiềm được thực hiện theo Sambrook et al. (Molecular Cloning - A Laboratory Manual (1989) Cold Spring Harbor Laboratory Press). Trừ khi được mô tả khác đi, quá trình biến nạp *Escherichia coli* được thực hiện theo Chung et al. (Proceedings of the National Academy of Sciences of the United States of America 86: 2172-2175 (1989)).

Trừ khi được mô tả theo cách khác, nhiệt độ ủ cho các chủng sản xuất và các thể biến nạp là 37°C.

Ví dụ thực hiện sáng chế

Ví dụ 1

Sàng lọc các đột biến dung nạp L-metionin

Chủng *E. coli* sản xuất L-metionin ECM2 dựa trên có sở chủng kiểu dại K12 MG1655. Như mô tả trong EP2205754A2, EP2326724A1 và EP12156052.8, chủng ECM2 này mang alen metA có tính kháng cơ chế ức chế ngược, làm khuyết gen metJ, yncA, pykA và pykF, biến thể của gen spoT, và trình tự khởi đầu tăng cường ở phía trước mỗi trong số các gen metH, metF, gcvT, cysP và cysJ.

1.1 Nhân dòng gen serC vào plasmit pUC18

Gen serC từ *Escherichia coli* MG1655 được khuếch đại nhờ phản ứng chuỗi polymeaza (PCR) và sau đó được tách dòng vào plasmit pUC18 (Fermentas GmbH, St. Leon-Rot, Germany).

Các đoạn mồi PCR serCF(XbaI) và serCR(HindIII) có ở đầu 5' của chúng trong mỗi trường hợp 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết các endonucleaza giới hạn XbaI (TCTAGA) và HindIII (AAGCTT) tương ứng. Các nucleotit 13 đến 38 của serCF(XbaI) gắn kết trong hệ gen MG1655 *E. Coli* từ vị trí 956619 đến 956644. Các nucleotit 13 đến 37 của serCR(HindIII) gắn kết trong hệ gen MG1655 của *E. coli* từ vị trí 958028 đến 958004.

serCF(XbaI) (SEQ ID NO: 25)

5' AGGTGCTCTAGAGTCCGCGCTGTGCAAATCCAGAATGG 3'

serCR(HindIII) (SEQ ID NO: 26)

5' TACACCAAGCTTAACTCTCTACAACAGAAATAAAAAC 3'

Gen serC được khuếch đại bằng cách sử dụng phản ứng chuỗi polymeaza (PCR) bằng các đoạn mồi serCF(XbaI) và serCR(HindIII) và bằng ADN polymeaza Phusion (Finnzymes Oy, Espoo, Finland). ADN hệ gen của *E. coli* MG1655 đóng vai trò làm khuôn mẫu. Mảnh ADN thu được có kích cμ 1434 bp. Mảnh này được phân cắt bằng các endonucleaza giới hạn XbaI và HindIII và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Plasmit ADN không được metyl hóa pUC18 được phân cắt bằng các endonucleaza giới hạn XbaI và HindIII và được tinh

chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Sau đó, plasmit đã được phân cắt được ghép nối với sản phẩm PCR và được biến nạp vào *E. coli* DH5α. Các dòng plasmit chứa gen serC được nhận diện bằng cách phân cắt giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pUC18-serC.

1.2 Nhân dòng gen serA vào plasmit pUC18-serC

Gen serA từ *Escherichia coli* MG1655 được khuếch đại bằng phản ứng chuỗi polymeaza (PCR) và sau đó được tách dòng vào plasmit pUC18-serC.

Đoạn mồi PCR serAF(XbaI) có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn XbaI (TCTAGA). Các nucleotit 12 đến 33 gắn kết trong hệ gen MG1655 của *E. Coli* từ vị trí 3055199 đến 3055220. Đoạn mồi PCR serAR(SHSSNB) có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết các endonucleaza giới hạn SacI, HindIII, SphI, SmaI, NotI và BglII. Các nucleotit 49 đến 58 gắn kết trong hệ gen MG1655 của *E. coli* từ vị trí 3056880 đến 3056861.

serAF(XbaI) (SEQ ID NO: 27)

5' CTGTAGTCTAGATTAGTACAGCAGACGGGCGCG 3'

serAR(SHSSNB) (SEQ ID NO: 28)

5'
CAAGAGCTCAAGCTTGCATGCGATTCCCGGGCGGCCGCAATAAGATCTCC
GTCAGGGCGTGGTGACCG 3'

Gen serA được khuếch đại bằng cách sử dụng phản ứng chuỗi polymeaza (PCR) bằng các đoạn mồi serAF(XbaI) và serAR(SHSSNB) và bằng polymeaza ADN Phusion (Finnzymes Oy, Espoo, Finland). ADN của hệ gen của MG1655 *E. coli* đóng vai trò làm khuôn mẫu. Mảnh ADN thu được có kích cụ 1731 bp.

Mảnh này được phân cắt bằng các endonucleaza giới hạn XbaI và SacI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Tương tự, plasmit pUC18-serC được phân cắt bằng các endonucleaza giới hạn XbaI và SacI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Sau đó,

Plasmit đã được phân cắt này được ghép nối với sản phẩm PCR và được biến nạp vào *E. coli* DH5α. Dòng plasmit chứa gen serA được nhận diện bằng cách phân cắt giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pUC18-serAC.

1.3 Nhân dòng gen serB vào plasmit pUC18-serAC

Gen serB từ *Escherichia coli* MG1655 được khuếch đại bằng phản ứng chuỗi polymeaza (PCR) và sau đó được tách dòng vào plasmit pUC18-serAC.

Đoạn mồi PCR serB(SphI) có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn SphI (GCATGC). Nucleotit 13 đến 34 gắn kết trong hệ gen *E. coli* MG1655 từ vị trí 4622816 đến 4622837.

đoạn mồi PCR serB(SmaI) có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết các endonucleaza giới hạn SalI (GTCGAC) và SmaI (CCCGGG). Nucleotit 54 đến 75 gắn kết trong hệ gen *E. coli* MG1655 từ vị trí 4623887 đến 4623866.

serB(SphI) (SEQ ID NO: 29)

5' CCATGCGCATGCCCACCCTTTGAAAATTTGAGAC 3'

serB(SmaI) (SEQ ID NO: 30)

5'

CCGCATGTCGACATCCCGGGGCAGAAAGGCCCACCCGAAGGTGAGCCAGT
GTG

ATTACTTCTGATTCAGGCTGCC 3'

Gen serB được khuếch đại bằng cách sử dụng phản ứng chuỗi polymeaza (PCR) bằng các đoạn mồi serB(SphI) và serB(SmaI) và bằng polymeaza ADN Phusion (Finnzymes Oy, Espoo, Finland). ADN của hệ gen của *E. coli* MG1655 đóng vai trò làm khuôn mẫu. Mảnh ADN thu được có kích cụ 1137 bp.

Mảnh này được phân cắt bởi các endonucleaza giới hạn SphI và SmaI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Tương tự, plasmit pUC18-serAC được phân cắt bằng các endonucleaza giới hạn SphI và SmaI, được khử phosphoryl bằng phosphataza kiềm và được tinh chế bằng kit tinh chế PCR

QIAquick (Qiagen, Hilden, Germany). Khi đó, plasmit đã được phân cắt này được ghép nối với sản phẩm PCR và được biến nạp vào *E. coli* DH5α. Các dòng plasmit chứa gen serB được nhận diện bằng cách phân cắt giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pUC18-serBAC.

1.4 Nhân dòng gen glyA vào plasmit pUC18-serBAC

Gen glyA từ *Escherichia coli* MG1655 được khuếch đại bằng phản ứng chuỗi polymeaza (PCR) và sau đó được tách dòng vào plasmit pUC18-serBAC.

Đoạn mồi PCR nằm sau glyA có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn BglII (AGATCT). Nucleotit 13 đến 35 gắn kết trong hệ gen *E. coli* MG1655 từ vị trí 2682063 đến 2682085.

Đoạn mồi PCR nằm trước glyA có ở đầu 5' của nó 6 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn NotI (GCGGCCGC). Nucleotit 15 đến 33 gắn kết trong hệ gen *E. coli* MG1655 từ vị trí 2683762 đến 2683744.

nằm sau glyA (SEQ ID NO: 31)

5' ATCTAAAGATCTGTTACGACAGATTTGATGGCGCG 3'

nằm trước glyA (SEQ ID NO: 32)

5' TTCATCGCGGCCGCGAAAGAATGTGATGAAGTG 3'

Gen glyA được khuếch đại bằng cách sử dụng phản ứng chuỗi polymeaza (PCR) bằng đoạn mồi nằm sau glyA và nằm trước glyA và bằng polymeaza ADN Phusion (Finnzymes Oy, Espoo, Finland). ADN của hệ gen của *E. coli* MG1655 đóng vai trò làm khuôn mẫu. Thu được mảnh ADN có kích cụ 1726 bp.

Mảnh này được phân cắt bằng các endonucleaza giới hạn BglII và NotI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Tương tự, plasmit pUC18-serBAC được phân cắt bằng các endonucleaza giới hạn BglII và NotI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Sau đó, plasmit đã được phân cắt này được ghép nối với sản phẩm PCR và được biến nạp vào *E. coli* DH5α. Dòng plasmit chứa gen glyA được nhận diện bằng cách phân cắt giới hạn và xác định trình tự ADN. Plasmit thu được được gọi là pUC18-serB-glyA-serAC.

1.5 Nhân dòng các gen serB-glyA-serAC từ pUC18-serB-glyA-serAC vào pCC1-BAC

Plasmit pUC18-serB-glyA-serAC được phân cắt bằng endonucleaza giới hạn HindIII, và các mảnh AND này được tách phân đoạn bằng cách điện di trên gel agaroza. Mảnh ADN 5,9 kb chứa các gen serB, glyA, serA và serC được phân lập từ gel này. Mảnh này được ghép nối với vật truyền sẵn dàng nhân dòng là plasmit pCC1BAC (Hind III) từ Epicentre /Madison, USA), mà đã được phân cắt bằng HindIII, và được biến nạp vào *E. coli* EPI300. Các dòng plasmit chứa mảnh ADN chứa serB, glyA, serA và serC được nhận diện bằng cách phân cắt giới hạn và xác định trình tự ADN. Plasmit sản xuất thu được được gọi là pCC3.

1.6 Nhân dòng plasmit sản xuất pME-RDL2a

Plasmit sản xuất pME-RDL2a được tách dòng như được mô tả trong đơn EP số 11151526.8. Plasmit này bao gồm của gen cysE *Escherichia coli*, các alen tính kháng cơ chế ức chế ngược của các gen thrA và metA của *Escherichia coli*, và gen RDL2a mã hoá thiosulphat sulphurtransferaza RDL2p của *Saccharomyces cerevisiae*. Ngoài ra, plasmit này bao gồm gen kháng streptomycin.

1.7 Biến nạp chủng ECM2 bằng plasmit sản xuất

Chủng ECM2 được biến nạp bằng plasmit sản xuất pCC3 của Ví dụ 1.5, và các thể biến nạp được chọn lọc bằng 20 mg/l chloramphenicol. Sau đó, các tế bào này được biến nạp bằng plasmit pME-RDL2a của Ví dụ 1.6 và các thể biến nạp thu được được chọn lọc bằng 20 mg/l chloramphenicol + 100 mg/l streptomycin. Chủng thu được được gọi là ECM2/pCC3/pME-RDL2a.

1.8 Sàng lọc và xác định trình tự các đột biến dung nạp L-metionin

Các đột biến dung nạp L-metionin được chọn lọc bằng cách dàn mỏng tiền giống cấy chủng ECM2/pCC3/pME-RDL2a trên các đĩa chứa môi trường tối thiểu PC1 (Bảng 1, với thạch 14 g/l) chứa 75g/l L-metionin (Merck, Frankfurt, Germany). Tiền giống cấy này chứa 10 ml môi trường tiền nuôi cấy (10% môi trường LB chứa 2,5 g/l glucoza và 90% môi trường tối thiểu PC1) được cấy 100µl dịch nuôi cấy tế bào và được nuôi cấy ở nhiệt độ 37°C trong thời gian 10 giờ.

Bảng 1: môi trường tối thiểu PC1

Chất	Nồng độ
ZnSO4 * 7 H2O	4 mg/l
CuCl2 * 2 H2O	2 mg/l
MnSO4 * H2O	20 mg/l
H3BO3	1 mg/l
Na2MoO4 * 2 H2O	0,4 mg/l
MgSO4 * 7 H2O	1 g/l
Axit xitric * 1 H2O	6,56 g/l
CaCl2 * 2 H2O	40 mg/l
K2HPO4	8,02 g/l
Na2HPO4	2 g/l
(NH4)2HPO4	8 g/l
NH4Cl	0,13 g/l
(NH4)2SO3	5,6 g/l
MOPS	5 g/l
NaOH 10M	được điều chỉnh đến pH 6,8
FeSO4 * 7 H2O	40 mg/l
Thiamin hyđroclorua	10 mg/l
Vitamin B12	10 mg/l
Glucoza	10 g/l
Isopropyl-thio-β-galactosit (IPTG)	2,4 mg/l
Spectinomycin	50 mg/l
Chloramphenicol	20 mg/l

Sau 5 ngày ủ, các khuẩn lạc đơn của các đột biến dung nạp L-metionin được phân lập từ các đĩa.

Để chuẩn bị cho việc xác định trình tự, các dẫn xuất của các đột biến dung nạp L-metionin của chủng ECM2/pCC3/pME-RDL2a mà không còn chứa plasmit pCC3 và pME-RDL2a được phân lập sau nhân trong môi trường LB không chứa chất kháng sinh cho khoảng sáu thế hệ. Các chủng thu được này là nhạy với streptomycin- và chloramphenicol.

Phân lập ADN nhiễm sắc thể từ 8 đột biến. ADN này được sử dụng cho việc xác định trình tự toàn bộ hệ gen (GATC, Constance, Germany).

So với chủng ECM2 này bắt đầu, nhận thấy chỉ có đột biến ở gen *proP*. Bảng 2 sau đây liệt kê các đột biến này. Trình tự của các alen *proP* được thể hiện ở SEQ ID No: 9 đến SEQ ID No: 24.

Bảng 2:

Chủng	đột biến ở gen *proP*	Loại đột biến	Tên đột biến
DM2321	E412*	Thay thế AA	M8 (SEQ ID NO:9-10)
DM2322	Cài xen sau C nucleotit vị trí 842	Cài xen nucleotit	M3 (SEQ ID NO:11-12)
DM2323	Làm khuyết nằm sau A nucleotit vị trí 854	Làm khuyết nucleotit	M4 (SEQ ID NO:13-14)
DM2324	Y467H	Thay thế AA	M11 (SEQ ID NO:15-16)
DM2325	Làm khuyết 51bp từ vị trí nucleotit 1173 đến 1223	Làm khuyết nucleotit	M7 (SEQ ID NO:17-18)
DM2326	cài xen 19 bp nằm sau nucleotit vị trí 973	Cài xen nucleotit	M6 (SEQ ID NO:19-20)
DM2327	R324L	Thay thế AA	M5 (SEQ ID NO:21-22)
DM2328	cài xen 1359 bp nằm sau vị trí nucleotit 183	Cài xen nucleotit	M2 (SEQ ID NO:23-24)

Theo hiệp ước Budapest, các chủng DM2321, DM2322, DM2323, DM2324, DM2325, DM2326, DM2327 và DM2328 được nộp nộp lưu ở DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Inhoffenstraòe 7 B, 38124 Brunswick, Germany) với dữ liệu DSM DSM 25095 (= DM2321), 25096 (= DM2322), 25097 (= DM2323), 25098 (= DM2324), 25099 (= DM2325), 25100 (= DM2326), 25101 (= DM2327), và 25102 (= DM2328) ngày 23/8/2011.

Ví dụ 2

Phát hiện hoạt tính của đột biến *proP* dung nạp L-metionin

Hoạt tính ProP của đột biến *proP* được xác định bằng cách phân tích ví dụ sự hấp thu prolin của chủng DM2328 so sánh với chủng ban đầu ECM2. Cho mục đích này, 10 ml tiền giống cấy 1 được ủ bằng cách lắc trong môi trường LB ở nhiệt độ 37°C trong ngày. 1 ml tiền giống cấy được chuyển vào 20 ml môi trường M9 là tiền giống cấy 2 và được nuôi cấy qua đêm. Tiền giống cấy 2 được cấy đến OD600 bằng 0,2 trong 50 ml môi trường M9 mới. Sau đó, các tế bào này được nuôi cấy trong thời gian 3-4 giờ, tiếp theo rửa ba lần bằng dung dịch đệm M9 được làm lạnh bằng đá, được điều chỉnh đến OD600 bằng 2 và được bảo quản trên đá.

Sự hấp thu prolin được xác định bằng [14C(U)]L-prolin được đánh dấu phóng xạ (hoạt tính đặc hiệu: 1,85 MBq; Hartmann Analytic, Germany). Cho mục đích này, trong mỗi trường hợp 2,3 ml huyền phù tế bào được cho vào thùng có khuấy và được bổ sung năng lượng ở nhiệt độ 37°C trong thời gian 2 phút bằng cách bổ sung 10 mM glucoza. Việc đo lường vận chuyển được bắt đầu cùng với việc khuấy bằng cách bổ sung 20 µl L-[14C]-prolin ở nồng độ 24,6 mM. Ở các thời điểm khác nhau, (0, 15, 30, 45, 60, 75, 90, 105 và 120 giây) lấy ra 200 µl trong mỗi trường hợp và được hút bằng pipet trên các bộ lọc bằng sợi thủy tinh (Millipore). Môi trường xung quanh được loại ra bằng cách hút bằng cách xử lý bằng cách lọc chân không, và sau đó, các tế bào này được rửa hai lần bằng 2,5 ml dung dịch 0,1 M LiCl. Các bộ lọc được xử lý bằng 3,8 ml dung dịch nhấp nháy (Roth, Karlsruhe, Germany), và đo hoạt tính phóng xạ bằng máy đếm nhấp nháy (LS 6500, Beckman Instruments, Munich, Germany). Tổng hoạt tính trong hỗn hợp phản ứng được xác định bằng cách đo mẫu trực tiếp mà không cần lọc. Đo mức rã trong một phút (dpm) cho mỗi mẫu. Tính hoạt tính hấp thu theo nmol/phút (mg chất khô) (0,36 = hệ thức chất khô của *E. coli* [mg/ml OD=1]). Mức vận chuyển theo nmol/mg*phút là thu được từ phần tuyến tính của các động học hấp thu. Bảng 3 sau đây cho số trung bình của ba đo lường song song.

Bảng 3:

Chủng	Mức vận chuyển L-prolinnmol/mg*phút)
ECM2	2,39
DM2328	1,27

Ví dụ 3

Biến nạp các chủng DM2321, DM2322, DM2323, DM2324, DM2325, DM2326, DM2327 và DM2328 bằng plasmit sản xuất

Các chủng DM2321, DM2322, DM2323, DM2324, DM2325, DM2326, DM2327 và DM2328 được biến nạp bằng plasmit sản xuất pCC3 của Ví dụ 1.5, và các thể biến nạp được chọn lọc bằng 20 mg/l chloramphenicol. Sau đó, các tế bào này được biến nạp bằng plasmit pME-RDL2a của Ví dụ 1.6, và các thể biến nạp thu được được chọn lọc bằng 20 mg/l chloramphenicol và 100 mg/l streptomycin. Các chủng thu được được gọi là DM2321/pCC3/pME-RDL2a, DM2322/pCC3/pME-RDL2a, DM2323/pCC3/pME-RDL2a, DM2324/pCC3/pME-RDL2a, DM2325/pCC3/pME-RDL2a, DM2326/pCC3/pME-RDL2a, DM2327/pCC3/pME-RDL2a và DM2328/pCC3/pME-RDL2a.

Ví dụ 4

Thử nghiệm đặc tính trong thử nghiệm bình lắc-bình thót cổ

Đặc tính của chủng sản xuất L-metionin *E. coli* được đánh giá bằng các thử nghiệm sản xuất trong bình thót cổ hình nón dung tích 100 ml. Tiền giống cấy trong mỗi trường hợp 10 ml môi trường tiền nuôi cấy (10% môi trường LB chứa 2,5 g/l glucoza và 90% môi trường tối thiểu PC1) đã được cấy 100 µl nuôi cấy tế bào được nuôi cấy ở nhiệt độ 37°C trong thời gian 10 giờ. Sau đó, các tiền giống cấy này được dùng để cấy trong mỗi trường hợp 10 ml môi trường tối thiểu PC1 (xem Bảng 1 ở Ví dụ 1) đến OD 600 nm bằng 0,2 (Eppendorf Bio-Photometer; Eppendorf AG, Hamburg, Germany) và môi trường nuôi cấy được nuôi cấy ở nhiệt độ 37°C trong thời gian 24 giờ. Nồng độ L-metionin ngoại bào được xác định bằng phương pháp sắc ký trao đổi ion và được tạo dẫn xuất sau cột bằng cách phát hiện bằng ninhyđrin bằng cách sử

dụng thiết bị phân tích axit amin (Sykam GmbH, Eresing, Germany). Nồng độ glucoza ngoại bào được xác định bằng cách sử dụng thiết bị phân tích glucoza YSI 2700 Select Glucose Analyzer (YSI Life Sciences, Yellow Springs, Ohio, USA). Các kết quả được đưa ra trong Bảng 4. Sau 24 giờ glucoza đã được sử dụng hoàn toàn trong cả hai nuôi cấy.

Bảng 4: Nồng độ L-metionin trong môi trường lên men lỏng của các chủng *E. coli* được thử nghiệm

Chủng	OD(600nm)	L-Metionin (g/l)
ECM2/pCC3/pME-RDL2a	3,27	1,71
DM2321/pCC3/pME-RDL2a	3,04	1,83
DM2322/pCC3/pME-RDL2a	3,38	1,83
DM2323/pCC3/pME-RDL2a	3,20	1,83
DM2324/pCC3/pME-RDL2a	3,38	1,82
DM2325/pCC3/pME-RDL2a	3,44	1,83
DM2326/pCC3/pME-RDL2a	3,23	1,87
DM2327/pCC3/pME-RDL2a	3,22	1,99
DM2328/pCC3/pME-RDL2a	3,12	1,86

Ví dụ 5

Nhân dòng các alen *proP* M8, M4 và M6 vào plasmit pMAK705

Các alen *proP* M8 từ DM2321, M4 từ DM2323 và M6 từ DM2326 được khuếch đại bằng phản ứng chuỗi polymeaza (PCR) và sau đó được tách dòng vào plasmit pMAK705.

Cặp đoạn mồi được thiết kế trên cơ sở các trình tự thu được cho các alen *proP* (xem Ví dụ 1), trong mỗi trường hợp làm cho mảnh chứa đột biến nhất định được khuếch đại. Sau đó, mảnh này có thể được tách dòng vào plasmit pMAK705 (Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR (1989); J Bacteriol.; 171(9): 4617-4622).

Đoạn mồi PCR proPmut1(NotI) có ở đầu 5´ của nó 4 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn HindIII.

Đoạn mồi PCR proPmut2(BamHI) có ở đầu 5´ của nó 4 nucleotit ngẫu nhiên tiếp đó là trình tự nhận biết endonucleaza giới hạn BamHI.

Các nucleotit 13 đến 32 của proPmut1(NotI) gắn kết từ vị trí 4328800 đến 4328819 trong hệ gen của *E. coli* MG1655. Các nucleotit 13 đến 37 của proPmut2(NotI) gắn kết từ vị trí 4330303 đến 4330322 ở hệ gen *E. coli* MG1655.

proPmut1(HindIII) (SEQ ID NO: 33)

5' GTCAAAGCTT ATATGGTCGCCAGAAGATCC 3'

proPmut2(BamHI) (SEQ ID NO: 34)

5' GTCAGGATCC TCAGCCGCATTACACAGTTG 3'

ADN hệ gen của các chủng DM2321, DM2323 và DM2328 (xem Ví dụ 1) đóng vai trò làm khuôn mẫu.

Trong mỗi trường hợp, mảnh nối đột biến tương ứng ở gen *proP* được khuếch đại bằng cách sử dụng phản ứng chuỗi polymeaza (PCR) bằng polymeaza ADN Phusion (Finnzymes Oy, Espoo, Finland). Mảnh từ DM2321 (mang proP-M8), chứa đột biến E412*, có kích cμ 1564bp (SEQ ID NO: 35). Mảnh từ DM2323 (mang proP-

M4), chứa khuyết một "A" nằm sau vị trí nucleotit 854, có kích cμ 1542bp (SEQ ID NO: 36). Mảnh từ DM2326 (mang proP-M6), chứa đoạn cài xen 19bp nằm sau vị trí nucleotit 973, có kích cμ 1563bp (SEQ ID NO: 37).

Tất cả 3 mảnh này được phân cắt bằng bằng endonucleaza giới hạn HindIII und BamHI và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden, Germany). Sau đó, các mảnh này được sử dụng để nối với plasmit pMAK705.

Plasmit pMAK705 được phân cắt bởi endonucleaza giới hạn HindIII und BamHI, được khử phosphoryl bằng phosphataza kiềm (Alkaline Phosphatase, Calf Intestinal, New England Biolabs, Frankfurt a.M., Germany) và được tinh chế bằng kit tinh chế PCR QIAquick (Qiagen, Hilden). Sau đó, plasmit này được phối trộn với các *proP* nhất định và được ghép nối bằng ligaza Quick DNA (New England BioLabs, Frankfurt, Germany). Các hỗn hợp nối này được biến nạp vào *E.coli* DH5α (Grant et al.; Proceedings of the National Academy of Sciences USA, 87 (1990) 4645-4649). Các dòng plasmit chính xác được chọn lọc bằng 20 mg/l chloramphenicol và được nhận diện bằng cách phân cắt giới hạn và sau đó xác định trình tự các đoạn xen. Các plasmit thu được như vậy được gọi là pMAK_proP-M4, pMAK_proP-M6 và pMAK_proP-M8. pMAK_proP-M8 được thể hiện bằng cách ví dụ trên Fig 3.

Ví dụ 6

Gây đột biến gen *proP* ở chủng *E.coli* ECM2

Alen kiểu dại *proP* trên nhiễm sắc thể ở chủng *E.coli* ECM2 sản xuất L-metionin (xem Ví dụ 1) được thay thế trong mỗi trường hợp bằng các alen *proP* M4, M6 và M8 mà làm trung gian tính kháng metionin. Cho mục đích này, chủng này được biến nạp trong mỗi trường hợp bằng phương pháp đục lỗ điện bằng các plasmit pMAK_proP-M4, pMAK_proP-M6 và pMAK_proP-M8 (xem Ví dụ 5). Các plasmit pMAK có gen có tính kháng chloramphenicol và điểm khởi đầu sao chép nhạy cảm nhiệt độ. Plasmit này được sao chép bằng *E.coli* ở nhiệt độ 30°C, nhưng không phải ở 44°C. Các hỗn hợp biến nạp được dàn mỏng trong mỗi trường hợp trên thạch LB chứa 20mg/l chloramphenicol và được ủ ở nhiệt độ 30°C trong thời gian 40 giờ. Sau đó, các tế bào này được bỏ ra bằng cách sử dụng vòng cấy, tái tạo huyền phù trong môi trường

LB và pha loãng 10000 lần trong môi trường LB. 100 µl pha loãng này được dàn mỏng lên thạch LB chứa 20mg/l chloramphenicol và được ủ ở nhiệt độ 44°C trong thời gian 24 giờ nữa. Kết quả là, chọn lọc các khuẩn lạc trong đó các plasmit pMAK_proP-M4, pMAK_proP-M6 và pMAK_proP-M8 được hợp nhất vào nhiễm sắc thể trong mỗi trường hợp. Trong mỗi trường hợp một trong các khuẩn lạc này được dàn mỏng trên thạch LB chứa 20mg/l chloramphenicol bằng cách sử dụng vòng cấy và được ủ ở nhiệt độ 44°C trong thời gian 24 giờ. Đột biến đưa vào được phát hiện bằng phương pháp PCR chuẩn (Innis et al. (1990) PCR Protocols. A guide to methods and applications, Academic Press) bằng cách sử dụng cặp đoạn mồi sau (xem Ví dụ 11):

proPmut1(HindIII) (SEQ ID NO: 33)

5' GTCAAAGCTT ATATGGTCGCCAGAAGATCC 3'

proPmut2(BamHI) (SEQ ID NO: 34)

5' GTCAGGATCC TCAGCCGCATTACACAGTTG 3'

Sản phẩm PCR thu được trong mỗi trường hợp được xác định trình tự bằng cách sử dụng cả hai đoạn mồi ở GATC (Constance, Germany). Tất cả ba alen *proP* được chuyển nạp trong mỗi trường hợp vào chủng *E.coli* ECM2, với các chủng thu được được gọi là ECM2_proP-M4, ECM2_proP-M6 và ECM2_proP-M8.

Ví dụ 7

Biến nạp các chủng ECM2_proP-M4, ECM2_proP-M6 và ECM2_proP-M8 bằng các plasmit sản xuất

Các chủng ECM2_proP-M4, ECM2_proP-M6 và ECM2_proP-M8 được biến nạp bằng plasmit sản xuất pCC3 của Ví dụ 1.5 và các thể biến nạp được chọn lọc bằng 20 mg/l chloramphenicol. Sau đó, các tế bào này được biến nạp bằng plasmit pME-RDL2a của Ví dụ 1.6 và các thể biến nạp thu được được chọn lọc bằng 20 mg/l chloramphenicol và 100 mg/l streptomycin. Các chủng thu được được gọi là ECM2_proP-M4/pCC3/pME-RDL2a, ECM2_proP-M6/pCC3/pME-RDL2a và ECM2_proP-M8/pCC3/pME-RDL2a.

Ví dụ 8

Thử nghiệm đặc tính trong thử nghiệm bình lắc-bình thót cổ

Đặc tính của các chủng sản xuất L-metionin *E. coli* ECM2_proP-M4/pCC3/pME-RDL2a, ECM2_proP-M6/pCC3/pME-RDL2a và ECM2_proP-M8/pCC3/pME-RDL2a được đánh giá bằng thử nghiệm sản xuất trong bình thót cổ hình nón dung tích 100 ml. Tiền giống cấy trong mỗi trường hợp gồm 10 ml môi trường tiền nuôi cấy (10% môi trường LB chứa 2,5 g/l glucoza và 90% môi trường tối thiểu PC1) đã được cấy 100 µl nuôi cấy tế bào được nuôi cấy ở nhiệt độ 37°C trong thời gian 10 giờ. Sau đó, các tiền giống cấy này được dùng để cấy trong mỗi trường hợp 10 ml môi trường tối thiểu PC1 (xem Bảng 1 ở Ví dụ 1) đến OD 600 nm bằng 0,2 (Eppendorf Bio-Photometer; Eppendorf AG, Hamburg, Germany) và môi trường nuôi cấy được nuôi cấy ở nhiệt độ 37°C trong thời gian 24 giờ. Nồng độ L-metionin ngoại bào được xác định bằng phương pháp sắc ký trao đổi ion và được tạo dẫn xuất sau cột bằng cách phát hiện bằng ninhyđrin bằng cách sử dụng thiết bị phân tích axit amin (Sykam GmbH, Eresing, Germany). Nồng độ glucoza ngoại bào được xác định bằng cách sử dụng thiết bị phân tích YSI 2700 Select Glucose Analyzer (YSI Life Sciences, Yellow Springs, Ohio, USA). Các kết quả được thê rhiện trong Bảng 5. Sau 24 giờ glucoza đã được sử dụng hoàn toàn trong cả hai nuôi cấy.

Bảng 5: Nồng độ L-metionin trong môi trường lên men lỏng của chủng *E. coli* được thử nghiệm

Chủng	OD(600nm)	L-metionin (g/l)
ECM2/pCC3/pME-RDL2a	3,27	1,71
ECM2_proP-M4/pCC3/pME-RDL2a	3,38	1,83
ECM2_proP-M6/pCC3/pME-RDL2a	3,04	1,87
ECM2_proP-M8/pCC3/pME-RDL2a	3,22	1,83

YÊU CẦU BẢO HỘ

1. Quy trình sản xuất L-axit amin hoặc chất phụ gia thức ăn chứa L-axit amin bằng cách lên men vi sinh vật thuộc họ *Enterobacteriaceae*, khác biệt ở chỗ, sử dụng vi sinh vật trong đó gen *proP* được làm suy giảm, với kết quả là hoạt tính hay nồng độ của protein ProP được làm giảm xuống còn từ 0 đến 75% hoạt tính hay nồng độ của protein trong vi sinh vật mà ko phải là tái tổ hợp đối với protein tương ứng.

2. Quy trình theo điểm 1, khác biệt ở chỗ, gen *proP* đã được tắt đi.

3. Quy trình theo điểm 1 và 2, khác biệt ở chỗ, vi sinh vật sản xuất quá mức L-axit amin.

4. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, vi sinh vật cho L-axit amin tích lũy trong môi trường.

5. Quy trình theo điểm bất kỳ trong số các điểm từ 1 đến 4, khác biệt ở chỗ, vi sinh vật cho L-axit amin tích lũy trong tế bào.

6. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, L-axit amin là axit amin chứa lưu huỳnh, tốt hơn nếu là L-metionin, L-xystein, L-xystin, L-homoxystein hoặc L-homoxystin.

7. Quy trình theo điểm 6, khác biệt ở chỗ, axit amin chứa lưu huỳnh này là L-metionin.

8. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, vi sinh vật có khả năng chống chịu metionin tăng lên so với vi sinh vật không có gen *proP* được làm suy giảm, khả năng chống chịu metionin tăng lên được hiểu có nghĩa là vi sinh vật vẫn có khả năng tăng trưởng ở nồng độ L-metionin là 50gam/lít.

9. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, vi sinh vật thuộc họ *Enterobacteriaceae* là vi khuẩn thuộc giống *Escherichia*, *Erwinia*, *Providencia* hoặc *Serratia*, đặc biệt là thuộc giống *Escherichia*, đặc biệt tốt hơn nếu là *Escherichia coli*.

10. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, gen *proP* là gen có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là

ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự polynucleotit của SEQ ID NO: 1, 3, 5 hoặc 7, đặc biệt tốt hơn nếu so với trình tự polynucleotit của SEQ ID NO: 1.

11. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, gen *proP* được làm suy giảm là polynucleotit có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt tốt hơn nếu ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự polynucleotit của SEQ ID NO: 1 và mà ngoài ra nhất thiết có một hoặc nhiều đột biến sau so với polynucleotit của SEQ ID NO: 1:

a. thay thế bộ ba mã hoá L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b. thay thế bộ ba mã hoá L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c. thay thế bộ ba mã hoá L-axit glutamic ở vị trí 412 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá cođon dừng;

d. làm khuyết nucleobazơ ađenin ở vị trí 854 của gen *proP* của SEQ ID NO:1;

e. làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của gen *proP* của SEQ ID NO:1;

f. cài xen nucleobazơ xytosin ở vị trí 842 của gen *proP* của SEQ ID NO:1;

g. cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của gen *proP* của SEQ ID NO:1;

h. cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của gen *proP* của SEQ ID NO:1.

12. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, quá trình lên men được thực hiện trong môi trường chứa nguồn lưu huỳnh vô cơ.

13. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, L-axit amin được gây tích lũy trong môi trường lên men lỏng thu được và sau đó, nếu thích hợp, được phân lập, được gom và/hoặc được tinh chế.

14. Quy trình theo điểm bất kỳ trong số các điểm nêu trên, khác biệt ở chỗ, L-axit amin được phân lập hoặc được gom cùng với các thành phần từ môi trường lên men lỏng và/hoặc sinh khối.

15. Vi sinh vật thuộc họ *Enterobacteriaceae*, mang gen *proP* được làm suy giảm, khác biệt ở chỗ vi sinh vật này sản xuất L-axit amin và tốt hơn nếu tiết các chất này vào môi trường này, khác biệt ở chỗ, vi sinh vật này có một hoặc nhiều dấu hiệu được chọn từ nhóm sau đây:

a) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống vận chuyển CysPUWA thiosulphat/sulphat (EC 3.6.3.25),

b) polynucleotit được biểu hiện quá mức mã hoá CysH 3'-phosphoađenosin 5'-phosphosulphat ređuctaza (EC 1.8.4.8),

c) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của CysJI sulphit ređuctaza (EC 1.8.1.2),

d) polynucleotit được biểu hiện quá mức mã hoá CysK xystein synthaza A (EC 2.5.1.47),

e) polynucleotit được biểu hiện quá mức mã hoá CysM xystein synthaza B (EC 2.5.1.47),

f) polynucleotit được biểu hiện quá mức mã hoá CysE serin axetyltransferaza (EC 2.3.1.30),

g) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của hệ thống phân cắt GcvTHP-Lpd glyxin (EC 2.1.2.10, EC 1.4.4.2, EC 1.8.1.4),

h) polynucleotit được biểu hiện quá mức mã hoá LipA lipoyl synthaza (EC 2.8.1.8),

i) polynucleotit được biểu hiện quá mức mã hoá LipB lipoyl-protein ligaza (EC 2.3.1.181),

j) polynucleotit được biểu hiện quá mức mã hoá SerA phosphoglyxerat đehyđrogenaza (EC 1.1.1.95),

k) polynucleotit được biểu hiện quá mức mã hoá SerB 3-phosphoserin phosphataza (EC 3.1.3.3),

l) polynucleotit được biểu hiện quá mức mã hoá SerC 3-phosphoserin/phosphohyđroxythreonin aminotransferaza (EC 2.6.1.52),

m) polynucleotit được biểu hiện quá mức mã hoá GlyA serin hyđroxymetyltransferaza (EC 2.1.2.1),

n) polynucleotit được biểu hiện quá mức mã hoá ThrA aspartokinaza I và homoserin đehyđrogenaza I (EC 2.7.2.4, EC 1.1.1.3),

o) tăng cường hoạt động của enzym aspartate kinaza, đặc biệt là polynucleotit được biểu hiện quá mức mã hoá LysC aspartat kinaza (EC 2.7.2.4),

p) polynucleotit được biểu hiện quá mức mã hoá Hom homoserin đehyđrogenaza (EC 1.1.1.3),

q) polynucleotit được biểu hiện quá mức mã hoá MetX homoserin O-axetyltransferaza (EC 2.3.1.31),

r) polynucleotit được biểu hiện quá mức mã hoá MetA homoserin O-succinyltransferaza (EC 2.3.1.46),

s) polynucleotit được biểu hiện quá mức mã hoá MetB xystathionin gama synthaza (EC 2.5.1.48),

t) polynucleotit được biểu hiện quá mức mã hoá AecD β-C-S-lyaza (EC 4.4.1.8, còn được gọi là beta-lyaza),

u) polynucleotit được biểu hiện quá mức mã hoá MetC xystathionin beta-lyaza (EC 4.4.1.8),

v) polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza độc lập với MetE B12 (EC 2.1.1.14),

w)polynucleotit được biểu hiện quá mức mã hoá homoxystein S-metyltransferaza phụ thuộc MetH B12 (EC 2.1.1.13),

x) polynucleotit được biểu hiện quá mức mã hoá MetF metylen tetrahyđrofolat ređuctaza (EC 1.5.1.20),

y) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất BrnFE L-metionin ra khỏi *Corynebacterium glutamicum*,

z) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của chất xuất YgaZH valin *Escherichia coli* (b2682, b2683),

aa) polynucleotit được biểu hiện quá mức mã hoá chất vận chuyển *Escherichia coli* YjeH giả định (b4141),

bb) polynucleotit được biểu hiện quá mức mã hoá một hoặc nhiều hợp phần của PntAB pyriđin nucleotit transhyđrogenaza (EC 1.6.1.2),

cc) polynucleotit được biểu hiện quá mức mã hoá MetZ O-succinylhomoserin sulphhyđrylaza (EC 2.5.1.48),

dd) polynucleotit được biểu hiện quá mức mã hoá Pyc phosphoenolpyruvat carboxylaza (EC 4.1.1.31),

ee) polynucleotit được biểu hiện quá mức mã hoá RDL2 thiosulphat sulphurtransferaza (EC 2.8.1.1),

ff)polynucleotit được biểu hiện quá mức mã hoá thiosulphat—thiol sulphurtransferaza (EC 2.8.1.3),

gg) polynucleotit được biểu hiện quá mức mã hoá thiosulphat—đithiol sulphurtransferaza (EC 2.8.1.5),

hh) hoạt tính gia tăng của MetA homoserin O-succinyltransferaza (EC 2.3.1.46),

ii) hoạt tính gia tăng của CysE serin axetyltransferaza (EC 2.3.1.30),

jj) làm suy giảm gen *metJ* mã hoá chất điều hoà phiên mã sinh tổng hợp L-metionin (MetJ) (b3938, ECK3930),

kk) làm suy giảm gen pgi mã hoá glucoza-6-phosphat isomeraza (Pgi, EC No. 5.3.1.9) (b4025, ECK4017),

ll) làm suy giảm gen thrB mã hoá homoserin kinaza (ThrB, EC 2.7.1.39) (b0003, ECK0003),

mm) làm suy giảm gen metK mã hoá S-ađenosylmetionin synthaza (MetK, EC No. 2.5.1.6) (b2942, ECK2937),

nn) làm suy giảm gen dapA mã hoá đihyđrođipicolinat synthaza (DapA, EC No. 4.2.1.52) (b2478, ECK2474),

oo) làm suy giảm gen pck mã hoá phosphoenolpyruvat carboxykinaza (Pck, EC No. 4.1.1.49) (b3403, ECK3390),

pp) làm suy giảm gen purU mã hoá formyltetrahyđrofolat hydrolaza (PurU, EC No. 3.5.1.10) (b1232, ECK1227),

qq) làm suy giảm gen pykA mã hoá pyruvat kinaza II (PykA, EC No. 2.7.1.40) (b1854, ECK1855)

rr) làm suy giảm gen pykF mã hoá pyruvat kinaza I (PykF, EC 2.7.1.40) (b1676, ECK1672),

ss) làm suy giảm gen metQ mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0197, ECK0197),

tt) làm suy giảm gen metI mã hoá cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0198, ECK0198),

uu) làm suy giảm gen mã hoá metN cấu trúc siêu phân tử của chất vận chuyển L-metionin (MetQNI) (b0199, ECK0199),

vv) làm suy giảm gen dcd mã hoá đeoxyxytiđin-5'-triphosphat đeaminaza (Dcd, EC No. 3.5.4.13) (b2065, ECK2059),

ww) làm suy giảm gen yncA mã hoá N-axyltransferaza giả định (YncA, Khám phá trao đổi chất: Metabolic Explorer WO2010/020681) (b1448, ECK1442),

xx) làm suy giảm gen rpoS mã hoá yếu tố RpoS sigma (b2741, ECK2736),

yy) làm suy giảm chất điều hòa protein MetJ.

zz) làm suy giảm S-adenosylmetionin synthaza MetK,

trong đó "làm suy giảm" có nghĩa là hoạt tính hay nồng độ của protein tương ứng được giảm xuống còn từ 0 đến 75% hoạt tính hay nồng độ của protein trong vi sinh vật mà không phải là tái tổ hợp đối với protein tương ứng và trong đó "biểu hiện quá mức", "tăng cường" hoặc "hoạt tính gia tăng" có nghĩa là hoạt tính của protein tương ứng được tăng lên so với chủng loại hoang dại với hệ số ít nhất 2 lần.

16. Vi sinh vật theo điểm 15, khác biệt ở chỗ, vi sinh vật này làm cho L-axit amin tích lũy trong tế bào.

17. Vi sinh vật theo điểm 15, khác biệt ở chỗ, vi sinh vật này làm cho L-axit amin tích lũy trong môi trường.

18. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 15 đến 17, khác biệt ở chỗ, gen *proP* đã được tắt đi.

19. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 15 đến 18, khác biệt ở chỗ, vi sinh vật này có khả năng chống chịu metionin tăng lên so với vi sinh vật mà không gen *proP* được làm suy giảm, trong đó khả năng chống chịu metionin tăng lên được hiểu có nghĩa là vi sinh vật vẫn có khả năng tăng trưởng ở nồng độ L-metionin là 50gam/lít.

20. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 15 đến 19, khác biệt ở chỗ, vi sinh vật thuộc họ *Enterobacteriaceae* là vi khuẩn thuộc giống *Escherichia*, *Erwinia*, *Providencia* hoặc *Serratia*, tốt hơn nếu thuộc giống *Escherichia*, đặc biệt tốt hơn nếu là *Escherichia coli*.

21. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 15 đến 20, khác biệt ở chỗ, gen *proP* là gen có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự polynucleotit của SEQ ID NO: 1, của SEQ ID NO: 3, của SEQ ID NO: 5 hoặc của SEQ ID NO: 7, đặc biệt tốt hơn nếu của SEQ ID NO: 1.

22. Vi sinh vật theo điểm bất kỳ trong số các điểm từ 15 đến 21, khác biệt ở chỗ, gen *proP* được làm suy giảm là polynucleotit có mức đồng nhất trình tự ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự đầy đủ của polynucleotit của SEQ ID NO: 1 và mà ngoài ra nhất thiết có một hoặc nhiều đột biến sau so với polynucleotit của SEQ ID NO: 1:

a. thay thế bộ ba mã hoá L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b. thay thế bộ ba mã hoá L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c. thay thế bộ ba mã hoá L-axit glutamic ở vị trí 412 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá codon dừng;

d. làm khuyết nucleobazơ ađenin ở vị trí 854 của gen *proP* của SEQ ID NO:1;

e. làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của gen *proP* của SEQ ID NO:1;

f. cài xen nucleobazơ xytosin ở vị trí 842 của gen *proP* của SEQ ID NO:1;

g. cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của gen *proP* của SEQ ID NO:1;

h. cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của gen *proP* của SEQ ID NO:1.

23. Polynucleotit có mức đồng nhất ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự của SEQ ID NO: 1, khác biệt ở chỗ, polynucleotit này nhất thiết có một hoặc nhiều đột biến so với polynucleotit của SEQ ID NO: 1, được chọn từ:

a. thay thế bộ ba mã hoá L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b. thay thế bộ ba mã hoá L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c. thay thế bộ ba mã hoá L-axit glutamic ở vị trí 412 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng bộ ba mã hoá cođon dừng;

d. làm khuyết nucleobazơ ađenin ở vị trí 854 của gen *proP* của SEQ ID NO:1;

e. làm khuyết một hoặc nhiều nucleobazơ từ vị trí 1173 đến vị trí 1223, tốt hơn nếu làm khuyết tất cả nucleobazơ từ vị trí 1173 đến vị trí 1223, của gen *proP* của SEQ ID NO:1;

f. cài xen nucleobazơ xytosin ở vị trí 842 của gen *proP* của SEQ ID NO:1;

g. cài xen một hoặc nhiều nucleobazơ ở vị trí 973, tốt hơn nếu cài xen 19 nucleobazơ ở vị trí 973, của gen *proP* của SEQ ID NO:1;

h. cài xen một hoặc nhiều nucleobazơ ở vị trí 183, tốt hơn nếu cài xen 1359 nucleobazơ ở vị trí 183, của gen *proP* của SEQ ID NO:1.

24. Polynucleotit theo điểm 23, khác biệt ở chỗ, polynucleotit này được lai với một hoặc nhiều polynucleotit được chọn từ nhóm bao gồm polynucleotit bổ trợ với SEQ ID NO:1, polynucleotit bổ trợ với SEQ ID NO:3 và polynucleotit bổ trợ với SEQ ID NO:5, tốt hơn nếu polynucleotit bổ trợ với SEQ ID NO:1.

25. Vật truyền chứa polynucleotit theo điểm 23 hoặc 24.

26. Polypeptit có mức đồng nhất ít nhất 80%, tốt hơn nếu ít nhất 90%, đặc biệt là ít nhất 95%, đặc biệt là ít nhất 98%, 99% hoặc 100%, so với trình tự của SEQ ID NO: 2, khác biệt ở chỗ, polypeptit này nhất thiết có một hoặc nhiều đột biến so với polypeptit của SEQ ID NO: 2, được chọn từ:

a. thay thế L-arginin ở vị trí 324 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng axit amin được chọn từ nhóm bao gồm L-leuxin, L-isoleuxin và L-valin, tốt hơn nếu là L-leuxin;

b. thay thế L-tyrosin ở vị trí 467 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin bằng axit amin được chọn từ nhóm bao gồm L-lysin, L-arginin và L-histiđin, tốt hơn nếu là L-histiđin;

c. làm khuyết 17 axit amin từ vị trí 392 đến vị trí 408 của SEQ ID NO: 2 hoặc vị trí tương đương của trình tự axit amin;

d. làm khuyết lên đến 420 axit amin ở đầu tận C, tính theo trình tự của SEQ ID NO: 2, tốt hơn nếu làm khuyết 88, 163, 202, 212 hoặc 420 axit amin ở đầu tận C của trình tự axit amin của SEQ ID NO: 2.

27. Vi sinh vật tái tổ hợp mang polynucleotit theo điểm 23 hoặc 24 và/hoặc vật truyền theo điểm 25 và/hoặc polypeptit theo điểm 26.

Fig. 1: Bản đồ plasmit pCC3

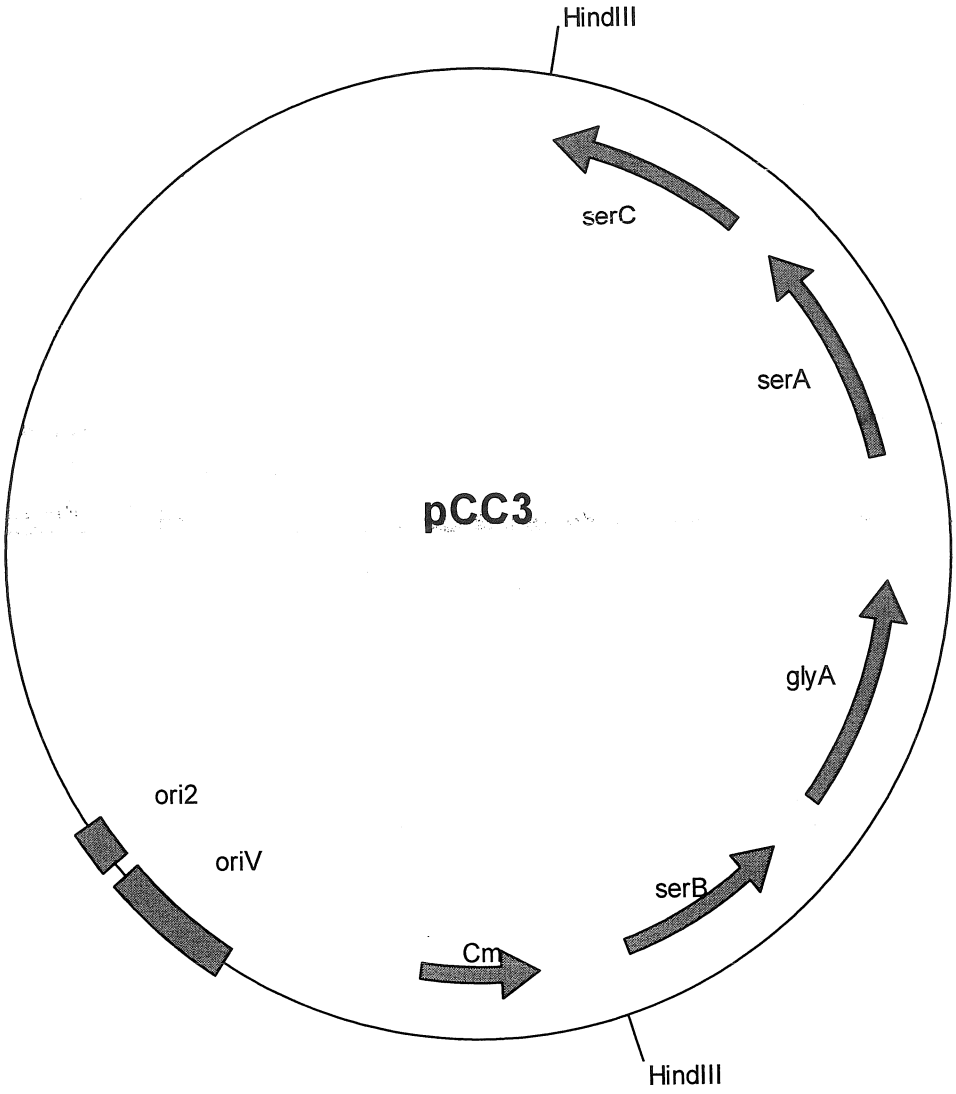


Fig. 2: Bản đồ plasmit pME-RDL2a

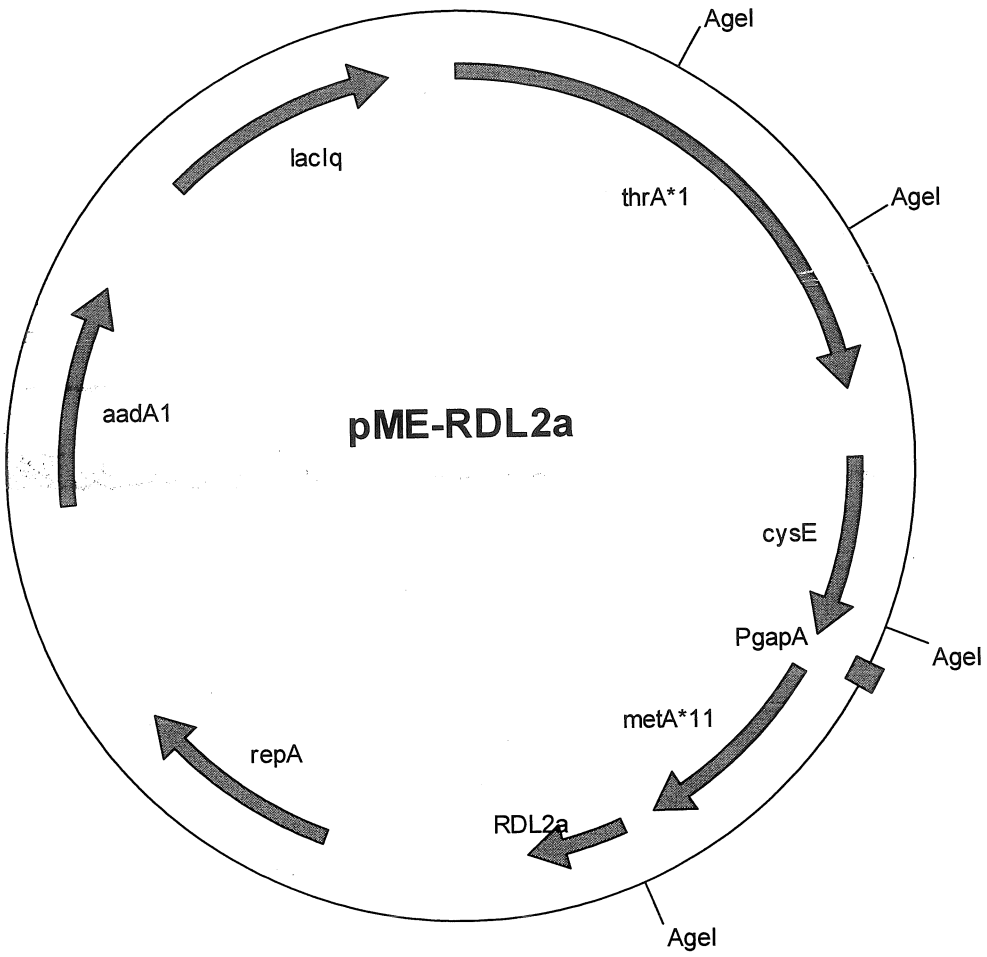
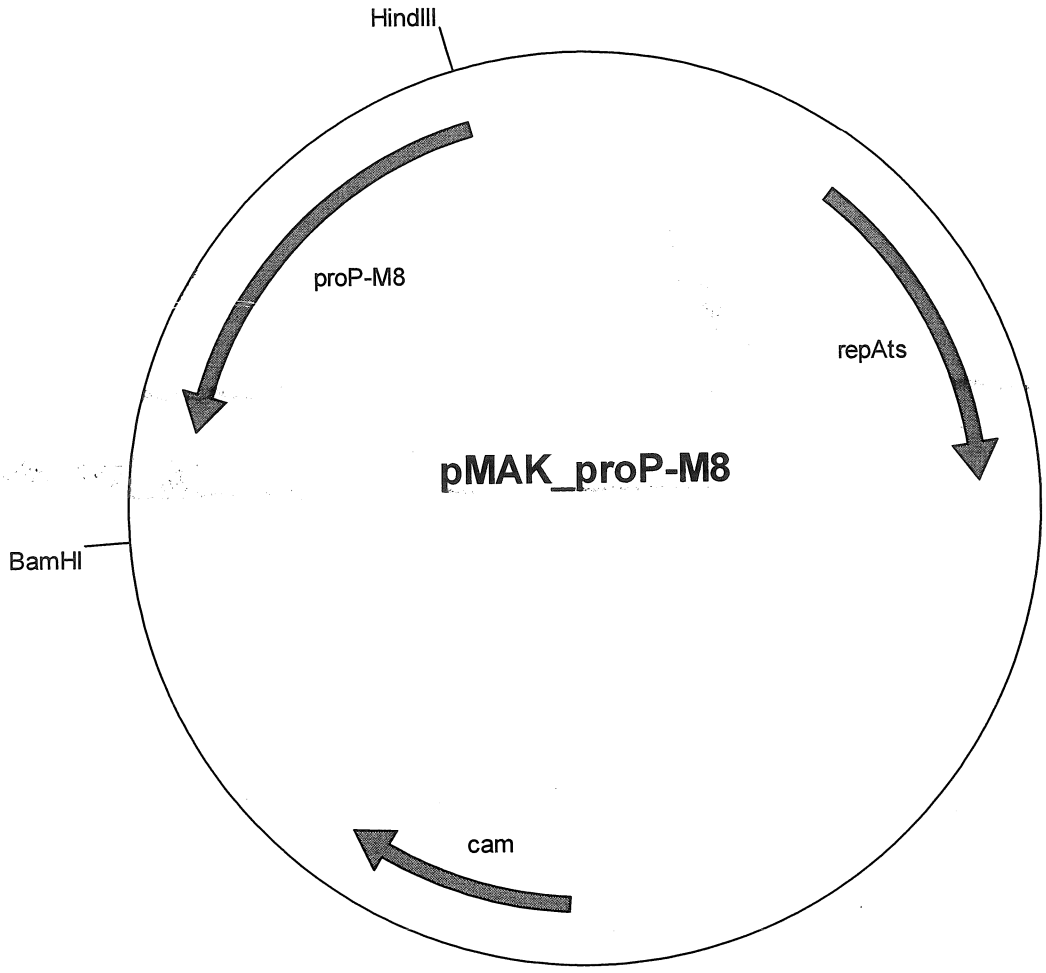


Fig. 3: Bản đồ plasmit pMAK_proP-M8



Danh mục trình tự

<110> Evonik Degussa GmbH

<120> QUY TRÌNH SẢN XUẤT L-AXIT AMIN HOẶC CHẤT PHỤ GIA THỨC ĂN GIA SÚC BẰNG CÁCH LÊN MEN VI SINH VẬT THUỘC HỌ *ENTEROBACTERIACEAE* VÀ VI SINH VẬT THUỘC HỌ *ENTEROBACTERIACEAE*

<130> 2011E00314

<160> 38

<170> PatentIn version 3.5

<210> 1
<211> 1503
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1503)
<223> Vùng mã hóa proP

<400> 1

```
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc          48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                  10                  15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg          96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct         144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag         192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg         240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag         288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt         336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att         384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa         432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa         480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160
```

```
cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt     1008
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
                325                 330                 335

gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc     1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt     1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc     1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcg gca ttt aat att tcg gtg ctg gtt gcc     1200
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385                 390                 395                 400

ggt ctg acg cca acg ctg gcg gcc tgg ctg gtc gaa agc tcg cag aat     1248
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
                405                 410                 415

ctg atg atg cct gcc tat tac ctg atg gta gtg gcg gtg gtt ggt tta     1296
```

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
420 425 430

atc acc ggc gta acc atg aaa gag acg gca aat cgt ccg ttg aaa ggt 1344
Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

gcg aca ccg gcg gcg tca gat ata cag gaa gcg aag gaa att ctc gtc 1392
Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
450 455 460

gag cat tac gat aat atc gag cag aaa atc gat gat att gac cac gag 1440
Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465 470 475 480

att gcc gat ttg cag gcg aaa cgt acc cgc ctg gtg cag caa cat ccg 1488
Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
485 490 495

cga att gat gaa taa 1503
Arg Ile Asp Glu
500

<210> 2
<211> 500
<212> PRT
<213> Escherichia coli

<400> 2

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

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn

405 410 415

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
420 425 430

Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
450 455 460

Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465 470 475 480

Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
485 490 495

Arg Ile Asp Glu
500

<210> 3
<211> 1503
<212> ADN
<213> Salmonella enterica

<220>
<221> CDS
<222> (1)..(1503)
<223> Vùng mã hóa proP

<400> 3

atg ctg aaa agg aaa aaa ata aaa ccg att aca ctg ggc gat gtg acc 48
Met Leu Lys Arg Lys Lys Ile Lys Pro Ile Thr Leu Gly Asp Val Thr
1 5 10 15

atc att gat gat ggt aaa ctt cgc aaa gcg att acc gcc gcc tcg ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggc aac gcg atg gag tgg ttt gat ttt ggt gtt tat gga ttt gtt gcc 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gcg ttg ggt aaa gtc ttt ttc ccc ggc gcc gat ccc agc gtc cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

atg att gcc gcg ctg gcc acg ttt tcc gtt ccc ttc ctg att cgt ccg 240
Met Ile Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65 70 75 80

ctc ggc ggg tta ttc ttt ggt atg ctc ggc gat aaa tac ggg cgc cag 288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

aag atc ctg gcg atc acg att gtg att atg tcg atc agt acc ttc tgt 336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

```
atc ggg tta atc ccc tct tac gcg acg atc ggt atc tgg gcg cca ata      384
Ile Gly Leu Ile Pro Ser Tyr Ala Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ttg ttg ctg tgt aaa atg gcg cag ggc ttc tcg gtt ggc ggg gaa      432
Leu Leu Leu Leu Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggc gcg tcg atc ttt gtc gcg gaa tat tcg ccg gat cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgc gga ttt atg gga agc tgg ctg gat ttt ggt tct atc gcc gga ttc      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggc gcg ggc gtg gtg gtc ttg atc tcg acg att gtc ggc gag      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gag aat ttc ctt gag tgg ggc tgg cgt att ccg ttc ttt atc gcc ctg      624
Glu Asn Phe Leu Glu Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

cca ttg ggg att att ggt ctc tac tta cgc cat gcg ctg gag gag acg      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

cca gcg ttt cag cag cac gtg gat aaa ctg gag cag ggc gac cgc gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggg ttg cag gat ggg ccg aaa gtc tcc ttt aaa gag att gcc acc aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

cac tgg cgt agc ctg ttg tca tgt atc ggt ctg gtg att gcc acc aac      816
His Trp Arg Ser Leu Leu Ser Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg acc tac tac atg ctg ctc acc tac atg ccg agc tac ctg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cac tat tct gaa gat cac ggc gtg ttg att atc atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggg atg ctg ttt gtg cag ccg gtg atg ggg ctg ctg agc gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttc ggt cga cgt cca ttt gtg att atg ggc agc att gcg ctg ttc     1008
Arg Phe Gly Arg Arg Pro Phe Val Ile Met Gly Ser Ile Ala Leu Phe
                325                 330                 335

gcg ctg gcg atc ccg gcc ttc atc ctg att aac agt aac gtt att ggc     1056
Ala Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gca ggt ttg ttg atg ctg gcg gtg att ctg aac tgc ttt     1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acc ggg gtg atg gcc tcg aca ttg ccg gcg atg ttt ccg acg cat att     1152
```

```
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tat agc gcg ctg gcg gcg gct ttt aat atc tct gta ttg att gcc      1200
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Ile Ala
385                 390                 395                 400

ggt ctg acg cca acg ctg gcg gcc tgg ctg gtg gaa agc tcg cag gat      1248
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asp
                405                 410                 415

ctg atg atg ccg gcg tat tat ttg atg gtc atc gcg gtg ata ggc ttg      1296
Leu Met Met Pro Ala Tyr Tyr Leu Met Val Ile Ala Val Ile Gly Leu
            420                 425                 430

att acc ggt att tcc atg aaa gag acg gcc aat cgt ccg tta aaa ggc      1344
Ile Thr Gly Ile Ser Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
        435                 440                 445

gca acg cca gcg gcg tcg gac atc cag gaa gcg aag gaa att ctg ggc      1392
Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Gly
    450                 455                 460

gag cat tac gat aat att gag cag aaa atc gac gac atc gat cag gaa      1440
Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp Gln Glu
465                 470                 475                 480

att gcg gag ctg cag gtc aaa cgt tcg cgt ctg gta cag caa cat ccg      1488
Ile Ala Glu Leu Gln Val Lys Arg Ser Arg Leu Val Gln Gln His Pro
                485                 490                 495

cgt atc gat gaa taa                                                 1503
Arg Ile Asp Glu
            500
```

<210> 4
<211> 500
<212> PRT
<213> Salmonella enterica

<400> 4

```
Met Leu Lys Arg Lys Lys Ile Lys Pro Ile Thr Leu Gly Asp Val Thr
1               5                   10                  15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95
```

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Ala Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Leu Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Glu Asn Phe Leu Glu Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

His Trp Arg Ser Leu Leu Ser Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Ile Met Gly Ser Ile Ala Leu Phe
325 330 335

Ala Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe

355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Ile Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asp
405 410 415

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Ile Ala Val Ile Gly Leu
420 425 430

Ile Thr Gly Ile Ser Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Gly
450 455 460

Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp Gln Glu
465 470 475 480

Ile Ala Glu Leu Gln Val Lys Arg Ser Arg Leu Val Gln Gln His Pro
485 490 495

Arg Ile Asp Glu
500

<210> 5
<211> 1503
<212> ADN
<213> Shigella sonnei

<220>
<221> CDS
<222> (1)..(1503)
<223> Vùng mã hóa proP

<400> 5
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

```
atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg      240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc ggg ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag      288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt      336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att      384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa      432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gac tgg ggc tgg cgt att ccg ttc ttt att gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ttg gaa gaa act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtt gat aaa ctg gaa cag ggc gac cgc gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt     1008
```

```
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
                325                 330                 335

gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc      1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt      1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc      1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcg gca ttt aat att tcg gtg ctg gtt gcc      1200
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385                 390                 395                 400

ggt ctg acg cca aca ctg gcg gcc tgg ctg gtc gaa agc tcg cag aat      1248
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
                405                 410                 415

ctg atg atg cct gcc tat tac ctg atg gta gtg gcg gtg att ggt tta      1296
Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Ile Gly Leu
            420                 425                 430

atc acc ggc gta acc atg aaa gag acg gca aat cgt ccg ttg aaa ggt      1344
Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
        435                 440                 445

gcg aca ccg gcg gcg tca gat ata cag gaa gcg aag gaa att ctc gtc      1392
Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
    450                 455                 460

gag cat tac gat aat atc gag cag aaa atc gat gat att gac cac gag      1440
Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465                 470                 475                 480

att gcc gat ttg cag gcg aaa cgt acc cgc ctg gtg cag caa cat ccg      1488
Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
                485                 490                 495

cga att gat gaa taa                                                 1503
Arg Ile Asp Glu
            500
```

<210> 6
<211> 500
<212> PRT
<213> Shigella sonnei

<400> 6

```
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                   10                  15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45
```

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp

305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
405 410 415

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Ile Gly Leu
420 425 430

Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
450 455 460

Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465 470 475 480

Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
485 490 495

Arg Ile Asp Glu
500

<210> 7
<211> 1506
<212> ADN
<213> Erwinia pyrifoliae

<220>
<221> CDS
<222> (1)..(1506)
<223> Vùng mã hóa proP

<400> 7
atg aaa tta cgt agg aag cgt gtt aag cct atc gga tta aag gac gtc 48
Met Lys Leu Arg Arg Lys Arg Val Lys Pro Ile Gly Leu Lys Asp Val
1 5 10 15

```
acc att att gac gat gcc aga tta cgt aag gcg att aca gct gcc tca      96
Thr Ile Ile Asp Asp Ala Arg Leu Arg Lys Ala Ile Thr Ala Ala Ser
            20                  25                  30

ttg ggc aat gcc atg gag tgg ttc gac ttt ggc gtt tat ggt ttt gtt     144
Leu Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val
        35                  40                  45

gcc tac gca ctg ggg caa gtt ttc ttc ccc ggc gcc gat cca ggg acg     192
Ala Tyr Ala Leu Gly Gln Val Phe Phe Pro Gly Ala Asp Pro Gly Thr
    50                  55                  60

cag atg att gcc gcc ctg gca acc ttc tcc gtg ccc ttc ctg atc cgc     240
Gln Met Ile Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg
65                  70                  75                  80

ccg tta ggc ggc ctg ttc ttt ggg tcg ctg ggg gat aaa tac ggc cgc     288
Pro Leu Gly Gly Leu Phe Phe Gly Ser Leu Gly Asp Lys Tyr Gly Arg
                85                  90                  95

cag aaa ata ctg tcg ata acc att att atc atg tcg gtc agt acg ttc     336
Gln Lys Ile Leu Ser Ile Thr Ile Ile Ile Met Ser Val Ser Thr Phe
            100                 105                 110

tgt att ggt tta atc ccg tct tat gcc tcg att ggt atc tgg gca ccg     384
Cys Ile Gly Leu Ile Pro Ser Tyr Ala Ser Ile Gly Ile Trp Ala Pro
        115                 120                 125

atc ctg cta ttg ctg tgt aaa atg gcg cag ggc ttc tcg gtg ggt ggt     432
Ile Leu Leu Leu Leu Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly
    130                 135                 140

gaa tat acc ggg gct tcc atc ttc gtt gct gaa tac tca ccg gat cgc     480
Glu Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg
145                 150                 155                 160

aaa cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tcc atc gcc gga     528
Lys Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly
                165                 170                 175

ttt gtg ctg ggt gcc gga ctg gtg gtg ctg att tca gcg gtt atc ggt     576
Phe Val Leu Gly Ala Gly Leu Val Val Leu Ile Ser Ala Val Ile Gly
            180                 185                 190

gaa gcg agt ttc ctt gaa tgg ggc tgg cgt att ccg ttc ttc gtg gcg     624
Glu Ala Ser Phe Leu Glu Trp Gly Trp Arg Ile Pro Phe Phe Val Ala
        195                 200                 205

cta ccg ctg ggt atc atc ggg ctt tat ctg cgc cac gcg ctt gaa gag     672
Leu Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu
    210                 215                 220

act ccg gcg ttc cag cag cat gtc gac aag ctg gaa aag ggc gat cgg     720
Thr Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Lys Gly Asp Arg
225                 230                 235                 240

gaa gga ttg gcc gat ggg ccg caa gtc tca ttt aaa gag att gcc act     768
Glu Gly Leu Ala Asp Gly Pro Gln Val Ser Phe Lys Glu Ile Ala Thr
                245                 250                 255

aag cac tgg aaa agc ctg ctg gcc tgc atc ggt ctg gtg att gcc acc     816
Lys His Trp Lys Ser Leu Leu Ala Cys Ile Gly Leu Val Ile Ala Thr
            260                 265                 270

aac gtg acc tat tac atg ctg ctg acc tac atg ccg agc tac ctg tcg     864
```

```
Asn Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser
        275                 280                 285

cat aac ctc cat tat tcg gaa gat cat ggc gtg atg atc att atc gcc      912
His Asn Leu His Tyr Ser Glu Asp His Gly Val Met Ile Ile Ile Ala
    290                 295                 300

att atg ttg ggg atg ctg ttt gtg cag ccg gtg atg ggc ctg atg agc      960
Ile Met Leu Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Met Ser
305                 310                 315                 320

gac aaa ttc ggt cgt cgc ccg ttt gtt att atc ggc agt atc gcg ctg     1008
Asp Lys Phe Gly Arg Arg Pro Phe Val Ile Ile Gly Ser Ile Ala Leu
                325                 330                 335

ctg acg ctg tca gta ccg tgc ttt atg ctg atc aac agc ggc gtg atg     1056
Leu Thr Leu Ser Val Pro Cys Phe Met Leu Ile Asn Ser Gly Val Met
            340                 345                 350

ggt ctg att ttt gcc ggg ctg ctg acg ctg gcg gtg atc ctt aac agc     1104
Gly Leu Ile Phe Ala Gly Leu Leu Thr Leu Ala Val Ile Leu Asn Ser
        355                 360                 365

ttc acc ggg gtc atg gca tca agc ctg ccg gcg atg ttc ccc act cat     1152
Phe Thr Gly Val Met Ala Ser Ser Leu Pro Ala Met Phe Pro Thr His
    370                 375                 380

atc cgc tac agt gcg ctg gcc agt gcc ttt aac atc tcg gtg ctg gtt     1200
Ile Arg Tyr Ser Ala Leu Ala Ser Ala Phe Asn Ile Ser Val Leu Val
385                 390                 395                 400

gcc ggc ctg acg ccg acc gct gcc gcc tgg ctg gta gaa acg acc agc     1248
Ala Gly Leu Thr Pro Thr Ala Ala Ala Trp Leu Val Glu Thr Thr Ser
                405                 410                 415

aat ttg tat atg cca gct tat tat ctg atg gtc gtc gcg gtg att ggt     1296
Asn Leu Tyr Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Ile Gly
            420                 425                 430

ctg gta acc ggg att atg atg aag gaa acg gcc aat ctg ccg ctg cgt     1344
Leu Val Thr Gly Ile Met Met Lys Glu Thr Ala Asn Leu Pro Leu Arg
        435                 440                 445

ggc gct gcg cct gcg gct tcg gat atg gct gaa gcg aaa gag atc ctg     1392
Gly Ala Ala Pro Ala Ala Ser Asp Met Ala Glu Ala Lys Glu Ile Leu
    450                 455                 460

cag gaa cat cac gat aat atc gaa cat aag atc gca gat att aac gag     1440
Gln Glu His His Asp Asn Ile Glu His Lys Ile Ala Asp Ile Asn Glu
465                 470                 475                 480

cag atc gct gag ctt gaa gca aaa cgc tcg cat ctt att tat caa cat     1488
Gln Ile Ala Glu Leu Glu Ala Lys Arg Ser His Leu Ile Tyr Gln His
                485                 490                 495

ccg cgt atc aac gag taa                                            1506
Pro Arg Ile Asn Glu
            500
```

<210> 8
<211> 501
<212> PRT
<213> Erwinia pyrifoliae

<400> 8

```
Met Lys Leu Arg Arg Lys Arg Val Lys Pro Ile Gly Leu Lys Asp Val
1               5                  10                  15

Thr Ile Ile Asp Asp Ala Arg Leu Arg Lys Ala Ile Thr Ala Ala Ser
            20                  25                  30

Leu Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val
        35                  40                  45

Ala Tyr Ala Leu Gly Gln Val Phe Phe Pro Gly Ala Asp Pro Gly Thr
    50                  55                  60

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

Pro Leu Gly Gly Leu Phe Phe Gly Ser Leu Gly Asp Lys Tyr Gly Arg
                85                  90                  95

Gln Lys Ile Leu Ser Ile Thr Ile Ile Ile Met Ser Val Ser Thr Phe
            100                 105                 110

Cys Ile Gly Leu Ile Pro Ser Tyr Ala Ser Ile Gly Ile Trp Ala Pro
        115                 120                 125

Ile Leu Leu Leu Leu Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly
    130                 135                 140

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

Lys Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly
                165                 170                 175

Phe Val Leu Gly Ala Gly Leu Val Val Leu Ile Ser Ala Val Ile Gly
            180                 185                 190

Glu Ala Ser Phe Leu Glu Trp Gly Trp Arg Ile Pro Phe Phe Val Ala
        195                 200                 205

Leu Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu
    210                 215                 220

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

Glu Gly Leu Ala Asp Gly Pro Gln Val Ser Phe Lys Glu Ile Ala Thr
                245                 250                 255

Lys His Trp Lys Ser Leu Leu Ala Cys Ile Gly Leu Val Ile Ala Thr
```

260 265 270

Asn Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser
275 280 285

His Asn Leu His Tyr Ser Glu Asp His Gly Val Met Ile Ile Ile Ala
290 295 300

Ile Met Leu Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Met Ser
305 310 315 320

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

Leu Thr Leu Ser Val Pro Cys Phe Met Leu Ile Asn Ser Gly Val Met
340 345 350

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

Phe Thr Gly Val Met Ala Ser Ser Leu Pro Ala Met Phe Pro Thr His
370 375 380

Ile Arg Tyr Ser Ala Leu Ala Ser Ala Phe Asn Ile Ser Val Leu Val
385 390 395 400

Ala Gly Leu Thr Pro Thr Ala Ala Ala Trp Leu Val Glu Thr Thr Ser
405 410 415

Asn Leu Tyr Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Ile Gly
420 425 430

Leu Val Thr Gly Ile Met Met Lys Glu Thr Ala Asn Leu Pro Leu Arg
435 440 445

Gly Ala Ala Pro Ala Ala Ser Asp Met Ala Glu Ala Lys Glu Ile Leu
450 455 460

Gln Glu His His Asp Asn Ile Glu His Lys Ile Ala Asp Ile Asn Glu
465 470 475 480

Gln Ile Ala Glu Leu Glu Ala Lys Arg Ser His Leu Ile Tyr Gln His
485 490 495

Pro Arg Ile Asn Glu
500

<210> 9
<211> 1503
<212> ADN

<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1236)
<223> Vùng mã hóa alen proP M8 thay thế g bằng t ở vị trí 1234 gây ra thay thế AA E412*)

<400> 9

```
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc          48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                  10                  15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg          96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct         144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag         192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg         240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag         288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt         336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att         384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa         432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa         480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt         528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa         576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg         624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act         672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220
```

```
ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt     1008
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
                325                 330                 335

gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc     1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt     1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc     1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcg gca ttt aat att tcg gtg ctg gtt gcc     1200
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385                 390                 395                 400

ggt ctg acg cca acg ctg gcg gcc tgg ctg gtc taa agctcgcaga          1246
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val
                405                 410

atctgatgat gcctgcctat tacctgatgg tagtggcggt ggttggttta atcaccggcg    1306

taaccatgaa agagacggca aatcgtccgt tgaaaggtgc gacaccggcg gcgtcagata    1366

tacaggaagc gaaggaaatt ctcgtcgagc attacgataa tatcgagcag aaaatcgatg    1426

atattgacca cgagattgcc gatttgcagg cgaaacgtac ccgcctggtg cagcaacatc    1486

cgcgaattga tgaataa                                                  1503
```

<210> 10
<211> 411
<212> PRT
<213> Escherichia coli

<400> 10

```
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                  10                  15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270
```

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val
405 410

<210> 11
<211> 1504
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(867)
<223> Vùng mã hóa alen proP M3 (xen đoạn vùng dưới C của vị trí nucleotit 842)

<400> 11
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

```
atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg      240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag      288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt      336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att      384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa      432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tac tat gcc gag tta ttt gtc gca      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Tyr Ala Glu Leu Phe Val Ala
        275                 280                 285

taa cctgcattac tccgaagacc acggggtgct gattattatc gccattatga           917

tcggtatgct gtttgtccag ccggtgatgg gcttgctgag tgaccgtttt ggccgtcgtc    977

cgtttgtgct acttggtagt gttgccctgt ttgtgttggc gatcccggcg tttattctga   1037

ttaacagtaa cgtcatcggc ctgatttttg ccgggttact gatgctggcg gtgatcctta   1097

actgctttac gggcgttatg gcttctacct tgccagcgat gttcccgacg catatccgtt   1157
```

acagcgcgct ggcggcggca tttaatattt cggtgctggt tgccggtctg acgccaacgc 1217

tggcggcctg gctggtcgaa agctcgcaga atctgatgat gcctgcctat tacctgatgg 1277

tagtggcggt ggttggttta atcaccggcg taaccatgaa agagacggca aatcgtccgt 1337

tgaaaggtgc gacaccggcg gcgtcagata tacaggaagc gaaggaaatt ctcgtcgagc 1397

attacgataa tatcgagcag aaaatcgatg atattgacca cgagattgcc gatttgcagg 1457

cgaaacgtac ccgcctggtg cagcaacatc cgcgaattga tgaataa 1504

<210> 12
<211> 288
<212> PRT
<213> Escherichia coli

<400> 12

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

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu

180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Tyr Ala Glu Leu Phe Val Ala
275 280 285

<210> 13
<211> 1502
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(897)
<223> Vùng mã hóa alen proP M4 (khuyết đoạn vùng nằm dưới A của vị trí nucleotit 854)

<400> 13
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg 240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65 70 75 80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag 288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt 336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys

```
                100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att      384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa      432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt ttt tgt cgc ata      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Phe Cys Arg Ile
        275                 280                 285

acc tgc att act ccg aag acc acg ggg tgc tga ttattatcgc cattatgatc      917
Thr Cys Ile Thr Pro Lys Thr Thr Gly Cys
    290                 295

ggtatgctgt ttgtccagcc ggtgatgggc ttgctgagtg accgttttgg ccgtcgtccg      977

tttgtgctac ttggtagtgt tgccctgttt gtgttggcga tcccggcgtt tattctgatt     1037

aacagtaacg tcatcggcct gatttttgcc gggttactga tgctggcggt gatccttaac     1097

tgctttacgg gcgttatggc ttctaccttg ccagcgatgt tcccgacgca tatccgttac     1157

agcgcgctgg cggcggcatt taatatttcg gtgctggttg ccggtctgac gccaacgctg     1217

gcggcctggc tggtcgaaag ctcgcagaat ctgatgatgc ctgcctatta cctgatggta     1277

gtggcggtgg ttggtttaat caccggcgta accatgaaag agacggcaaa tcgtccgttg     1337

aaaggtgcga caccggcggc gtcagatata caggaagcga aggaaattct cgtcgagcat     1397
```

tacgataata tcgagcagaa aatcgatgat attgaccacg agattgccga tttgcaggcg 1457

aaacgtaccc gcctggtgca gcaacatccg cgaattgatg aataa 1502

<210> 14
<211> 298
<212> PRT
<213> Escherichia coli

<400> 14

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

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Phe Cys Arg Ile
275 280 285

Thr Cys Ile Thr Pro Lys Thr Thr Gly Cys
290 295

<210> 15
<211> 1503
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1503)
<223> Vùng mã hóa alen proP M11 (thay thế t bằng c ở vị trí 1399 gây ra thay thế AA Y467H)

<400> 15
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg 240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65 70 75 80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag 288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt 336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att 384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

```
ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa      432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt     1008
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
                325                 330                 335

gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc     1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt     1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc     1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcg gca ttt aat att tcg gtg ctg gtt gcc     1200
```

```
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385                 390                 395                 400

ggt ctg acg cca acg ctg gcg gcc tgg ctg gtc gaa agc tcg cag aat      1248
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
                405                 410                 415

ctg atg atg cct gcc tat tac ctg atg gta gtg gcg gtg gtt ggt tta      1296
Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
            420                 425                 430

atc acc ggc gta acc atg aaa gag acg gca aat cgt ccg ttg aaa ggt      1344
Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
        435                 440                 445

gcg aca ccg gcg gcg tca gat ata cag gaa gcg aag gaa att ctc gtc      1392
Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
    450                 455                 460

gag cat cac gat aat atc gag cag aaa atc gat gat att gac cac gag      1440
Glu His His Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465                 470                 475                 480

att gcc gat ttg cag gcg aaa cgt acc cgc ctg gtg cag caa cat ccg      1488
Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
                485                 490                 495

cga att gat gaa taa                                                 1503
Arg Ile Asp Glu
            500
```

<210> 16
<211> 500
<212> PRT
<213> Escherichia coli

<400> 16

```
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                   10                  15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110
```

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile

370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
405 410 415

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
420 425 430

Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
450 455 460

Glu His His Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465 470 475 480

Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
485 490 495

Arg Ile Asp Glu
500

<210> 17
<211> 1452
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1452)
<223> Vùng mã hóa alen proP M7 (khuyết đoạn 51bp từ vị trí nucleotit 1173 đến 1223

<400> 17
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg 240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro

65 70 75 80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag 288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt 336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att 384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa 432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa 480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt 528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa 576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg 624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act 672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa 720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225 230 235 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa 768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac 816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat 864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att 912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac 960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt 1008
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

```
gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc      1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt      1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc      1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcc tgg ctg gtc gaa agc tcg cag aat ctg      1200
Arg Tyr Ser Ala Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn Leu
385                 390                 395                 400

atg atg cct gcc tat tac ctg atg gta gtg gcg gtg gtt ggt tta atc      1248
Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu Ile
                405                 410                 415

acc ggc gta acc atg aaa gag acg gca aat cgt ccg ttg aaa ggt gcg      1296
Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly Ala
            420                 425                 430

aca ccg gcg gcg tca gat ata cag gaa gcg aag gaa att ctc gtc gag      1344
Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val Glu
        435                 440                 445

cat tac gat aat atc gag cag aaa atc gat gat att gac cac gag att      1392
His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu Ile
    450                 455                 460

gcc gat ttg cag gcg aaa cgt acc cgc ctg gtg cag caa cat ccg cga      1440
Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro Arg
465                 470                 475                 480

att gat gaa taa                                                     1452
Ile Asp Glu
```

<210> 18
<211> 483
<212> PRT
<213> Escherichia coli

<400> 18

```
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                   10                  15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80
```

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn Leu
385 390 395 400

Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu Ile
405 410 415

Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly Ala
420 425 430

Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val Glu
435 440 445

His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu Ile
450 455 460

Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro Arg
465 470 475 480

Ile Asp Glu

<210> 19
<211> 1522
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1014)
<223> Vùng mã hóa alen proP M6 (xen đoạn 19bp nằm dưới vị trí nucleotit 973)

<400> 19
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc 48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1 5 10 15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg 96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct 144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag 192

```
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg      240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag      288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt      336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att      384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa      432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa      480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt      528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa      576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg      624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act      672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa      720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa      768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255

tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320
```

cgt ttt ggc cgt cgt ccg ttt gtg cta ctt gcg tcc gtt tgt gct act 1008
Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Ala Ser Val Cys Ala Thr
325 330 335

tgg tag tgttgccctg tttgtgttgg cgatcccggc gtttattctg attaacagta 1064
Trp

acgtcatcgg cctgattttt gccgggttac tgatgctggc ggtgatcctt aactgcttta 1124

cgggcgttat ggcttctacc ttgccagcga tgttcccgac gcatatccgt tacagcgcgc 1184

tggcggcggc atttaatatt tcggtgctgg ttgccggtct gacgccaacg ctggcggcct 1244

ggctggtcga aagctcgcag aatctgatga tgcctgccta ttacctgatg gtagtggcgg 1304

tggttggttt aatcaccggc gtaaccatga aagagacggc aaatcgtccg ttgaaaggtg 1364

cgacaccggc ggcgtcagat atacaggaag cgaaggaaat tctcgtcgag cattacgata 1424

atatcgagca gaaaatcgat gatattgacc acgagattgc cgatttgcag gcgaaacgta 1484

cccgcctggt gcagcaacat ccgcgaattg atgaataa 1522

<210> 20
<211> 337
<212> PRT
<213> Escherichia coli

<400> 20

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

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu

130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Arg Arg Pro Phe Val Leu Leu Ala Ser Val Cys Ala Thr
325 330 335

Trp

<210> 21
<211> 1503
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1503)
<223> Vùng mã hóa alen proP M5 (thay thế g bằng t ở vị trí 971 gây ra thay thế AA R324L)

<400> 21

```
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc      48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                  10                  15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg      96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct     144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc agc gtg cag     192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
    50                  55                  60

atg gtt gct gca ctt gcc act ttc tcc gtt ccc ttt ctg att cga ccg     240
Met Val Ala Ala Leu Ala Thr Phe Ser Val Pro Phe Leu Ile Arg Pro
65                  70                  75                  80

ctt ggc gga ctc ttc ttt ggt atg ttg ggc gat aaa tat ggt cgc cag     288
Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
                85                  90                  95

aag atc ctc gct atc act att gtg att atg tcg atc agt acg ttc tgt     336
Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
            100                 105                 110

att ggc tta ata ccg tcc tac gac acg att ggt att tgg gca ccg att     384
Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
        115                 120                 125

ctg ctg ttg atc tgt aag atg gca caa ggt ttc tcg gtc ggc ggt gaa     432
Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
    130                 135                 140

tat acc ggg gcg tcg ata ttt gtt gcg gaa tac tcc cct gac cgt aaa     480
Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145                 150                 155                 160

cgt ggc ttt atg ggc agc tgg ctg gac ttc ggt tct att gcc ggg ttt     528
Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
                165                 170                 175

gtg ctg ggt gcg ggc gtg gtg gtg tta att tcg acc att gtc ggc gaa     576
Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
            180                 185                 190

gcg aac ttc ctc gat tgg ggc tgg cgt att ccg ttc ttt atc gct ctg     624
Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
        195                 200                 205

ccg tta ggg att atc ggg ctt tac ctg cgc cat gcg ctg gaa gag act     672
Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
    210                 215                 220

ccg gcg ttc cag cag cat gtc gat aaa ctg gaa cag ggc gac cgt gaa     720
Pro Ala Phe Gln Gln His Val Asp Lys Leu Glu Gln Gly Asp Arg Glu
225                 230                 235                 240

ggt ttg cag gat ggc ccg aaa gtc tcg ttt aaa gag att gcc act aaa     768
Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
                245                 250                 255
```

```
tac tgg cgc agc ctg ttg aca tgt att ggt ctg gta att gcc acc aac      816
Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
            260                 265                 270

gtg act tac tac atg ttg ctg acc tat atg ccg agt tat ttg tcg cat      864
Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
        275                 280                 285

aac ctg cat tac tcc gaa gac cac ggg gtg ctg att att atc gcc att      912
Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
    290                 295                 300

atg atc ggt atg ctg ttt gtc cag ccg gtg atg ggc ttg ctg agt gac      960
Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305                 310                 315                 320

cgt ttt ggc ctt cgt ccg ttt gtg cta ctt ggt agt gtt gcc ctg ttt     1008
Arg Phe Gly Leu Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
                325                 330                 335

gtg ttg gcg atc ccg gcg ttt att ctg att aac agt aac gtc atc ggc     1056
Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
            340                 345                 350

ctg att ttt gcc ggg tta ctg atg ctg gcg gtg atc ctt aac tgc ttt     1104
Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
        355                 360                 365

acg ggc gtt atg gct tct acc ttg cca gcg atg ttc ccg acg cat atc     1152
Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
    370                 375                 380

cgt tac agc gcg ctg gcg gcg gca ttt aat att tcg gtg ctg gtt gcc     1200
Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385                 390                 395                 400

ggt ctg acg cca acg ctg gcg gcc tgg ctg gtc gaa agc tcg cag aat     1248
Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
                405                 410                 415

ctg atg atg cct gcc tat tac ctg atg gta gtg gcg gtg gtt ggt tta     1296
Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
            420                 425                 430

atc acc ggc gta acc atg aaa gag acg gca aat cgt ccg ttg aaa ggt     1344
Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
        435                 440                 445

gcg aca ccg gcg gcg tca gat ata cag gaa gcg aag gaa att ctc gtc     1392
Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
    450                 455                 460

gag cat tac gat aat atc gag cag aaa atc gat gat att gac cac gag     1440
Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465                 470                 475                 480

att gcc gat ttg cag gcg aaa cgt acc cgc ctg gtg cag caa cat ccg     1488
Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
                485                 490                 495

cga att gat gaa taa                                                1503
Arg Ile Asp Glu
            500
```

<210> 22

<211> 500
<212> PRT
<213> Escherichia coli

<400> 22

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

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ser Val Gln
50 55 60

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

Leu Gly Gly Leu Phe Phe Gly Met Leu Gly Asp Lys Tyr Gly Arg Gln
85 90 95

Lys Ile Leu Ala Ile Thr Ile Val Ile Met Ser Ile Ser Thr Phe Cys
100 105 110

Ile Gly Leu Ile Pro Ser Tyr Asp Thr Ile Gly Ile Trp Ala Pro Ile
115 120 125

Leu Leu Leu Ile Cys Lys Met Ala Gln Gly Phe Ser Val Gly Gly Glu
130 135 140

Tyr Thr Gly Ala Ser Ile Phe Val Ala Glu Tyr Ser Pro Asp Arg Lys
145 150 155 160

Arg Gly Phe Met Gly Ser Trp Leu Asp Phe Gly Ser Ile Ala Gly Phe
165 170 175

Val Leu Gly Ala Gly Val Val Val Leu Ile Ser Thr Ile Val Gly Glu
180 185 190

Ala Asn Phe Leu Asp Trp Gly Trp Arg Ile Pro Phe Phe Ile Ala Leu
195 200 205

Pro Leu Gly Ile Ile Gly Leu Tyr Leu Arg His Ala Leu Glu Glu Thr
210 215 220

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

Gly Leu Gln Asp Gly Pro Lys Val Ser Phe Lys Glu Ile Ala Thr Lys
245 250 255

Tyr Trp Arg Ser Leu Leu Thr Cys Ile Gly Leu Val Ile Ala Thr Asn
260 265 270

Val Thr Tyr Tyr Met Leu Leu Thr Tyr Met Pro Ser Tyr Leu Ser His
275 280 285

Asn Leu His Tyr Ser Glu Asp His Gly Val Leu Ile Ile Ile Ala Ile
290 295 300

Met Ile Gly Met Leu Phe Val Gln Pro Val Met Gly Leu Leu Ser Asp
305 310 315 320

Arg Phe Gly Leu Arg Pro Phe Val Leu Leu Gly Ser Val Ala Leu Phe
325 330 335

Val Leu Ala Ile Pro Ala Phe Ile Leu Ile Asn Ser Asn Val Ile Gly
340 345 350

Leu Ile Phe Ala Gly Leu Leu Met Leu Ala Val Ile Leu Asn Cys Phe
355 360 365

Thr Gly Val Met Ala Ser Thr Leu Pro Ala Met Phe Pro Thr His Ile
370 375 380

Arg Tyr Ser Ala Leu Ala Ala Ala Phe Asn Ile Ser Val Leu Val Ala
385 390 395 400

Gly Leu Thr Pro Thr Leu Ala Ala Trp Leu Val Glu Ser Ser Gln Asn
405 410 415

Leu Met Met Pro Ala Tyr Tyr Leu Met Val Val Ala Val Val Gly Leu
420 425 430

Ile Thr Gly Val Thr Met Lys Glu Thr Ala Asn Arg Pro Leu Lys Gly
435 440 445

Ala Thr Pro Ala Ala Ser Asp Ile Gln Glu Ala Lys Glu Ile Leu Val
450 455 460

Glu His Tyr Asp Asn Ile Glu Gln Lys Ile Asp Asp Ile Asp His Glu
465 470 475 480

Ile Ala Asp Leu Gln Ala Lys Arg Thr Arg Leu Val Gln Gln His Pro
485 490 495

Arg Ile Asp Glu
500

<210> 23
<211> 2852
<212> ADN
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(243)
<223> Vùng mã hóa alen proP M2 (xen đoạn 1359bp nằm dưới vị trí nucleotit 183)

<400> 23

```
atg ctg aaa agg aaa aaa gta aaa ccg att acc ctt cgt gat gtc acc      48
Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr
1               5                  10                  15

att att gat gac ggt aaa ctg cgt aaa gcc att acc gca gca tca ctg      96
Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
            20                  25                  30

ggt aat gca atg gaa tgg ttc gat ttt ggt gtt tat ggt ttt gtt gct     144
Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
        35                  40                  45

tac gca tta ggt aaa gtt ttt ttc ccg ggg gct gac ccc ata agc gct     192
Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ile Ser Ala
    50                  55                  60

aac tta agg gtt gaa cca tct gaa gaa tgc gac gcc tcg gtg cct cgt     240
Asn Leu Arg Val Glu Pro Ser Glu Glu Cys Asp Ala Ser Val Pro Arg
65                  70                  75                  80

taa gacgatgcct cgcgttcttc aattgcgttt tgtaggctgt cagggatact          293

gtcccacgaa tggccacctg taagctccag atgaccattt ttgttattct ccacaacgag  353

ttagttcttc ttttcggatc cggcacttct ggggggggaaa tccagcgatg gctggattat  413

gtcgtcaatt aaaaatgcgg cgagtagatt agcaaatatc cacgctttcg cgagttcagg  473

ttcctttgca cgcaaagcat ccaggtgcag caaacttttg agccgcttaa aagccagttc  533

aatttgccat cgcagacggt aacaatcagc cacttgctct gctgaatatt catcttccgg  593

taatgatgtt agcaatagca catggcccgc tgcttccagc gtttccgcct gaactactcg  653

tccttttcga cgattctcgc tgagcagtcg ggttttactg attaatgctt tttcgggagg  713

aagtgatacg gcaatgagac gtgccggaaa gggagctccg gctttttttat tacctgaatt  773

gcctatcatt acagtggttt caccgttctt accgcaatcc agcccgcgca gaaaacccat  833

catgtcaaag cgcattcctt ctgcagttaa ccagcgcaat cctcgccagt gaacccggac  893

gatataatca gcttctccaa aagcaagtga gcggatacat tcgggacgcg aaccgaatcc  953

ccggtcagca atgcgtatct cgtctgccgt ttgcgcaaat cggtccagcc gttcagcgtc  1013

tctgctgtcg gttagctcaa aatcagtgaa ctgacaggta tgaggatcat atcccatatg  1073

tagtcgccat tcagcgctgc cgcccccggg cgcactgatt gctgttccat cgacaagacg  1133

caatctcttt ccgcttgtac aacccgtaac tgcggcgcgt acagcaagtg tttgtgcggc  1193
```

aagtatgcca aaccagtcgg cggcattccg cagccgcttc aggagagcca cgtcagataa 1253

tgttgcaacg tcatggagct gagcccatgc agtgacttca cgtaatgaca tccccccggg 1313

gccgtaagcc agccccagac gtagcagagt tgcagcatca cgaatttcgc ggcggcgggt 1373

tagagccccg gcattacgtg ccgaagtatc cagttcttcg ggcttaccaa tatgggccag 1433

aattgctgac cagttatcgt gagagtaatt catcggcacg ttaaatcata tcaggcgtaa 1493

taccacaacc cttaagttag cgcttatggg gctgacccca gcgtgcagat ggttgctgca 1553

cttgccactt tctccgttcc ctttctgatt cgaccgcttg gcggactctt ctttggtatg 1613

ttgggcgata aatatggtcg ccagaagatc ctcgctatca ctattgtgat tatgtcgatc 1673

agtacgttct gtattggctt aataccgtcc tacgacacga ttggtatttg ggcaccgatt 1733

ctgctgttga tctgtaagat ggcacaaggt ttctcggtcg gcggtgaata taccggggcg 1793

tcgatatttg ttgcggaata ctcccctgac cgtaaacgtg gctttatggg cagctggctg 1853

gacttcggtt ctattgccgg gtttgtgctg ggtgcgggcg tggtggtgtt aatttcgacc 1913

attgtcggcg aagcgaactt cctcgattgg ggctggcgta ttccgttctt tatcgctctg 1973

ccgttaggga ttatcgggct ttacctgcgc catgcgctgg aagagactcc ggcgttccag 2033

cagcatgtcg ataaactgga acagggcgac cgtgaaggtt tgcaggatgg cccgaaagtc 2093

tcgtttaaag agattgccac taaatactgg cgcagcctgt tgacatgtat tggtctggta 2153

attgccacca acgtgactta ctacatgttg ctgacctata tgccgagtta tttgtcgcat 2213

aacctgcatt actccgaaga ccacggggtg ctgattatta tcgccattat gatcggtatg 2273

ctgtttgtcc agccggtgat gggcttgctg agtgaccgtt ttggccgtcg tccgtttgtg 2333

ctacttggta gtgttgccct gtttgtgttg gcgatcccgg cgtttattct gattaacagt 2393

aacgtcatcg gcctgatttt tgccgggtta ctgatgctgg cggtgatcct taactgcttt 2453

acgggcgtta tggcttctac cttgccagcg atgttcccga cgcatatccg ttacagcgcg 2513

ctggcggcgg catttaatat ttcggtgctg gttgccggtc tgacgccaac gctggcggcc 2573

tggctggtcg aaagctcgca gaatctgatg atgcctgcct attacctgat ggtagtggcg 2633

gtggttggtt taatcaccgg cgtaaccatg aaagagacgg caaatcgtcc gttgaaaggt 2693

gcgacaccgg cggcgtcaga tatacaggaa gcgaaggaaa ttctcgtcga gcattacgat 2753

aatatcgagc agaaaatcga tgatattgac cacgagattg ccgatttgca ggcgaaacgt 2813

acccgcctgg tgcagcaaca tccgcgaatt gatgaataa 2852

<210> 24
<211> 80
<212> PRT
<213> Escherichia coli

<400> 24

Met Leu Lys Arg Lys Lys Val Lys Pro Ile Thr Leu Arg Asp Val Thr

1 5 10 15

Ile Ile Asp Asp Gly Lys Leu Arg Lys Ala Ile Thr Ala Ala Ser Leu
20 25 30

Gly Asn Ala Met Glu Trp Phe Asp Phe Gly Val Tyr Gly Phe Val Ala
35 40 45

Tyr Ala Leu Gly Lys Val Phe Phe Pro Gly Ala Asp Pro Ile Ser Ala
50 55 60

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

<210> 25
<211> 38
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(38)
<223> Đoạn mồi serCF(XbaI)

<400> 25
aggtgctcta gagtccgcgc tgtgcaaatc cagaatgg 38

<210> 26
<211> 37
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(37)
<223> Đoạn mồi serCR(HindIII)

<400> 26
tacaccaagc ttaactctct acaacagaaa taaaaac 37

<210> 27
<211> 33
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(33)
<223> Đoạn mồi serAF(XbaI)

<400> 27
ctgtagtcta gattagtaca gcagacgggc gcg 33

<210> 28
<211> 68

<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(68)
<223> Đoạn mồi serAR(SHSSNB)

<400> 28
caagagctca agcttgcatg cgattcccgg gcggccgcaa taagatctcc gtcagggcgt 60

ggtgaccg 68

<210> 29
<211> 34
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(34)
<223> Đoạn mồi serB(SphI)

<400> 29
ccatgcgcat gcccacccctt tgaaaatttg agac 34

<210> 30
<211> 75
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(75)
<223> Đoạn mồi serB(SmaI)

<400> 30
ccgcatgtcg acatcccggg gcagaaaggc ccacccgaag gtgagccagt gtgattactt 60

ctgattcagg ctgcc 75

<210> 31
<211> 35
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(35)
<223> Đoạn mồi nằm dưới glyA

<400> 31
atctaaagat ctgttacgac agatttgatg gcgcg 35

<210> 32
<211> 33
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(33)
<223> nằm trên glyA

<400> 32
ttcatcgcgg ccgcgaaaga atgtgatgaa gtg 33

<210> 33
<211> 30
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(30)
<223> Đoạn mồi proPmut1(HindIII)

<400> 33
gtcaaagctt atatggtcgc cagaagatcc 30

<210> 34
<211> 30
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (1)..(30)
<223> Đoạn mồi proPmut2(BamHI)

<400> 34
gtcaggatcc tcagccgcat tacacagttg 30

<210> 35
<211> 1564
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (5)..(10)
<223> Trình tự nhận biết HindIII

<220>
<221> đột biến
<222> (969)..(969)
<223> M8 đột biến: thay thế a bằng t ở vị trí 1234, gây ra E412*

<220>
<221> dấu hiệu misc
<222> (1534)..(1539)
<223> Trình tự nhận biết BamHI

<400> 35
gtcaaagctt atatggtcgc cagaagatcc tcgctatcac tattgtgatt atgtcgatca 60

gtacgttctg tattggctta ataccgtcct acgacacgat tggtatttgg gcaccgattc 120

tgctgttgat ctgtaagatg gcacaaggtt tctcggtcgg cggtgaatat accggggcgt 180

cgatatttgt tgcggaatac tcccctgacc gtaaacgtgg ctttatgggc agctggctgg 240

acttcggttc tattgccggg tttgtgctgg gtgcgggcgt ggtggtgtta atttcgacca 300

ttgtcggcga agcgaacttc ctcgattggg gctggcgtat tccgttcttt atcgctctgc 360

cgttagggat tatcgggctt tacctgcgcc atgcgctgga agagactccg gcgttccagc 420

agcatgtcga taaactggaa cagggcgacc gtgaaggttt gcaggatggc ccgaaagtct 480

cgtttaaaga gattgccact aaatactggc gcagcctgtt gacatgtatt ggtctggtaa 540

ttgccaccaa cgtgacttac tacatgttgc tgacctatat gccgagttat ttgtcgcata 600

acctgcatta ctccgaagac cacggggtgc tgattattat cgccattatg atcggtatgc 660

tgtttgtcca gccggtgatg ggcttgctga gtgaccgttt tggccgtcgt ccgtttgtgc 720

tacttgcgtc cgtttgtgct acttggtagt gttgccctgt ttgtgttggc gatcccggcg 780

tttattctga ttaacagtaa cgtcatcggc ctgatttttg ccgggttact gatgctggcg 840

gtgatcctta actgctttac gggcgttatg gcttctacct tgccagcgat gttcccgacg 900

catatccgtt acagcgcgct ggcggcggca tttaatattt cggtgctggt tgccggtctg 960

acgccaacgc tggcggcctg gctggtcgaa agctcgcaga atctgatgat gcctgcctat 1020

tacctgatgg tagtggcggt ggttggttta atcaccggcg taaccatgaa agagacggca 1080

aatcgtccgt tgaaaggtgc gacaccggcg gcgtcagata tacaggaagc gaaggaaatt 1140

ctcgtcgagc attacgataa tatcgagcag aaaatcgatg atattgacca cgagattgcc 1200

gatttgcagg cgaaacgtac ccgcctggtg cagcaacatc cgcgaattga tgaataagct 1260

gaaacggatg gcctgatgtg acgctgtctt atcaggccaa ttgaactctt aaggttcact 1320

taatctctga cgcgcatact ctcctccagg ttaacggagg agagtgcaat gaaaaaccgt 1380

gtttatgaaa gtttaactac cgtgttcagc gtgctggtgg tcagcagctt tctttatatc 1440

tggtttgcca cgtactgatc tttcttcagc cgtacccagg cccgcgtgcc ggaagtctct 1500

tgccggtttt gcaggaaaaa ctgcccgtga tgcaactgtg taatgcggct gagcggccgc 1560

tcag 1564

<210> 36
<211> 1542
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (5)..(10)
<223> Trình tự nhận biết HindIII

<220>
<221> đột biến
<222> (588)..(589)

<223> đột biến M4: khuyết đoạn của vùng nằm dưới A của vị trí nucleotit 854

<220>
<221> dấu hiệu misc
<222> (1533)..(538)
<223> Trình tự nhận biết BamHI

<400> 36
gtcaaagctt atatggtcgc cagaagatcc tcgctatcac tattgtgatt atgtcgatca 60

gtacgttctg tattggctta ataccgtcct acgacacgat tggtatttgg gcaccgattc 120

tgctgttgat ctgtaagatg gcacaaggtt tctcggtcgg cggtgaatat accggggcgt 180

cgatatttgt tgcggaatac tcccctgacc gtaaacgtgg ctttatgggc agctggctgg 240

acttcggttc tattgccggg tttgtgctgg gtgcgggcgt ggtggtgtta atttcgacca 300

ttgtcggcga agcgaacttc ctcgattggg gctggcgtat tccgttcttt atcgctctgc 360

cgttagggat tatcgggctt tacctgcgcc atgcgctgga agagactccg gcgttccagc 420

agcatgtcga taaactggaa cagggcgacc gtgaaggttt gcaggatggc ccgaaagtct 480

cgtttaaaga gattgccact aaatactggc gcagcctgtt gacatgtatt ggtctggtaa 540

ttgccaccaa cgtgacttac tacatgttgc tgacctatat gccgagtttt tgtcgcataa 600

cctgcattac tccgaagacc acggggtgct gattattatc gccattatga tcggtatgct 660

gtttgtccag ccggtgatgg gcttgctgag tgaccgtttt ggccgtcgtc cgtttgtgct 720

acttggtagt gttgccctgt ttgtgttggc gatcccggcg tttattctga ttaacagtaa 780

cgtcatcggc ctgatttttg ccgggttact gatgctggcg gtgatcctta actgctttac 840

gggcgttatg gcttctacct tgccagcgat gttcccgacg catatccgtt acagcgcgct 900

ggcggcggca tttaatattt cggtgctggt tgccggtctg acgccaacgc tggcggcctg 960

gctggtcgaa agctcgcaga atctgatgat gcctgcctat tacctgatgg tagtggcggt 1020

ggttggttta atcaccggcg taaccatgaa agagacggca aatcgtccgt tgaaaggtgc 1080

gacaccggcg gcgtcagata tacaggaagc gaaggaaatt ctcgtcgagc attacgataa 1140

tatcgagcag aaaatcgatg atattgacca cgagattgcc gatttgcagg cgaaacgtac 1200

ccgcctggtg cagcaacatc cgcgaattga tgaataagct gaaacggatg gcctgatgtg 1260

acgctgtctt atcaggccaa ttgaactctt aaggttcact taatctctga cgcgcatact 1320

ctcctccagg ttaacggagg agagtgcaat gaaaaaccgt gtttatgaaa gtttaactac 1380

cgtgttcagc gtgctggtgg tcagcagctt tctttatatc tggtttgcca cgtactgatc 1440

tttcttcagc cgtacccagg cccgcgtgcc ggaagtctct tgccggtttt gcaggaaaaa 1500

ctgcccgtga tgcaactgtg taatgcggct gaggatcctg ac 1542

<210> 37
<211> 1563
<212> ADN
<213> Escherichia coli

<220>
<221> dấu hiệu misc
<222> (5)..(10)
<223> Trình tự nhận biết HindIII

<220>
<221> đột biến
<222> (722)..(740)
<223> Đột biến M6 : xen đoạn 19bp nằm dưới vị trí nucleotit 973

<220>
<221> dấu hiệu misc
<222> (1553)..(1558)
<223> Trình tự nhận biết BamHI

<400> 37
gtcaaagctt atatggtcgc cagaagatcc tcgctatcac tattgtgatt atgtcgatca 60

gtacgttctg tattggctta ataccgtcct acgacacgat tggtatttgg gcaccgattc 120

tgctgttgat ctgtaagatg gcacaaggtt tctcggtcgg cggtgaatat accggggcgt 180

cgatatttgt tgcggaatac tcccctgacc gtaaacgtgg ctttatgggc agctggctgg 240

acttcggttc tattgccggg tttgtgctgg gtgcgggcgt ggtggtgtta atttcgacca 300

ttgtcggcga agcgaacttc ctcgattggg gctggcgtat tccgttcttt atcgctctgc 360

cgttagggat tatcgggctt tacctgcgcc atgcgctgga agagactccg gcgttccagc 420

agcatgtcga taaactggaa cagggcgacc gtgaaggttt gcaggatggc ccgaaagtct 480

cgtttaaaga gattgccact aaatactggc gcagcctgtt gacatgtatt ggtctggtaa 540

ttgccaccaa cgtgacttac tacatgttgc tgacctatat gccgagttat ttgtcgcata 600

acctgcatta ctccgaagac cacggggtgc tgattattat cgccattatg atcggtatgc 660

tgtttgtcca gccggtgatg ggcttgctga gtgaccgttt tggccgtcgt ccgtttgtgc 720

tacttgcgtc cgtttgtgct acttggtagt gttgccctgt ttgtgttggc gatcccggcg 780

tttattctga ttaacagtaa cgtcatcggc ctgatttttg ccgggttact gatgctggcg 840

gtgatcctta actgctttac gggcgttatg gcttctacct tgccagcgat gttcccgacg 900

catatccgtt acagcgcgct ggcggcggca tttaatattt cggtgctggt tgccggtctg 960

acgccaacgc tggcggcctg gctggtcgaa agctcgcaga atctgatgat gcctgcctat 1020

tacctgatgg tagtggcggt ggttggttta atcaccggcg taaccatgaa agagacggca 1080

aatcgtccgt tgaaaggtgc gacaccggcg gcgtcagata tacaggaagc gaaggaaatt 1140

ctcgtcgagc attacgataa tatcgagcag aaaatcgatg atattgacca cgagattgcc 1200

gatttgcagg cgaaacgtac ccgcctggtg cagcaacatc cgcgaattga tgaataagct 1260

gaaacggatg gcctgatgtg acgctgtctt atcaggccaa ttgaactctt aaggttcact 1320

taatctctga cgcgcatact ctcctccagg ttaacggagg agagtgcaat gaaaaaccgt 1380

gtttatgaaa gtttaactac cgtgttcagc gtgctggtgg tcagcagctt tctttatatc 1440

tggtttgcca cgtactgatc tttcttcagc cgtacccagg cccgcgtgcc ggaagtctct 1500

tgccggtttt gcaggaaaaa ctgcccgtga tgcaactgtg taatgcggct gagggatcct 1560

cag 1563

<210> 38
<211> 3503
<212> ADN
<213> Escherichia coli

<220>
<221> gene
<222> (1001)..(2503)
<223> Vùng mã hóa proP

<400> 38
ttctttcgcg tgtgcagctt cagacatgga cgaacggcgg catgttaaat gcgccgctta 60

gcctgcgtct gacattggtg gaaaaactgg cgtcgatgct ggatcccggt catctggcac 120

tgacgcagat cgcgcagcat ctggcgctgc tgcaaaaaat ggatcaccgc cagcactctg 180

ctttcccgga gctcccccag caaattgccg ccttgtatga gtggttttca gcccgttgtc 240

gctggaagga aaaggcgtta acgcaacgag gcctactggt gcaggcaggt gatcagagcg 300

agcaaatttt tacccgctgg cgtgctgggg cgtataacgc ctggtcgttg cctgggcgct 360

gttttatcgt tctggaggag ttgcgctggg gggcatttgg cgatgcctgc cgtctgggaa 420

gcccgcaagc ggtggcgttg ttgctgggtg atttgctcga gaaagcgaca caacatctgg 480

cagagagtat caatgcggca ccgaccacgc gtcactatta ccatcagtgg tttgcctctt 540

cgaccgttcc gacgggcggg gagcatgctg atttttaag ttggctggga aagtggacca 600

cggcagataa acaacccgtt tgctggtcag tgacccaacg ctggcaaact gtcgcgctgg 660

ggatgccacg actctgttca gcgcagcgtc tggcgggggc aatgctcgag gaaatcttct 720

ctgtaaattt ggcgtaaata atcagttaca tcaatgagtc ctaaacgaaa tccatgtgtg 780

aagttgatca caaatttaaa cactggtagg gtaaaaaggt cattaactgc ccaattcagg 840

cgtcaactgg tttgattgta cattccttaa ccggagggtg taagcaaacc cgctacgctt 900

gttacagaga ttgcatcctg caattcccgc tccccttttg cggccgtcgc gctgattttt 960

ctggcgtttg cggaaatggg ccaactctgc gaggaaagct atgctgaaaa ggaaaaaagt 1020

aaaaccgatt acccttcgtg atgtcaccat tattgatgac ggtaaactgc gtaaagccat 1080

taccgcagca tcactgggta atgcaatgga atggttcgat tttggtgttt atggttttgt 1140

tgcttacgca ttaggtaaag ttttttttccc gggggctgac cccagcgtgc agatggttgc 1200

tgcacttgcc actttctccg ttccctttct gattcgaccg cttggcggac tcttctttgg 1260

tatgttgggc gataaatatg gtcgccagaa gatcctcgct atcactattg tgattatgtc 1320

gatcagtacg ttctgtattg gcttaatacc gtcctacgac acgattggta tttgggcacc 1380

gattctgctg ttgatctgta agatggcaca aggtttctcg gtcggcggtg aatataccgg 1440

ggcgtcgata tttgttgcgg aatactcccc tgaccgtaaa cgtggcttta tgggcagctg 1500

gctggacttc ggttctattg ccgggtttgt gctgggtgcg ggcgtggtgg tgttaatttc 1560

gaccattgtc ggcgaagcga acttcctcga ttggggctgg cgtattccgt tctttatcgc 1620

tctgccgtta gggattatcg ggctttacct gcgccatgcg ctggaagaga ctccggcgtt 1680

ccagcagcat gtcgataaac tggaacaggg cgaccgtgaa ggtttgcagg atggcccgaa 1740

agtctcgttt aaagagattg ccactaaata ctggcgcagc ctgttgacat gtattggtct 1800

ggtaattgcc accaacgtga cttactacat gttgctgacc tatatgccga gttatttgtc 1860

gcataacctg cattactccg aagaccacgg ggtgctgatt attatcgcca ttatgatcgg 1920

tatgctgttt gtccagccgg tgatgggctt gctgagtgac cgttttggcc gtcgtccgtt 1980

tgtgctactt ggtagtgttg ccctgtttgt gttggcgatc ccggcgttta ttctgattaa 2040

cagtaacgtc atcggcctga tttttgccgg gttactgatg ctggcggtga tccttaactg 2100

ctttacgggc gttatggctt ctaccttgcc agcgatgttc ccgacgcata tccgttacag 2160

cgcgctggcg gcggcattta atatttcggt gctggttgcc ggtctgacgc caacgctggc 2220

ggcctggctg gtcgaaagct cgcagaatct gatgatgcct gcctattacc tgatggtagt 2280

ggcggtggtt ggtttaatca ccggcgtaac catgaaagag acggcaaatc gtccgttgaa 2340

aggtgcgaca ccggcggcgt cagatataca ggaagcgaag gaaattctcg tcgagcatta 2400

cgataatatc gagcagaaaa tcgatgatat tgaccacgag attgccgatt tgcaggcgaa 2460

acgtacccgc ctggtgcagc aacatccgcg aattgatgaa taagctgaaa cggatggcct 2520

gatgtgacgc tgtcttatca ggccaattga actcttaagg ttcacttaat ctctgacgcg 2580

catactctcc tccaggttaa cggaggagag tgcaatgaaa aaccgtgttt atgaaagttt 2640

aactaccgtg ttcagcgtgc tggtggtcag cagctttctt tatatctggt ttgccacgta 2700

ctgatctttc ttcagccgta cccaggcccg cgtgccggaa gtctcttgcc ggttttgcag 2760

gaaaaactgc ccgtgatgca actgtgtaat gcggctgaca atacttaacc ccagaccaat 2820

cccgccataa cggctgtcca tacgtacaaa cgctttactc aactccccgc atttactctc 2880

atcaatacct ggtccttcat cttcaactgc catgaccgct ccgtcatctt cttgcagctt 2940

aatcataatg ttgctgcctt gcgggctgta acgatgggcg ttttctacca ggtttcgcaa 3000

taacatccgc agcagggttg catcaccctg aacggtgatg tcggcggcgc tctctggcaa 3060

tagcagggtt tgctgtcgct ggtcgagcat ggtactgagt tcgtcatacg aggggagaat 3120

gacatcttcc agcagtttta catgttgata attaccggaa gaaaatgact gtccggcacg 3180

cgccagttgc agcagctggg agacgctctc catcatctga tcaagccgtg ccactaacgg 3240

tgctacatca atgtgatgcg ttttcgccag cagttccaga tgcaaacgca cccccgccag 3300

tggcgttcgc agttcgtgcg cgacgtcagc ggtaaacaac ctttcgttat ccagcgtgct 3360

ggtcaggcga ctgaccagat cgtttaacgc cgaaaccacc gcttcgattt cgagggtggc 3420

gctgtgaatg gcaatgggcg ttaagttgtc ggcggtgcgc gcttccagct ctttttgcag 3480

ctccgccagc gggcgggtga tgc 3503

5