



(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)



1-0041505

(51)<sup>2020.01</sup> C12N 15/70; C12P 13/00; C12N 9/88; (13) B  
C12N 15/77; C12N 15/81

(21) 1-2020-01284

(22) 13/08/2018

(86) PCT/KR2018/009246 13/08/2018

(87) WO 2019/035612 21/02/2019

(30) 10-2017- 0103795 16/08/2017 KR

(45) 25/10/2024 439

(43) 25/06/2020 387

(73) CJ CHEILJEDANG CORPORATION (KR)

330, Dongho-ro, Jung-gu, Seoul 04560, Republic of Korea

(72) KIM, Sol (KR); LEE, Kyusung (KR); LEE, Joo Hee (KR); SEOK, Jong-cheol (KR);  
JANG, Jae Woo (KR).

(74) Công ty TNHH Sáng chế ACTIP (ACTIP PATENT LIMITED)

(54) VI SINH VẬT SẢN SINH CÁC AXIT AMIN DẠNG MYCOSPORIN VÀ  
PHƯƠNG PHÁP ĐỂ SẢN XUẤT CÁC AXIT AMIN DẠNG MYCOSPORIN  
BẰNG VI SINH VẬT NÀY

(57) Sáng chế đề cập đến vi sinh vật sản sinh các axit amin dạng mycosporin, và phương pháp sản xuất các axit amin dạng mycosporin bằng vi sinh vật này. Vi sinh vật của sáng chế thể hiện khả năng được cải thiện sản sinh các axit amin dạng mycosporin và do đó có thể được sử dụng hiệu quả trong sản xuất các axit amin dạng mycosporin.

### **Lĩnh vực kỹ thuật được đề cập**

Sáng chế đề cập đến vi sinh vật sản sinh các axit amin dạng mycosporin, và phương pháp sản xuất các axit amin dạng mycosporin bằng vi sinh vật này.

### **Tình trạng kỹ thuật của sáng chế**

Bức xạ tia cực tím được phát ra từ mặt trời chứa UV-A (tia cực tím A, bước sóng trong khoảng từ 320 nm đến 400 nm), UV-B (tia cực tím B, bước sóng trong khoảng từ 290 nm đến 320 nm), và UV-C (tia cực tím C, bước sóng trong khoảng từ 100 nm đến 280 nm). Trong số các tia mặt trời, khoảng 6% UV-A và UV-B chiếu đến bề mặt trái đất, mặt khác, UV-C không chiếu đến bề mặt trái đất vì nó được hấp thụ và phân tán trong tầng ôzôn của trái đất và bầu khí quyển.

Mặc dù các tia UV này có thể cung cấp các khía cạnh có lợi, như là tổng hợp vitamin D trong cơ thể, điều trị các bệnh về da, các tác dụng khử trùng, v.v., các tia này cũng có các khía cạnh tác động có hại như là cháy nắng, ung thư da, lão hóa, bệnh da nhạy cảm với ánh sáng, và đột biến. Đã được biết rằng UV-A thâm nhập ngay cả vào lớp hạ bì, chủ yếu gây ra nám và lão hóa da và có liên quan đến phát triển các bệnh da nhạy cảm với ánh sáng, trong khi UV-B có liên quan đến việc gây cháy nắng, nám và ung thư da do thâm nhập vào phần trên của các lớp biểu bì và hạ bì với các tia năng lượng cao.

Các nỗ lực đã được thực hiện để ngăn chặn ánh sáng mặt trời nhằm ngăn chặn các tác động phụ gây ra bởi các tia cực tím. Theo đó, các loại chất chống nắng có thể được phân loại thành chất chống nắng hóa học và chất chống nắng vật lý. Các chất chống nắng hóa học chủ yếu ngăn chặn sự xâm nhập của ánh sáng mặt trời thông qua sự hấp thụ, trong khi các chất chống nắng vật lý ngăn chặn ánh sáng mặt trời thông qua sự phản xạ và tán xạ ánh sáng mặt trời.

Các chất chống nắng hóa học chứa một hoặc nhiều thành phần hấp thụ tia cực tím, như là PABA, PABA este (amyl dimetyl PABA, octyl dimetyl PABA), xinamat (xinoxat), salixylat (Homomentyl salixylat), long não, v.v. là các chất chủ yếu hấp thụ

UV-B, và benzophenon (oxybenzon, đioxybenzon, sulisobenzon), đibenzoyl metan, anthranilat, v.v., là các chất chủ yếu hấp thụ UV-A. Mặc dù các chất chống nắng hóa học này có thể mang lại hiệu quả chống tia UV bằng cách hấp thụ các tia UV, một số trong đó có thể gây kích ứng ở da hoặc mắt. Đặc biệt, PABA, PABA este, benzophenon, xinnamat, v.v., được biết là gây viêm da tiếp xúc. Ngoài ra, các vấn đề đã được báo cáo rằng một số nghiên cứu khác có liên quan đến việc phát triển các phản ứng quá mẫn ở da, do đó, việc sử dụng các chất chống nắng hóa học và lượng sử dụng chúng bị hạn chế ở một số quốc gia.

Các chất chống nắng vật lý có chứa các thành phần tồn tại trong tự nhiên, bảo vệ da bằng cách phản xạ và tán xạ các tia UV thâm nhập vào da. Ví dụ, các chất chống nắng vật lý như là titan đioxit, talc (magiê silicat), magiê oxit, kẽm oxit, cao lanh, v.v. có thể mang lại hiệu quả chống tia UV đối với cả UV-A và UV-B. Hơn nữa, các chất này có ưu điểm ở chỗ chúng không có các tác dụng phụ, như là viêm da tiếp xúc, và không dễ dàng bị rửa trôi bởi nước. Ngược lại, chúng cũng có nhược điểm là khó có thể duy trì lượng hiệu quả của các chất chống nắng vật lý trong khi thực hiện theo công thức mong muốn của chúng, và chúng để lại một lớp trắng khi thoa lên da.

Các axit amin dạng mycosporin (Mycosporine-like amino acids: MAAs) là những chất được tìm thấy trong các sinh vật tự nhiên và được biết là hấp thụ hiệu quả UV-A (bước sóng trong khoảng 320 nm đến 400 nm) và UV-B (bước sóng trong khoảng 290 nm đến 320 nm). Có 35 loại MAAs hoặc nhiều hơn trong tự nhiên theo các loại tiền chất axit amin, vòng cyclohexenon hoặc cyclohexenimin (Mar. Biol., 1991, 108: 157-166; Tạp chí Cây thuốc 2015, 81: 813-820). Gần đây, MAAs glycosyl hóa khác nhau được tìm thấy trong vi tảo, đã được báo cáo là đóng vai trò như chất chống oxy hóa tốt (Tạp chí Quang hóa và Quang sinh học, 2015, 142: 154-168). Ngoài ra, MAAs được biết đến không chỉ cung cấp khả năng chống UV mà còn cho thấy khả năng chống oxy hóa, thẩm thấu và căng thẳng nhiệt (So sánh Sinh hóa và Sinh lý học. phần C, Độc tính và Dược lý, 2007, 146: 60-78; Tạp chí Quang hóa và Quang sinh học, phần B, 2007, 89: 29-35).

Tuy nhiên, lượng MAAs được sản xuất trong vi tảo là rất nhỏ ở mức vài  $\mu\text{g}$ , và các điều kiện để phân tách, tách chiết và tinh sạch MAAs bằng cách nuôi cấy vi tảo rất phức tạp, và do đó, sẽ khó sản xuất MAAs trong sản xuất quy mô lớn.

Tài liệu kỹ thuật liên quan

Tài liệu non-patent

Tài liệu non-patent 1: So sánh Sinh hóa và Sinh lý học. phần B, 1995, 112: 105-114.

Tài liệu non-patent 2: Báo cáo của Liên đoàn hiệp hội Vi sinh vật Châu Âu. 2007, 269: 1-10.

Tài liệu non-patent 3: Đánh giá thường niên về Sinh lý học. 2002, 64: 223-262.

Tài liệu non-patent 4: Sinh vật biển. 1991, 108: 157-166.

Tài liệu non-patent 5: Tạp chí Quang hóa và Quang sinh học, phần B: Sinh học. 2015, 142: 154-168

Tài liệu non-patent 6: Đánh giá sinh học. 1999, 74: 311-345.

Tài liệu non-patent 7: Sinh học phân tử và tiến hóa 2006, 23: 1437-1443.

Tài liệu non-patent 8: Khoa học, 2010, 329: 1653-1656.

Tài liệu non-patent 9: Hệ gen 2010, 95: 120-128.

Tài liệu non-patent 10: Địa sinh học. J. 1997. 14: 231-241.

Tài liệu non-patent 11: So sánh sinh hóa và sinh lý học, phần C: Độc tính và Dược lý. 2007. 146: 60-78.

Tài liệu non-patent 12: Tạp chí thực vật học Canada. 2003. 81: 131-138.

Tài liệu non-patent 13: Tạp chí Quang hóa và Quang sinh học, phần B. 2007, 89: 29-35.

Tài liệu non-patent 14: Tạp chí Vi khuẩn học. 2011. 193(21): 5923-5928.

Tài liệu non-patent 15: Tạp chí về Cây thuốc. 2015. 81: 813-820

Tài liệu non-patent 16: Vật liệu ứng dụng & giao diện của Hiệp hội hóa học Hoa Kỳ 2015. 7: 16558-16564

Tài liệu non-patent 17: Ứng dụng và Vi sinh môi trường. 2016, 82(20): 6167-6173

Tài liệu non-patent 18: Tạp chí Hóa học và Sinh học. 2015, 16: 320-327

Tài liệu non-patent 19: Các phương pháp trong sinh học phân tử. 2013, 1073: 43-7

Tài liệu non-patent 20: Công nghệ enzym và vi sinh vật., tháng 1, năm 2016, 82: 96-104

Tài liệu non-patent 21: Đánh giá thiên nhiên, 2011, 9: 791-802

Các tác giả sáng chế đã nỗ lực rất nhiều để tăng sản xuất MAAs trong các vi sinh vật. Kết quả là, đã xác nhận rằng việc sản xuất MAAs đã được tăng lên thông qua các nghiên cứu khác nhau về việc bất hoạt hoạt động 3-dehydroquinate dehydratase trong vi sinh vật để sản sinh MAAs, nhờ đó hoàn thành sáng chế.

### **Bản chất kỹ thuật của sáng chế**

Mục đích của sáng chế là cung cấp vi sinh vật để sản sinh các axit amin dạng mycosporin, trong đó hoạt động của 3-dehydroquinate dehydratase bị bất hoạt so với vi sinh vật không biến đổi.

Mục đích khác của sáng chế là đề xuất phương pháp để sản xuất các axit amin dạng mycosporin, bao gồm:

nuôi cấy vi sinh vật; và

thu hồi các axit amin dạng mycosporin từ vi sinh vật được nuôi cấy hoặc môi trường nuôi cấy.

### **Hiệu quả đạt được của sáng chế**

Vi sinh vật của sáng chế cho thấy khả năng được cải thiện để sản sinh các axit amin dạng mycosporin, và do đó có thể được sử dụng hiệu quả trong sản xuất các axit amin dạng mycosporin.

### **Mô tả chi tiết sáng chế**

Các chi tiết cụ thể của sáng chế có thể được giải thích như sau.

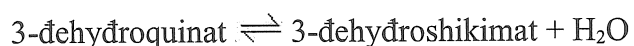
Trong khi, các giải thích và phương án được bộc lộ trong sáng chế có thể được áp dụng cho các giải thích và phương án khác, tương ứng. Do đó, tất cả sự kết hợp của các yếu tố khác nhau được tiết lộ ở đây thuộc về phạm vi sáng chế. Ngoài ra, phạm vi của sáng chế không nên bị giới hạn bởi các phần mô tả cụ thể được mô tả dưới đây. Hơn nữa, người có hiểu biết trung bình trong cùng lĩnh vực kỹ thuật có thể nhận ra

hoặc xác nhận, chỉ sử dụng thí nghiệm thông thường, nhiều đương lượng với các khía cạnh cụ thể của sáng chế được mô tả trong đơn sáng chế này. Theo đó, các đương lượng này cũng được dự định thuộc phạm vi của sáng chế.

Để đạt được các mục đích trên, khía cạnh của sáng chế cung cấp vi sinh vật để sản sinh các axit amin dạng mycosporin, trong đó hoạt động 3-đehydroquinat dehydrataza bị bất hoạt so với vi sinh vật không biến đổi.

Như được sử dụng ở đây, thuật ngữ “3-đehydroquinat dehydrataza” đề cập đến enzyme xúc tác cho phản ứng thuận nghịch được biểu thị bởi sơ đồ phản ứng dưới đây, và đặc biệt có thể chuyển 3-đehydroquinat thành 3-đehydroshikimat, nhưng sáng chế không giới hạn ở đây.

Sơ đồ phản ứng]



Như được sử dụng ở đây, thuật ngữ “bất hoạt” đề cập đến trường hợp trong đó hoạt động của protein enzyme ban đầu ở vi sinh vật bị suy yếu so với hoạt động nội sinh của chúng hoặc hoạt động trước khi biến đổi; trường hợp trong đó protein không được biểu hiện theo bất kỳ cách nào; hoặc trường hợp trong đó protein được biểu hiện nhưng biểu hiện không hoạt động. Bất hoạt là khái niệm bao gồm trường hợp hoạt động của chính enzyme bị suy yếu hoặc bị loại bỏ so với hoạt động của enzyme ban đầu được sở hữu bởi vi sinh vật, do sự biến đổi trong polynucleotit mã hóa enzyme, v.v.; trường hợp trong đó mức độ hoạt động của enzyme tổng thể trong tế bào bị giảm hoặc loại bỏ so với vi sinh vật kiểu dại, do sự ức chế biểu hiện gen mã hóa enzyme, hoặc ức chế dịch mã, v.v.; trường hợp trong đó một phần hoặc toàn bộ gen mã hóa enzyme bị xóa; và sự kết hợp của các trường hợp trên, nhưng sáng chế không giới hạn ở đây. Thuật ngữ “vi sinh vật không biến đổi” đề cập đến vi sinh vật có hoạt tính của protein cụ thể ban đầu được sở hữu bởi chủng bố mẹ trước khi sửa đổi chúng, khi các đặc điểm của vi sinh vật được biến đổi thông qua sửa đổi gen trong protein cụ thể được sở hữu bởi vi sinh vật để so sánh do các yếu tố tự nhiên hoặc nhân tạo. Như được sử dụng ở đây, “vi sinh vật không biến đổi” có thể được sử dụng thay thế cho “vi sinh vật có hoạt động nội sinh”.

Sự bất hoạt hoạt động enzyme có thể đạt được bằng các phương pháp khác nhau được biết đến trong kỹ thuật. Các ví dụ của các phương pháp bao gồm 1) phương pháp

xóa một phần hoặc toàn bộ gen mã hóa enzym trên nhiễm sắc thể; 2) phương pháp biến đổi trình tự điều hòa biểu hiện sao cho biểu hiện của gen mã hóa protein trên nhiễm sắc thể bị giảm; 3) phương pháp biến đổi trình tự gen mã hóa protein trên nhiễm sắc thể sao cho hoạt động của protein bị loại bỏ hoặc suy yếu; 4) phương pháp đưa oligonucleotit đối mã liên kết bổ sung với bản phiên mã của gen mã hóa protein trên nhiễm sắc thể (ví dụ, RNA đối mã); 5) phương pháp bổ sung nhân tạo trình tự bổ sung với trình tự Shine-Dalgarno của gen mã hóa protein trên nhiễm sắc thể với vùng ngược hướng của trình tự Shine-Dalgarno để tạo thành cấu trúc thứ cấp, do đó không thể gắn được với ribosom; và 6) phương pháp kỹ thuật sao chép ngược (reverse transcription engineering: RTE), trong đó thêm vùng gen khởi động được sao chép ngược vào đầu 3' của khung đọc mở (open reading frame: ORF) của chuỗi polynucleotit mã hóa protein, hoặc kết hợp của các phương pháp trên, nhưng sáng chế không bị giới hạn cụ thể ở đây.

Phương pháp xóa một phần hoặc toàn bộ gen mã hóa protein trên nhiễm sắc thể có thể được thực hiện bằng cách thay thế polynucleotit mã hóa protein đích nội sinh trong nhiễm sắc thể với polynucleotit có trình tự axit nucleic bị xóa một phần, hoặc gen đánh dấu thông qua vectơ để chèn vào nhiễm sắc thể của vi sinh vật. Ví dụ của phương pháp xóa một phần hoặc toàn bộ polynucleotit, phương pháp xóa polynucleotit bằng tái tổ hợp tương đồng có thể được sử dụng, nhưng sáng chế không giới hạn ở đây.

Phương pháp biến đổi trình tự điều hòa biểu hiện có thể được thực hiện bằng cách gây ra biến đổi trong trình tự điều hòa biểu hiện thông qua xóa, chèn, thay thế bảo thủ hoặc không bảo thủ, hoặc kết hợp các cách trên để làm suy yếu thêm hoạt động của trình tự điều hòa biểu hiện; hoặc bằng cách thay thế trình tự bằng trình tự axit nucleic có hoạt động yếu hơn. Trình tự điều hòa biểu hiện có thể bao gồm vùng gen khởi động, trình tự vùng vận hành, trình tự mã hóa vùng liên kết với ribosom, và trình tự để điều hòa kết thúc phiên mã và dịch mã, nhưng sáng chế không giới hạn ở đây.

Phương pháp biến đổi trình tự gen trên nhiễm sắc thể có thể được thực hiện bằng cách gây ra biến đổi trong trình tự gen thông qua xóa, chèn, thay thế bảo thủ hoặc không bảo thủ, hoặc kết hợp các cách trên để làm suy yếu thêm hoạt động của enzym; hoặc bằng cách thay thế trình tự bằng trình tự gen được biến đổi để có hoạt động yếu

hơn hoặc trình tự gen được biến đổi để không có hoạt động nào cả, nhưng sáng chế không giới hạn ở đây.

Polynucleotit có thể được mô tả như là gen trong trường hợp nó đề cập đến tập hợp các polynucleotit có khả năng thực hiện các chức năng. Theo sáng chế, các polynucleotit và các gen có thể được sử dụng thay thế cho nhau, và các trình tự polynucleotit và trình tự nucleotit có thể được sử dụng thay thế cho nhau.

Như được sử dụng ở đây, thuật ngữ “phần”, mặc dù nó có thể thay đổi tùy thuộc vào các loại polynucleotit, cụ thể có thể đề cập đến 1 đến 300 nucleotit, cụ thể hơn là 1 đến 100 nucleotit, và cụ thể hơn nữa là 1 đến 50 nucleotit, nhưng sáng chế không giới hạn cụ thể ở đây.

Trong vi sinh vật của sáng chế, hoạt động của ít nhất một enzym được chọn từ nhóm bao gồm 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza I/II, và 3-đehydroquinat synthaza, cụ thể là hoạt động của nhiều hơn một, nhiều hơn hai, nhiều hơn ba, hoặc tất cả enzym có thể được tăng cường so với vi sinh vật không biến đổi.

2-đehydro-3-đeoxyphosphoheptonat aldolaza đề cập đến enzym xúc tác cho phản ứng thuận nghịch được biểu thị bởi sơ đồ phản ứng dưới đây và đặc biệt có thể tổng hợp 3-đeoxy-arabino-heptulonat 7-phosphat, nhưng sáng chế không giới hạn ở đây.

Sơ đồ phản ứng

phosphoenolpyruvat + D-erythrose-4-phosphat + H<sub>2</sub>O  $\rightleftharpoons$  3-đeoxy-D-arabino-heptulosonat-7-phosphat + phosphat

Phosphoenolpyruvat synthetaza đề cập đến enzym xúc tác phản ứng thuận nghịch được biểu thị bằng sơ đồ phản ứng dưới đây, và đặc biệt có thể tổng hợp phosphoenolpyruvat, nhưng sáng chế không giới hạn ở đây.

Sơ đồ phản ứng

ATP + pyruvat + H<sub>2</sub>O  $\rightleftharpoons$  AMP + phosphoenolpyruvat + phosphat

Transketolaza I/II đề cập đến enzym xúc tác phản ứng thuận nghịch được biểu thị bằng sơ đồ phản ứng dưới đây.

Sơ đồ phản ứng

Sedoheptulos 7-phosphat + D-glyceraldehyt 3-phosphat = D-ribose 5-phosphat + D-xylulose 5-phosphat

3-đehydroquinat synthaza đề cập đến enzyme xúc tác phản ứng thuận nghịch được biểu thị bằng sơ đồ phản ứng dưới đây, và đặc biệt có thể tổng hợp 3-đehydroquinat (3-DHQ), nhưng sáng chế không giới hạn ở đây.

Sơ đồ phản ứng



Như được sử dụng ở đây, thuật ngữ “tăng cường hoạt động” có nghĩa là hoạt động của protein enzyme được đưa vào, hoặc hoạt động được tăng cường so với hoạt động nội sinh có ở vi sinh vật hoặc hoạt động trước khi biến đổi. “Sự đưa vào” của hoạt động nghĩa là hoạt động của polypeptit cụ thể mà vi sinh vật ban đầu không có được biểu hiện tự nhiên hoặc nhân tạo. Ví dụ, tăng cường hoạt động có thể bao gồm cả tăng cường bằng cách đưa vào 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza I/II, và/hoặc 3-đehydroquinat synthaza; hoặc tăng cường hoạt động nội sinh của 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza I/II, và/hoặc 3-đehydroquinat synthaza. Cụ thể, tăng cường hoạt động theo sáng chế có thể được thực hiện bằng các phương pháp sau:

- 1) phương pháp tăng cường số lượng bản sao của polynucleotit mã hóa các enzyme;
- 2) phương pháp biến đổi trình tự điều hòa biểu hiện sao cho biểu hiện của polynucleotit được tăng lên;
- 3) phương pháp biến đổi trình tự polynucleotit trên nhiễm sắc thể sao cho hoạt động của các enzyme được tăng cường; và
- 4) phương pháp biến đổi để tăng cường hoạt động bằng cách kết hợp các phương pháp trên, nhưng sáng chế không giới hạn ở đây.

Sự tăng số lượng bản sao polynucleotit trong phương pháp 1) ở trên có thể được thực hiện dưới dạng polynucleotit được liên kết hoạt động với vectơ hoặc bằng cách chèn vào nhiễm sắc thể của tế bào chủ, nhưng không bị giới hạn cụ thể ở đây. Ngoài ra, trong khía cạnh để tăng số lượng bản sao có thể được thực hiện bằng cách đưa polynucleotit ngoại lai để thể hiện hoạt tính của enzyme vào trong tế bào chủ, hoặc

polynucleotit được biến đổi trong đó các đơn vị mã hóa của polynucleotit ngoại lai đã được tối ưu hóa. Polynucleotit ngoại lai có thể được sử dụng mà không bị giới hạn nguồn gốc hoặc trình tự của nó miễn là nó thể hiện hoạt tính giống hệt hoặc tương tự enzym. Sự đưa vào có thể được thực hiện bởi những người có hiểu biết trung bình trong cùng lĩnh vực kỹ thuật bằng cách chọn phương pháp biến nạp phù hợp được biết trong kỹ thuật, và enzym có thể được sản xuất khi các polynucleotit đưa vào được biểu hiện trong tế bào chủ, do đó tăng cường hoạt động của nó.

Tiếp theo, sự biến đổi trình tự điều hòa biểu hiện sao cho biểu hiện của polynucleotit được tăng cường trong phương pháp 2) có thể được thực hiện bằng cách gây ra biến đổi trong trình tự thông qua xóa, chèn, thay thế bảo thủ hoặc không bảo thủ của trình tự axit nucleic, hoặc kết hợp các cách trên để tăng cường hơn nữa hoạt động của trình tự điều hòa biểu hiện, hoặc bằng cách thay thế trình tự axit nucleic có hoạt tính mạnh hơn, nhưng không bị giới hạn cụ thể ở đây. Ngoài ra, trình tự điều hòa biểu hiện có thể bao gồm vùng gen khởi động, trình tự vùng vận hành, trình tự mã hóa vùng liên kết ribosom, trình tự điều hòa kết thúc phiên mã và dịch mã, v.v., nhưng sáng chế không bị giới hạn cụ thể ở đây.

Cụ thể, vùng gen khởi động dị hợp mạnh có thể được liên kết với vùng ngược hướng của đơn vị biểu hiện của polynucleotit thay vì vùng gen khởi động ban đầu. Các ví dụ của vùng gen khởi động mạnh bao gồm vùng gen khởi động *CJ7*, vùng gen khởi động *lysCPI*, vùng gen khởi động *EF-Tu*, vùng gen khởi động *groEL*, vùng gen khởi động *aceA* hoặc *aceB*, v.v. Cụ thể hơn, tốc độ biểu hiện của polynucleotit mã hóa enzym có thể được tăng cường bằng cách liên kết hoạt động của polynucleotit với vùng gen khởi động *lysCPI* (công bố đơn quốc tế số WO2009/096689) hoặc vùng gen khởi động *CJ7* (công bố đơn quốc tế số WO2006/065095), nhưng sáng chế không giới hạn ở đây.

Hơn nữa, sự biến đổi trình tự polynucleotit trên nhiễm sắc thể trong phương pháp 3) có thể được thực hiện bằng cách gây ra biến đổi trong trình tự điều hòa biểu hiện thông qua xóa, chèn, thay thế bảo thủ hoặc không bảo thủ của trình tự axit nucleic, hoặc kết hợp các cách trên để tăng cường hơn nữa hoạt động của trình tự polynucleotit, hoặc bằng cách thay thế trình tự polynucleotit bằng trình tự polynucleotit được sửa đổi để có hoạt động mạnh hơn, nhưng không bị giới hạn cụ thể ở đây.

Cuối cùng, phương pháp biến đổi để tăng cường hoạt động bằng cách kết hợp các phương pháp 1) đến 3) trong phương pháp 4) có thể được thực hiện bằng cách áp dụng kết hợp của ít nhất một trong các phương pháp sau: tăng số lượng bản sao polynucleotit mã hóa protein; sửa đổi trình tự điều hòa biểu hiện sao cho biểu hiện của polynucleotit được tăng lên; sửa đổi trình tự polynucleotit trên nhiễm sắc thể, và sửa đổi polynucleotit ngoại lai thể hiện hoạt động enzym hoặc polynucleotit biến đổi được tối ưu hóa đơn vị mã hóa của chúng.

Như được sử dụng ở đây, thuật ngữ “vector” đề cập đến cấu trúc DNA chứa trình tự nucleotit của polynucleotit mã hóa protein đích, được liên kết hoạt động với trình tự điều hòa phù hợp để protein đích có thể được biểu hiện trong vật chủ thích hợp. Trình tự điều hòa bao gồm vùng gen khởi động có khả năng khởi động phiên mã, bất kỳ trình tự vùng vận hành nào để kiểm soát phiên mã, trình tự mã hóa vùng liên kết mRNA ribosom thích hợp, và trình tự điều hòa kết thúc phiên mã và dịch mã. Sau khi được biến nạp vào trong vật chủ thích hợp, vector có thể được sao chép hoặc hoạt động không phân biệt với bộ gen của vật chủ và có thể được tích hợp vào bộ gen của chính vật chủ.

Vector được sử dụng trong sáng chế không bị giới hạn cụ thể miễn là nó có thể được sao chép trong vật chủ và bất kỳ vector nào được biết đến trong kỹ thuật có thể được sử dụng. Các ví dụ về các vector được sử dụng thông thường có thể bao gồm các plasmid tự nhiên hoặc tái tổ hợp, các cosmid, các vi rút và các thực khuẩn thể. Ví dụ, như là vector phage hoặc vector cosmid, pWE15, M13, MBL3, MBL4, IXII, ASHII, APII, t10, t11, Charon4A, Charon21A, v.v., có thể được sử dụng, và vector plasmid, những vector dựa trên pBR, pUC, pBluescriptII, pGEM, pTZ, pCL, pET, v.v. có thể được sử dụng. Cụ thể, các vector pDZ, pACYC177, pACYC184, pCL, pECCG117, pUC19, pBR322, pMW118, pCC1BAC, v.v. có thể được sử dụng, nhưng vector không bị giới hạn ở đây.

Vector có thể được sử dụng trong sáng chế không bị giới hạn cụ thể, và vector biểu hiện đã biết có thể được sử dụng. Ngoài ra, polynucleotit mã hóa protein đích có thể được chèn vào trong nhiễm sắc thể thông qua vector để chèn vào nhiễm sắc thể. Việc chèn polynucleotit vào trong nhiễm sắc thể có thể được thực hiện bằng cách sử dụng bất kỳ phương pháp nào được biết trong kỹ thuật, ví dụ, bằng cách tái tổ hợp tương đồng, nhưng sáng chế không giới hạn ở đây. Dấu hiệu lựa chọn để xác nhận việc chèn

vào trong nhiễm sắc thể có thể được thêm vào. Dấu hiệu lựa chọn được sử dụng để lựa chọn tế bào được biến nạp vào trong vector, v.v., để xác nhận có hay không phân tử axit nucleic đích đã được chèn, và các dấu hiệu cung cấp các kiểu hình có thể lựa chọn, như là kháng thuốc, yêu cầu dinh dưỡng, kháng các tác nhân gây độc tế bào hoặc biểu hiện của các protein bề mặt có thể được sử dụng. Chỉ các tế bào biểu hiện các dấu hiệu lựa chọn có thể tồn tại hoặc biểu hiện các đặc điểm kiểu hình khác trong môi trường được xử lý bằng các tác nhân chọn lọc, do đó, các tế bào được biến nạp có thể được chọn dễ dàng.

Như được sử dụng ở đây, thuật ngữ “biến nạp” đề cập đến quá trình đưa vector chứa polynucleotit mã hóa polypeptit đích vào trong tế bào chủ, do đó cho phép biểu hiện protein được mã hóa bởi polynucleotit trong tế bào chủ. Miễn là polynucleotit được biến nạp có thể được biểu hiện trong tế bào chủ, việc nó được đưa vào nhiễm sắc thể của tế bào chủ và nằm ở đó, hoặc nằm bên ngoài nhiễm sắc thể, và cả hai trường hợp đều có thể được bao gồm. Ngoài ra, polynucleotit chứa DNA và RNA mã hóa polypeptit đích. Polynucleotit có thể được đưa vào dưới mọi hình thức miễn là nó có thể được đưa vào tế bào chủ và được biểu hiện trong đó. Ví dụ, polynucleotit có thể được đưa vào trong tế bào chủ dưới hình thức của catxet biểu hiện là cấu trúc gen bao gồm tất cả các yếu tố cần thiết để tự biểu hiện. Catxet biểu hiện có thể bao gồm thông thường vùng gen khởi động được liên kết hoạt động với polynucleotit, vùng kết thúc, vùng liên kết ribosom, và đơn vị mã hóa dừng. Catxet biểu hiện có thể dưới dạng vector biểu hiện có khả năng tự sao chép. Ngoài ra, polynucleotit có thể được đưa vào trong tế bào chủ vì nó được liên kết hoạt động với trình tự cần thiết để nó tự biểu hiện trong tế bào chủ, nhưng sáng chế không giới hạn ở đây. Phương pháp để biến nạp vector bao gồm bất kỳ phương pháp nào của việc đưa axit nucleic vào trong tế bào, và có thể được thực hiện bằng cách chọn kỹ thuật tiêu chuẩn phù hợp được biết trong kỹ thuật theo tế bào chủ. Ví dụ, biến nạp có thể được thực hiện thông qua kỹ thuật xung điện, kết tủa canxi phosphat ( $\text{CaPO}_4$ ), kết tủa canxi clorua ( $\text{CaCl}_2$ ), vi tiêm, kỹ thuật polyetylen glycol (PEG), kỹ thuật DEAE-dextran, kỹ thuật liposom cation, kỹ thuật liti axetat-DMSO, v.v., nhưng các phương pháp không bị giới hạn ở đây.

Hơn nữa, như được sử dụng ở trên, thuật ngữ “liên kết hoạt động” đề cập đến liên kết chức năng giữa trình tự polynucleotit ở trên và trình tự vùng gen khởi động bắt đầu và làm trung gian phiên mã của polynucleotit mã hóa protein đích của sáng chế. Liên

kết hoạt động có thể được điều chế bằng cách sử dụng kỹ thuật tái tổ hợp di truyền được biết đến trong kỹ thuật, tách DNA tại vị trí cụ thể và sự nối có thể được thực hiện bằng cách sử dụng các enzym như là lyses, ligaza, v.v. được biết đến trong kỹ thuật, nhưng sáng chế không giới hạn ở đây.

Thông tin di truyền của 3-đehydroquinat dehydrataza, 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza có thể thu được từ các cơ sở dữ liệu đã biết, và các ví dụ của các dữ liệu đã biết bao gồm Ngân hàng gen của Trung tâm thông tin công nghệ sinh học Quốc gia (NCBI), v.v., nhưng các cơ sở dữ liệu đã biết không bị giới hạn ở đây.

3-đehydroquinat dehydrataza, 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza có thể không bị giới hạn về nguồn gốc hoặc các trình tự của chúng, vì có thể có sự khác biệt trong trình tự axit amin của protein ức chế hoạt động tùy thuộc vào loại vi khuẩn hoặc vi sinh vật.

Cụ thể, 3-đehydroquinat dehydrataza có thể là protein chứa trình tự axit amin được biểu thị bởi SEQ ID NO: 72 hoặc SEQ ID NO: 80, và 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza có thể là protein chứa trình tự axit amin được biểu thị bởi các SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 78, và SEQ ID NO: 84, tương ứng, nhưng không bị giới hạn ở đây. Trong sáng chế, “protein chứa trình tự axit amin” có thể được sử dụng thay thế cho sự biểu hiện “protein có trình tự axit amin” hoặc “protein có trình tự axit amin”.

Ngoài ra, theo sáng chế, các enzym có thể không chỉ bao gồm protein chứa trình tự axit amin được biểu thị bằng SEQ ID NO được mô tả ở trên, nhưng cũng là protein có sự tương đồng lớn hơn 80%, cụ thể là lớn hơn 90%, lớn hơn 95%, thậm chí cụ thể hơn nữa là lớn hơn 99% so với các trình tự axit amin ở trên, miễn là protein có hoạt tính sinh học giống hệt hoặc tương ứng với từng enzym.

Hơn nữa, rõ ràng là, trình tự axit amin có sự tương đồng với các trình tự ở trên, trình tự axit amin với xóa, sửa đổi, thay thế, hoặc thêm một phần của các trình tự cũng nằm trong phạm vi của sáng chế miễn là trình tự axit amin có hoạt tính sinh học giống hệt nhau hoặc tương ứng với protein enzyme có trình tự axit amin được biểu thị bởi

SEQ ID NO được mô tả ở trên.

Polynucleotit mã hóa 3-đehydroquinat dehydrataza, 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza của sáng chế có thể bao gồm polynucleotit chứa trình tự axit amin được biểu thị bởi SEQ ID NO được mô tả ở trên, hoặc polynucleotit mã hóa protein có sự tương đồng lớn hơn 80%, cụ thể là lớn hơn 90%, lớn hơn 95%, thậm chí cụ thể hơn là lớn hơn 99% so với các trình tự axit amin, miễn là polynucleotit có hoạt tính sinh học giống hệt nhau hoặc tương ứng với từng enzym.

Ngoài ra, đối với polynucleotit mã hóa 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza, xem xét các đơn vị mã hóa được ưu tiên bởi các sinh vật để biểu hiện protein do đơn vị mã hóa thoái hóa, nhiều sửa đổi khác nhau có thể được thực hiện trên vùng mã hóa trong phạm vi mà không thay đổi trình tự axit amin của polypeptit. Theo đó, polynucleotit có thể chứa bất kỳ trình tự polypeptit nào mã hóa cho mỗi protein enzym mà không có giới hạn.

Hơn nữa, đầu dò có thể được điều chế từ các trình tự gen đã biết, ví dụ, bất kỳ trình tự lai với trình tự bổ sung với toàn bộ hoặc một phần trình tự nucleotit trong điều kiện nghiêm ngặt để mã hóa protein có hoạt động 3-đehydroquinat dehydrataza, 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza, và 3-đehydroquinat synthaza có thể được bao gồm nhưng không giới hạn ở đây.

Như được sử dụng ở đây, thuật ngữ “tương đồng” đề cập đến mức độ tương ứng với trình tự axit amin hoặc trình tự nucleotit nhất định và có thể được biểu thị bằng phần trăm. Theo sáng chế, trình tự tương đồng có hoạt tính giống hoặc tương tự với trình tự axit amin hoặc trình tự nucleotit đã cho có thể được chỉ định theo thuật ngữ “% tương đồng”. Ví dụ, tương đồng có thể được xác nhận bằng cách sử dụng phần mềm tiêu chuẩn để tính toán các tham số như là điểm số, giống hệt, và tương tự, cụ thể BLAST 2.0, hoặc bằng các trình tự so sánh thông qua thí nghiệm lai trong điều kiện nghiêm ngặt đã xác định, và các điều kiện lai thích hợp đã được xác định bởi người có hiểu biết trung bình trong cùng lĩnh vực kỹ thuật, và có thể được xác định bằng phương pháp đã được biết đến rộng rãi bởi những người có hiểu biết trung bình trong

cùng lĩnh vực kỹ thuật (Ví dụ, J. Sambrook và cộng sự., *Nhân dòng phân tử, Hướng dẫn thí nghiệm, Ấn bản lần 2, Phòng thí nghiệm Cold Spring Harbor*, Cold Spring Harbor, New York, 1989; F.M. Ausubel và cộng sự., *Các quy trình hiện tại trong Sinh học phân tử*, John Wiley & Sons, Inc., New York). Thuật ngữ “các điều kiện nghiêm ngặt” đề cập đến các điều kiện cho phép lai đặc hiệu giữa các polynucleotit. Các điều kiện như vậy được mô tả chi tiết trong tài liệu tham khảo (ví dụ, J. Sambrook và cộng sự).

Như được sử dụng ở đây, thuật ngữ “các axit amin dạng mycosporin (MAAs)” đề cập đến các hợp chất vòng hấp thụ các tia cực tím. Theo sáng chế, bất kỳ các axit amin dạng mycosporin có thể bao gồm nhưng không giới hạn miễn là nó có thể hấp thụ các tia UV, nhưng nó đặc biệt có thể là hợp chất chứa vòng xyclohexenon hoặc xyclohexenimin ở giữa; hoặc hợp chất trong đó các chất khác nhau, như là axit amin, v.v., được liên kết với vòng ở giữa. Cụ thể hơn là, nó có thể là mycosporin-2-glyxin, palythanol, axit palythenic, deoxygadusol, mycosporin-metylamin-threonin, mycosporin-glyxin-valin, palythin, asterina-330, shinorin, porphyra-334, euhalothece-362, mycosporin-glyxin, mycosporin-ornithin, mycosporin-lysin, mycosporin-glutamic, axit glyxin, mycosporin-metylamin-serin, mycosporin-taurin, palythen, palythin-serin, palythin-serin-sulfat, palythanol, usujiren, hoặc kết hợp các cách trên.

Trong sáng chế, các axit amin dạng mycosporin có thể được sử dụng thay thế cho MAA và MAAs.

Như được sử dụng ở đây, thuật ngữ “vi sinh vật để sản sinh các axit amin dạng mycosporin” đề cập đến vi sinh vật chứa gen của enzym tham gia trong sinh tổng hợp của các axit amin dạng mycosporin, hoặc cụm gen. Ngoài ra, như được sử dụng ở đây, thuật ngữ “gen sinh tổng hợp các axit amin dạng mycosporin” đề cập đến gen mã hóa enzym tham gia trong sinh tổng hợp các axit amin dạng mycosporin, và cũng chứa cụm gen. Gen sinh tổng hợp các axit amin dạng mycosporin bao gồm cả các gen ngoại sinh và/hoặc nội sinh của vi sinh vật miễn là vi sinh vật chứa gen có thể sản xuất mycosporin-như là các axit amin. Các gen ngoại sinh có thể là tương đồng và/hoặc dị hợp.

Gen sinh tổng hợp các axit amin dạng mycosporin có thể không bị giới hạn bởi các loài vi sinh vật có nguồn gốc từ các gen, miễn là vi sinh vật cùng tạo ra loại enzym

tham gia trong sinh tổng hợp các axit amin dạng mycosporin và do đó sản xuất các axit amin dạng mycosporin. Cụ thể, vi sinh vật có thể là *Anabaena variabilis*, *Nostoc punctiforme*, *Nodularia spumigena*, *Cyanothece* sp. PCC 7424, *Lyngbya* sp. PCC 8106, *Microcystis aeruginosa*, *Microcoleus chthonoplastes*, *Cyanothece* sp. ATCC 51142, *Crocospaera watsonii*, *Cyanothece* sp. CCY 0110, *Cylindrospermum stagnale* sp. PCC 7417, *Aphanothece halophytica* hoặc *Trichodesmium erythraeum* là vi khuẩn lam, hoặc *Magnaporthe oryzae*, *Pyrenophora tritici-repentis*, *Aspergillus clavatus*, *Nectria haematococca*, *Aspergillus nidulans*, *Gibberella zeae*, *Verticillium albo-atrum*, *Botryotinia fuckeliana*, hoặc *Phaeosphaeria nodorum* là các loài thuộc nấm, hoặc *Nematostella vectensis*, *Heterocapsa triquetra*, *Oxyrrhis marina*, *Karlodinium micrum*, *Actinosynnema mirum*, v.v., nhưng sáng chế không giới hạn ở đây.

Theo một phương án, vi sinh vật để sản sinh các axit amin dạng mycosporin của sáng chế có thể bao gồm gen sinh tổng hợp các axit amin dạng mycosporin.

Cụ thể, gen sinh tổng hợp các axit amin dạng mycosporin có thể không bị giới hạn bởi tên của các enzym hoặc có nguồn gốc vi sinh vật miễn là vi sinh vật có thể sản xuất các axit amin dạng mycosporin, và cụ thể có thể bao gồm ít nhất một, cụ thể một hoặc hai, hai hoặc ba, hoặc lớn hơn ba, hoặc tất cả các protein enzym được chọn từ nhóm bao gồm 2-demetyl 4-deoxygadusol synthaza, O-metyltransferaza, và C-N ligaza; hoặc gen mã hóa protein enzym có hoạt tính giống hệt và/hoặc tương tự như vậy.

Ví dụ, 2-demetyl 4-deoxygadusol synthaza là enzym chuyển đổi sedoheptulos-7-phosphat thành 2-demetyl-4-deoxygadusol. O-metyltransferaza là enzym chuyển đổi 2-demetyl-4-deoxygadusol thành 4-deoxygadusol, và glyxyl hóa của 4-deoxygadusol được xúc tác bởi C-N ligaza.

Ngoài ra, vi sinh vật để sản sinh các axit amin dạng mycosporin có thể chứa gen của enzym có hoạt tính gắn thêm nhóm dư axit amin với các axit amin dạng mycosporin, hoặc cụm các gen. Gen hoặc cụm các gen có thể không bị giới hạn bởi tên của các enzym hoặc nguồn gốc vi sinh vật miễn là vi sinh vật để sản sinh các axit amin dạng mycosporin có thể sản xuất các axit amin dạng mycosporin, đến hai hoặc nhiều nhóm dư axit amin được gắn vào, và cụ thể có thể chứa ít nhất một, cụ thể lớn hơn một, lớn hơn hai, lớn hơn ba, hoặc tất cả các protein enzym được chọn từ nhóm

bao gồm non-ribosomal peptid synthetaza (NRPS), enzym dạng non-ribosomal peptid synthetaza (enzym dạng NRPS), và D-Ala D-Ala ligaza (DDL); hoặc gen mã hóa protein enzym có hoạt tính giống hệt và/hoặc tương tự như vậy. Một số các axit amin dạng mycosporin có thể chứa nhóm dư axit amin thứ hai trong mycosporin-glyxin. Ít nhất một enzym được chọn từ nhóm bao gồm non-ribosomal peptid synthetaza, enzym dạng non-ribosomal peptid synthetaze, và D-Ala D-Ala ligaza có thể gắn nhóm dư axit amin thứ hai với mycosporin-glyxin.

Theo một phương án, vi sinh vật để sản sinh các axit amin dạng mycosporin có thể chứa bất kỳ enzym nào mà không giới hạn tên của enzym hoặc các loài vi sinh vật miễn là vi sinh vật có hoạt tính gắn axit amin thứ hai với mycosporin-glyxin, như là non-ribosomal peptid synthetaza, enzym dạng non-ribosomal peptid synthetaza, và D-Ala D-Ala ligaza.

Theo phương án của sáng chế, enzym dạng non-ribosomal peptid synthetaza (Ava\_3855) được tìm thấy trong *Anabaena variabilis* hoặc D-Ala D-Ala ligaza (NpF5597) được tìm thấy trong *Nostoc punctiforme* có thể sản xuất shinorin bằng cách gắn nhóm dư serin với mycosporin-glyxin. Trong phương án khác, mycosporin-2-glyxin có thể được tạo thành bằng cách gắn nhóm dư glyxin thứ hai thông qua D-Ala D-Ala ligaza homolog (Ap\_3855) được tìm thấy trong *Aphanothece halophytica*. Tương tự, trong *Actinosynnema mirum*, serin hoặc alanin có thể được gắn bởi D-Ala D-Ala ligaza để tạo thành shinorin hoặc mycosporin-glyxin-alanin. Vi sinh vật theo một phương án của sáng chế có thể chứa enzym phù hợp với sản xuất các axit amin dạng mycosporin mong muốn bằng cách chọn từ các enzym được mô tả ở trên hoặc các enzym có hoạt tính giống hệt và/hoặc tương tự như vậy.

2-đemetyl 4-đeoxygadusol synthaza, O-metyltransferaza, C-N ligaza, non-ribosomal peptid synthetaza, enzym dạng non-ribosomal peptid synthetaza, và/hoặc D-Ala D-Ala ligaza có thể được sử dụng trong sáng chế có thể không bị giới hạn bởi các loài có nguồn gốc từ các vi sinh vật, và bất kỳ enzym nào có thể được bao gồm không giới hạn miễn là nó được biết để làm các chức năng giống hệt và/hoặc tương tự và các hoạt tính. Ngoài ra, phạm vi tương đồng về số giữa các enzym có thể không bị giới hạn. Ví dụ, MylA, MylB, MylD, MylE và MylC của *Cylindrospermum stagnale* PCC 7417 là tương đồng với 2-đemetyl 4-đeoxygadusol synthaza, O-metyltransferaza, C-N ligaza, và D-Ala D-Ala ligaza có nguồn gốc từ *Anabaena variabilis* và *Nostoc*

*punctiforme*, và sự tương đồng giữa các tương đồng này là khoảng 61 đến 88% (*Appl Environ Microbiol*, 2016, 82(20), 6167-6173; *J Bacteriol*, 2011, 193(21), 5923-5928). Do đó, enzym có thể được sử dụng trong sáng chế có thể không bị giới hạn cụ thể bởi nguồn gốc các loài vi sinh vật hoặc trình tự tương đồng miễn là nó được biết đến để thể hiện các chức năng và tác dụng giống hệt và/hoặc tương tự. Hơn nữa, các tài liệu tham khảo non-patent được bộc lộ trong kỹ thuật đã có được kết hợp ở đây bằng cách tham khảo toàn bộ.

Ngoài ra, gen sinh tổng hợp các axit amin dạng mycosporin có thể là polynucleotit mã hóa protein chứa trình tự axit amin được biểu thị bằng các SEQ ID NO: 2, 4, 86, 88, 90, 92, 94, 96, 98, 100, 102, hoặc 104, nhưng sáng chế không giới hạn ở đây.

Ngoài ra, gen sinh tổng hợp các axit amin dạng mycosporin có thể gồm trình tự nucleotit mã hóa protein chứa trình tự axit amin có sự tương đồng của 50%, 60%, hoặc lớn hơn 70%, cụ thể là lớn hơn 80%, 90%, cụ thể hơn là lớn hơn 95%, cụ thể nhất là lớn hơn 99% so với trình tự axit amin được biểu thị bằng các SEQ ID NO: 2, 4, 86, 88, 90, 92, 94, 96, 98, 100, 102, hoặc 104, và bất kỳ trình tự nucleotit nào mã hóa protein có sự tương đồng lệch khỏi các giá trị ở trên có thể được bao gồm mà không giới hạn miễn là vi sinh vật sản sinh các axit amin dạng mycosporin. Cụ thể, gen sinh tổng hợp các axit amin dạng mycosporin có thể bao gồm, nhưng không bị giới hạn, trình tự nucleotit được biểu thị bởi các SEQ ID NO: 1, 3, 85, 84, 89, 91, 93, 95, 97, 99, 101, hoặc 103.

Ngoài ra, rõ ràng là, vì trình tự axit amin có sự tương đồng với trình tự axit amin ở trên, trình tự axit amin với xóa, sửa đổi, thay thế, hoặc thêm một phần của trình tự cũng nằm trong phạm vi của sáng chế miễn là trình tự axit amin có hoạt tính sinh học thực chất giống hệt nhau hoặc tương đương với hoạt tính sinh học của protein chứa trình tự axit amin được biểu thị bởi SEQ ID NO được mô tả ở trên.

Ngoài ra, xem xét các đơn vị mã hóa được ưu tiên bởi các sinh vật để biểu hiện protein do đơn vị mã hóa thoái hóa, các sửa đổi khác nhau có thể được thực hiện trên vùng mã hóa nằm trong phạm vi mà không thay đổi trình tự axit amin của polypeptit. Theo đó, gen sinh tổng hợp các axit amin dạng mycosporin có thể bao gồm nhưng không giới hạn bất kỳ trình tự nucleotit nào mã hóa protein tham gia trong quá trình

tổng hợp các axit amin dạng mycosporin.

Ngoài ra, đầu dò có thể được điều chế từ các trình tự gen đã biết, ví dụ, bất kỳ trình tự nào lai với trình tự bổ sung với toàn bộ hoặc một phần trình tự nucleotit trong các điều kiện nghiêm ngặt để mã hóa protein tham gia vào quá trình sinh tổng hợp mycosporin-như là các axit amin có thể được bao gồm nhưng không giới hạn ở đó.

Theo phương án, vi sinh vật để sản sinh các axit amin dạng mycosporin có thể chứa gen sinh tổng hợp các axit amin dạng mycosporin có nguồn gốc khác nhau.

Theo sáng chế, sự bất hoạt protein, tăng cường hoạt tính protein và/hoặc đưa gen ở đây có thể được thực hiện đồng thời, tuần tự, hoặc theo thứ tự ngược lại.

Như được sử dụng ở đây, thuật ngữ “vi sinh vật để sản sinh các axit amin dạng mycosporin” sở hữu gen gen nội sinh và/hoặc ngoại sinh được đưa vào sinh tổng hợp các axit amin dạng mycosporin, và do đó có thể sản xuất các axit amin dạng mycosporin, và ngoài ra, nó có thể là vi sinh vật trong đó có khả năng sản sinh các axit amin dạng mycosporin được tăng lên bằng cách bất hoạt hoạt động 3-dehydroquinate dehydratase nội sinh. Sự đưa vào gen sinh tổng hợp các axit amin dạng mycosporin và sự bất hoạt 3-dehydroquinate dehydratase có thể được thực hiện đồng thời, tuần tự, hoặc theo thứ tự ngược lại.

Ngoài ra, vi sinh vật của sáng chế có thể là vi sinh vật có nguồn gốc tự nhiên có gen sinh tổng hợp các axit amin dạng mycosporin; và vi sinh vật trong đó gen dị hợp và/hoặc tương đồng sinh tổng hợp mycosporins-dạng axit amin đã được đưa vào, nhưng sáng chế không giới hạn ở đây.

Ngoài ra, vi sinh vật của sáng chế có thể là vi sinh vật có hoạt động enzym được tăng cường được mã hóa bởi gen nội sinh và/hoặc được đưa vào liên quan đến sinh tổng hợp các axit amin dạng mycosporin, nhưng sáng chế không giới hạn ở đây.

Ngoài ra, vi sinh vật của sáng chế có thể không có giới hạn miễn là vi sinh vật có hoạt động 3-dehydroquinate dehydratase trước biến đổi, và cụ thể, nó có thể là vi sinh vật thuộc chi *Corynebacterium*, vi sinh vật thuộc chi *Escherichia*, hoặc nấm men.

Vi sinh vật thuộc chi *Corynebacterium* cụ thể có thể là *Corynebacterium glutamicum*, *Corynebacterium ammoniagenes*, *Brevibacterium lactofermentum*, *Brevibacterium flavum*, *Corynebacterium thermoaminogenes*, *Corynebacterium*

*efficiens*, v.v., và cụ thể hơn, vi sinh vật có thể là *Corynebacterium glutamicum*, nhưng sáng chế không giới hạn ở đây.

Vi sinh vật thuộc chi *Escherichia* cụ thể có thể là *Escherichia albertii*, *Escherichia coli*, *Escherichia fergusonii*, *Escherichia hermannii*, *Escherichia vulneris*, v.v., và cụ thể hơn, vi sinh vật có thể là *Escherichia coli*, nhưng sáng chế không giới hạn ở đây.

Nấm men cụ thể có thể là vi sinh vật thuộc Saccharomycotina, Taphrinomycotina của Ascomycota, hoặc Agaricomycotina, Pucciniomycotina của Basidiomycota, v.v., và cụ thể hơn, nấm men có thể là vi sinh vật thuộc chi *Saccharomyces*, vi sinh vật thuộc chi *Schizosaccharomyces*, vi sinh vật thuộc chi *Phaffia*, vi sinh vật thuộc chi *Kluyveromyces*, vi sinh vật thuộc chi *Pichia*, vi sinh vật thuộc chi *Candida*, và thậm chí cụ thể hơn, nấm men có thể là *Saccharomyces cerevisiae*, nhưng sáng chế không giới hạn ở đây.

Trong sáng chế, nấm men để sản sinh các axit amin dạng mycosporin có thể được đưa vào với gen mã hóa 3-dehydroquinat synthaza hoặc có thể có hoạt tính 3-dehydroquinat synthaza được tăng cường. Ví dụ, nếu một phần hoặc toàn bộ *ARO1* bị xóa để bất hoạt hoạt động 3-dehydroquinat synthaza trong nấm men, chức năng của 3-dehydroquinat synthaza bị mất, và do đó nó có thể khó khăn để tổng hợp 3-DHQ. Do đó, khi một phần hoặc toàn bộ *ARO1* trong nấm men bị xóa, gen mã hóa 3-dehydroquinat synthaza (ví dụ, gen *aroB*) có thể được đưa vào, nhưng sáng chế không giới hạn ở đây.

Khía cạnh khác của sáng chế đề xuất phương pháp để sản xuất các axit amin dạng mycosporin, bao gồm:

nuôi cấy vi sinh vật của sáng chế; và

thu hồi các axit amin dạng mycosporin từ vi sinh vật được nuôi cấy hoặc từ môi trường đã nuôi cấy vi sinh vật.

“Vi sinh vật” và “các axit amin dạng mycosporin” được mô tả như ở trên.

Như được sử dụng ở đây, thuật ngữ “nuôi cấy” đề cập đến phát triển vi sinh vật trong môi trường được điều chỉnh thích hợp. Quá trình nuôi cấy của sáng chế có thể đạt được theo các điều kiện môi trường và nuôi cấy đã được biết trong kỹ thuật. Quá

trình nuôi cấy có thể được điều chỉnh dễ dàng để sử dụng bởi những người có hiểu biết trung bình trong cùng lĩnh vực kỹ thuật theo đó vi sinh vật được chọn. Bước nuôi cấy vi sinh vật có thể được thực hiện bằng phương pháp nuôi cấy theo mẻ đã biết, phương pháp nuôi cấy liên tục, phương pháp theo mẻ-bổ sung cơ chất, v.v., nhưng không bị giới hạn cụ thể ở đây. Môi trường được sử dụng để nuôi cấy vi sinh vật của sáng chế và các điều kiện nuôi cấy khác không bị giới hạn cụ thể, nhưng bất kỳ môi trường nào được sử dụng để nuôi cấy vi sinh vật thông thường có thể được sử dụng. Cụ thể, vi sinh vật của sáng chế có thể được nuôi cấy trong môi trường thông thường chứa các nguồn các bon phù hợp, nguồn nitơ, nguồn photpho, hợp chất vô cơ, axit amin và/hoặc các vitamin, v.v., trong điều kiện hiếu khí trong khi điều chỉnh nhiệt độ, độ pH, v.v. Cụ thể, độ pH nuôi cấy có thể được điều chỉnh thành độ pH phù hợp (ví dụ, độ pH 5 đến 9, cụ thể là độ pH 6 đến 8, và cụ thể nhất là độ pH là 6,8) bằng cách sử dụng hợp chất bazơ (ví dụ, natri hydroxit, kali hydroxit, hoặc amoniac) hoặc hợp chất axit (ví dụ, axit phosphoric hoặc axit sunfuric), nhưng sáng chế không giới hạn ở đây. Ngoài ra, oxy hoặc khí có chứa oxy có thể được đưa vào nuôi cấy để duy trì trạng thái nuôi cấy; hoặc khí nitơ, khí hydro hoặc khí cacbon đioxit, hoặc không có khí có thể được bơm vào để duy trì tình trạng hiếu khí hoặc vi hiếu khí, nhưng sáng chế không giới hạn ở đây. Ngoài ra, nhiệt độ nuôi cấy có thể được duy trì ở 20 đến 45°C, cụ thể ở 25 đến 40°C, và nuôi cấy có thể được thực hiện trong khoảng 10 đến 160 giờ, nhưng nuôi cấy không bị giới hạn ở điều kiện này. Ngoài ra, sự tạo thành bọt có thể được ngăn chặn trong suốt quá trình nuôi cấy bằng cách sử dụng chất chống bọt như là este polyglycol axit béo, v.v., nhưng sáng chế không giới hạn ở đây.

Ngoài ra, nguồn cacbon cho môi trường nuôi cấy được sử dụng, đường và cacbohydrat (ví dụ, glucoza, sucroza, lactoza, frutoza, maltoza, ri đường, tinh bột, và xenluloza), dầu và chất béo (ví dụ, dầu đậu nành, dầu hướng dương, dầu đậu phộng, và dầu dừa), các axit béo (ví dụ, axit palmitic, axit stearic và axit linoleic), rượu (ví dụ, glyxerol và etanol), axit hữu cơ (ví dụ, axit axetic), v.v. có thể được sử dụng riêng lẻ hoặc kết hợp, nhưng sáng chế không giới hạn ở đây. Nguồn nitơ, các hợp chất hữu cơ chứa nitơ (ví dụ, pepton, dịch chiết nấm men, nước thịt, dịch chiết mạch nha, rượu ngô, bột đậu nành, và urê) hoặc các hợp chất vô cơ (ví dụ, amoniasunfat, amoni clorua, amoni phosphat, amoni cacbonat và amoni nitrat), v.v. có thể được sử dụng riêng lẻ hoặc kết hợp, nhưng sáng chế không giới hạn ở đây. Nguồn phospho, kali dihydrogen

phosphat, đikali hydrogên phosphat, các muối chứa natri tương ứng của chúng, v.v. có thể được sử dụng riêng lẻ hoặc kết hợp, nhưng sáng chế không giới hạn ở đây. Hơn nữa, các nguyên liệu thúc đẩy tăng trưởng thiết yếu như là các muối kim loại khác (ví dụ, magiê sulfat hoặc sắt sulfat), các axit amin, các vitamin, v.v. có thể được chứa trong môi trường.

MAAs được sản xuất bằng cách nuôi cấy có thể được tiết vào trong môi trường hoặc có thể được duy trì trong các tế bào.

Như được sử dụng ở đây, thuật ngữ “môi trường” đề cập đến sản phẩm thu được sau khi nuôi cấy vi sinh vật của sáng chế. Môi trường là khái niệm bao gồm cả dạng chứa vi sinh vật và dạng trong đó vi sinh vật được loại bỏ bằng ly tâm, lọc v.v., từ dịch nuôi cấy chứa vi sinh vật.

Bước thu hồi MAAs được sản xuất trong bước nuôi cấy của sáng chế, MAAs đích có thể được thu từ dịch nuôi cấy bằng cách sử dụng phương pháp thích hợp được biết trong kỹ thuật. Ví dụ, ly tâm, lọc, sắc ký trao đổi ion, kết tinh, HPLC-High Performance Liquid Chromatography (sắc ký lỏng hiệu năng cao), v.v. có thể được sử dụng, và MAAs đích có thể được thu từ vi sinh vật được nuôi cấy hoặc môi trường bằng cách sử dụng phương pháp thích hợp đã được biết trong kỹ thuật. Ngoài ra, bước thu hồi MAAs có thể còn bao gồm quá trình tách chiết và/hoặc bước tinh sạch.

### **Ví dụ thực hiện sáng chế**

Sáng chế sẽ được mô tả chi tiết hơn bằng các ví dụ. Tuy nhiên, các ví dụ này được đưa ra chỉ nhằm mục đích minh họa, và phạm vi sáng chế không nhằm mục đích giới hạn hoặc bởi các ví dụ này.

Xây dựng các vi sinh vật tái tổ hợp sản xuất MAAs có nguồn gốc từ *E.coli* và sản xuất MAAs cũng sử dụng vi sinh vật tái tổ hợp này.

Ví dụ 1: Xây dựng các vector biểu hiện quá mức gen sinh tổng hợp shinorin có nguồn gốc từ vi tảo.

Cụm gen sinh tổng hợp shinorin có nguồn gốc từ *A.variabilis* chứa bốn gen, đó là, 2-đemetyl 4-đeoxygadusol synthaza, O-metyltransferaza, C-N ligaza, và non-ribosomal peptit synthetaza, và *Nostoc punctiforme*, một số loài của vi khuẩn lam cũng có thể sản xuất shinorin bằng cách sử dụng các gen này. Cụm gen sinh tổng hợp

shinorin được xác định bằng cách sử dụng bộ gen DNA của *A. variabilis* ATCC29413 và *N. punctiforme* ATCC29133. Bốn vector, mỗi vector chứa các gen sinh tổng hợp shinorin (Ava\_ABCD và Npr\_ABCD) có nguồn gốc từ *A. variabilis* ATCC29413 và *N. punctiforme* ATCC29133, tương ứng, được cấu tạo bằng cách sử dụng hai loại vector pECCG 117\_Ptrc\_GFP\_terminator và pECCG 117\_Pcj1\_GFP\_terminator. Tên của bốn vector biểu hiện gen sinh tổng hợp shinorin, các mạch khuôn và các môi được sử dụng để điều chế các vector được tổng hợp trong Bảng 1 bên dưới.

Bảng 1

| Tên của các vector     | Các mạch khuôn được sử dụng                    | Các môi được sử dụng     |
|------------------------|------------------------------------------------|--------------------------|
| pECCG117_Ptrc_Ava_ABCD | Bộ gen DNA của <i>A. variabilis</i> ATCC29413  | SEQ ID NO: 5 (môi xuôi)  |
| pECCG117_Pcj1_Ava_ABCD |                                                | SEQ ID NO: 6 (môi ngược) |
| pECCG117_Ptrc_Npr_ABCD | Bộ gen DNA của <i>N. punctiforme</i> ATCC29133 | SEQ ID NO: 7 (môi xuôi)  |
| pECCG117_Pcj1_Npr_ABCD |                                                | SEQ ID NO: 8 (môi ngược) |

Sau khi các đoạn gen thu được bằng cách sử dụng các mạch khuôn và các cặp môi, mỗi đoạn gen được nối với các vector pECCG 117\_Ptrc\_GFP\_terminator và pECCG 117\_Pcj1\_GFP\_terminator đã được xử lý với enzym giới hạn EcoRV/XbaI bằng cách sử dụng bộ In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.). Các vector biểu hiện được định danh là pECCG117\_Ptrc\_Ava\_ABCD, pECCG117\_Pcj1\_Ava\_ABCD, pECCG117\_Ptrc\_Npr\_ABCD, và pECCG117\_Pcj1\_Npr\_ABCD, tương ứng, và mỗi vector biểu hiện được xác nhận bằng cách giải trình tự. Các trình tự nucleotit và các trình tự axit amin của Ava\_ABCD và Npr\_ABCD được biểu thị cụ thể bởi SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, và SEQ ID NO: 4, tương ứng.

Ví dụ 2: Đánh giá khả năng sản sinh Shinorin của các chủng được đưa vào các vector biểu hiện gen sinh tổng hợp Shinorin.

Để xác nhận khả năng sản sinh MAA trong *E. coli*, bốn plasmid được điều chế trong ví dụ 1 được đưa vào trong chủng W3110, *E. coli* kiểu dại được điều chế các chủng với sinh tổng hợp shinorin được tăng cường. Các chủng được điều chế như vậy được cấy trải vào trong môi trường LB đặc chứa kanamycin và sau đó được nuôi cấy qua đêm trong tủ ấm ở 37°C. Lấy một vòng que cấy bạch kim của mỗi chủng được

nuôi cấy qua đêm trong môi trường LB đặc được cấy chuyển vào trong 25 ml môi trường chuẩn độ (thành phần môi trường: 40 g/L glucoza, 0,3 g/L  $\text{KH}_2\text{PO}_4$ , 0,6 g/L  $\text{K}_2\text{HPO}_4$ , 15 g/L  $(\text{NH}_4)_2\text{SO}_4$ , 1 g/L  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 2,5 g/L NaCl, 1,2 g/L natri xitrat, 2,5 g/L dịch chiết nấm men, 40 g/L canxi cacbonat, độ pH 7,0), và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong Bảng 2 bên dưới:

Bảng 2

| Tên các chủng                 | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|-------------------------------|-------------|-------------------------|
| W3110                         | 22,3        | -                       |
| W3110/ pECCG117_Ptrc_Ava_ABCD | 20,1        | 121                     |
| W3110/ pECCG117_PCJ1_Ava_ABCD | 19,8        | 382                     |
| W3110/ pECCG117_Ptrc_Npr_ABCD | 21,0        | 96                      |
| W3110/ pECCG117_PCJ1_Npr_ABCD | 20,2        | 332                     |

Như thể hiện trong Bảng 2 bên trên, khi các gen sinh tổng hợp shinorin được đưa vào trong W3110, được xác nhận rằng sản xuất shinorin là có thể. Ngoài ra, đã được xác nhận rằng sản xuất shinorin được tăng lên bằng cách tăng mật độ của các vùng gen khởi động (vùng gen khởi động PCJ1 được đưa vào) thông qua việc tăng cường con đường sinh tổng hợp.

Ví dụ 3: Xây dựng các chủng trong đó 3-đehydroquinat dehydrataza bị bất hoạt.

Ava-A, gen đầu tiên để sinh tổng hợp MAAs trong vi tảo, chia sẻ và sử dụng như các cơ chất DHQ (3-đehydroquinat) trong con đường shikimat và SH-7P (sedoheptulos 7-phosphat) trong con đường petoza phosphat. Để điều chế các chủng trong đó 3-đehydroquinat dehydrataza bị bất hoạt bằng cách xóa gen *aroD*, phương pháp tái tổ hợp tương đồng bằng cách sử dụng lambda red recombinaza. Như là dấu hiệu chèn gen, gen kháng-chloramphenicol của pKD3 đã được sử dụng, và catxet *aroD* bị xóa chứa một phần của gen *aroD* và gen kháng chloramphenicol của plasmid pKD3 đã được xây dựng bằng PCR sử dụng cặp mồi được biểu thị bởi SEQ ID NO: 9 (mồi xuôi) và SEQ ID NO: 10 (mồi ngược). Sau khi điều chế các chủng (*E.coli* kiểu đại W3110) trong đó gen *aroD* bị xóa (SEQ ID NO: 71 và SEQ ID NO: 72), plasmid pKD46 chứa gen lambda red recombinaza được biến nạp vào trong các chủng, sau đó, biểu hiện gen được tạo ra bằng cách sử dụng arabinoza để điều chế các tế bào khả nạp. Catxet *aroD* bị xóa được đưa vào trong các tế bào khả nạp bằng cách xung điện, và

sau đó các tế bào được cấy trải trên môi trường LB đặc chứa 30 mg/L chloramphenicol. Các chủng thu được như vậy trải qua PCR bằng cách sử dụng cặp mồi được biểu thị bởi SEQ ID NO: 11 (mồi xuôi) và SEQ ID NO: 12 (mồi ngược), và gen *aroD* xóa được xác nhận bằng cách quan sát các đoạn khuếch đại 1300 bp.

Ví dụ 4: Đánh giá khả năng sản sinh Shinorin của các chủng trong đó 3-dehydroquinate dehydratase bị bất hoạt.

Hai plasmid có biểu hiện được kiểm soát bởi vùng gen khởi động PCJ1 trong số bốn plasmid được điều chế trong Ví dụ 1, được đưa vào trong các chủng, trong đó gen *aroD* bị xóa được điều chế trong (W3110  $\Delta$  *aroD*/pECCG117\_PCJ1\_Ava\_ABCD và W3110  $\Delta$  *aroD* /pECCG117\_PCJ1\_Npr\_ABCD), sau đó các chủng được cấy trải trên môi trường LB đặc chứa kanamycin. Tiếp theo, các chủng *aroD* bị xóa và các chủng không chứa *aroD* bị xóa được nuôi cấy tương ứng qua đêm trong tủ ấm ở 37°C, và lấy một vòng que cấy bạch kim của mỗi chủng được cấy vào trong 25 ml môi trường chuẩn độ [thành phần môi trường: 40 g/L glucoza, 0,3 g/L KH<sub>2</sub>PO<sub>4</sub>, 0,6 g/L K<sub>2</sub>HPO<sub>4</sub>, 15 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 2,5 g/L NaCl, 1,2 g/L natri xitrat, 2,5 g/L dịch chiết nấm men, 40 g/L canxi cacbonat, độ pH 7,0], và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong bảng 3 bên dưới:

Bảng 3

| Tên các chủng                                          | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|--------------------------------------------------------|-------------|-------------------------|
| W3110/ pECCG117_PCJ1_Ava_ABCD                          | 20,3        | 352                     |
| W3110 $\Delta$ <i>aroD</i> /<br>pECCG117_PCJ1_Ava_ABCD | 18,7        | 683                     |
| W3110/ pECCG117_PCJ1_Npr_ABCD                          | 18,9        | 331                     |
| W3110 $\Delta$ <i>aroD</i> /<br>pECCG117_PCJ1_Npr_ABCD | 17,9        | 601                     |

Như thể hiện trong bảng 3, nồng độ shinorin được sản xuất trong các chủng *aroD* bị xóa được tăng lên 194 % và 182% so với nồng độ shinorin được sản xuất trong các chủng không có *aroD* bị xóa, tương ứng. Theo đó, chủng W3110  $\Delta$  *aroD*/pECCG117\_PCJ1\_Ava\_ABCD và chủng W3110  $\Delta$  *aroD*/pECCG117\_PCJ1\_Npr\_ABCD là các chủng có *aroD* bị xóa, được định danh là CB06-0017 và CB06-0018 và được ký gửi tại Bảo tàng giống chuẩn Vi sinh vật Hàn

Quốc theo Hiệp ước Budapest vào ngày 26 tháng 06 năm 2017, với số đăng ký KCCM12044P và KCCM12045P, tương ứng.

Ví dụ 5: Xây dựng các chủng có hoạt động 2-đehydro-3-đeoxyphosphoheptonat aldolaza/Phosphoenolpyruvat synthetaza/Transketolaza I/II được tăng cường.

Để tăng khả năng sản sinh MAA của các vi sinh vật sản sinh MAA, hoạt động của 2-đehydro-3-đeoxyphosphoheptonat aldolaza/phosphoenolpyruvat synthetaza/transketolaza I/II được tăng cường. Cụ thể, ba gen có nguồn gốc từ *E.coli* W3110, cụ thể, *aroG* (2-đehydro-3-đeoxyphosphoheptonat aldolaza; SEQ ID NO: 73 và SEQ ID NO: 74), *ppsA* (phosphoenolpyruvat synthetaza; SEQ ID NO: 75 và SEQ ID NO: 76), và *tktA* (transketolaza I/II; SEQ ID NO: 77 và SEQ ID NO: 78) được đưa vào thêm. Các plasmid pSKH130- $\Delta$  fhuA-Pn-*aroG*-Pn-*ppsA*-Pn-*tktA* được xây dựng để tăng cường các gen *aroG*, *ppsA*, và *tktA*. Mạch khuôn và cặp môi được sử dụng trong xây dựng các plasmid pSKH130- $\Delta$  fhuA-Pn-*aroG*-Pn-*ppsA*-Pn-*tktA* được thể hiện trong bảng 4 bên dưới.

Bảng 4

| Các sản phẩm PCR | Mạch khuôn được sử dụng | Cặp môi được sử dụng (môi xuôi, môi ngược) |
|------------------|-------------------------|--------------------------------------------|
| Pn- <i>aroG</i>  | Bộ gen DNA của W3110    | SEQ ID NO: 13, SEQ ID NO: 14               |
| Pn- <i>ppsA</i>  |                         | SEQ ID NO: 15, SEQ ID NO: 16               |
| Pn- <i>tktA</i>  |                         | SEQ ID NO: 17, SEQ ID NO: 18               |

Bằng cách sử dụng mạch khuôn và cặp môi bên trên, các đoạn gen *aroG*, *ppsA*, và *tktA* đã được khuếch đại bằng cách sử dụng PCR, sau đó lần lượt được đưa vào trong vectơ pSKH130- $\Delta$  fhuA đã được cắt với enzym giới hạn BamH1-Pst1. Giải trình tự được áp dụng để xác nhận nhân dòng và các trình tự gen của các vectơ, sau đó các vectơ được biến nạp vào trong *E.coli* W3110 kiểu dại và *E. coli* W3110  $\Delta$  *aroD* bị xóa *aroD* bằng xung điện. Các gen được biến nạp được đưa vào trong nhiễm sắc thể bằng tái tổ hợp sơ cấp (trao đổi chéo), tiếp theo là cắt bỏ vùng plasmid khỏi nhiễm sắc thể bằng tái tổ hợp thứ cấp (trao đổi chéo). Việc đưa vào của các gen *aroG*, *ppsA*, và *tktA* được xác nhận trong các chủng *E.coli* được biến nạp, trong đó tái tổ hợp thứ cấp được hoàn thành, bằng cách sử dụng cặp môi được biểu thị bởi SEQ ID NO: 19 (môi xuôi) và SEQ ID NO: 20 (môi ngược).

Ví dụ 6: Đánh giá khả năng sản sinh Shinorin của các chủng có hoạt động 2-dehydro-3-deoxyphosphoheptonat aldolaza/Phosphoenolpyruvat synthetaza/Transketolaza I/II được tăng cường.

Hai plasmid có biểu hiện được kiểm soát bởi vùng gen khởi động PCJ1 trong số bốn plasmid được điều chế trong ví dụ 5 được lần lượt đưa vào trong các chủng với các gen *aroG*, *ppsA*, và *tktA* được điều chế trong ví dụ 5, sau đó các chủng được cấy trải trên môi trường LB đặc. Tiếp theo, các chủng được nuôi cấy qua đêm trong tủ ấm ở 37°C, và lấy một vòng que cấy bạch kim của mỗi chủng được cấy chuyển vào trong 25 ml môi trường chuẩn độ của Ví dụ 4, và được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong Bảng 5 bên dưới.

Bảng 5

| Tên các chủng                                                                                                                     | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|-----------------------------------------------------------------------------------------------------------------------------------|-------------|-------------------------|
| W3110 / pECCG117_PCJ1_Ava_ABCD                                                                                                    | 19,8        | 352                     |
| W3110 / pECCG117_PCJ1_Npr_ABCD                                                                                                    | 19,6        | 344                     |
| W3110 $\Delta$ <i>aroD</i> / pECCG117_PCJ1_Ava_ABCD                                                                               | 17,3        | 688                     |
| W3110 $\Delta$ <i>aroD</i> / pECCG117_PCJ1_Npr_ABCD                                                                               | 17,8        | 652                     |
| W3110 $\Delta$ <i>fhuA</i> ::Pn- <i>aroG</i> -Pn- <i>ppsA</i> -pn- <i>tktA</i> / pECCG117_PCJ1_Ava_ABCD                           | 18,9        | 1.163                   |
| W3110 $\Delta$ <i>fhuA</i> ::Pn- <i>aroG</i> -Pn- <i>ppsA</i> -pn- <i>tktA</i> / pECCG117_PCJ1_Npr_ABCD                           | 18,6        | 989                     |
| W3110 $\Delta$ <i>aroD <math>\Delta</math> <i>fhuA</i>::Pn-<i>aroG</i>-Pn-<i>ppsA</i>-pn-<i>tktA</i> / pECCG117_PCJ1_Ava_ABCD</i> | 17,3        | 1.928                   |
| W3110 $\Delta$ <i>aroD <math>\Delta</math> <i>fhuA</i>::Pn-<i>aroG</i>-Pn-<i>ppsA</i>-pn-<i>tktA</i> / pECCG117_PCJ1_Npr_ABCD</i> | 17,7        | 1.889                   |

Như thể hiện trong bảng 5, nồng độ shinorin được sản xuất trong các chủng, trong đó ba gen (*aroG*, *ppsA*, và *tktA*) được tăng cường, đã tăng lên khoảng 300% so với nhóm đối chứng.

Ví dụ 7: Xây dựng các vectơ được chèn *ava*\_ABCD vào nhiễm sắc thể và các chủng.

Để đưa các gen sinh tổng hợp shinorin vào trong *E.coli*, các plasmid pSKH130  $\Delta$  *pinR*::Ava-ABCD đã được điều chế. Ava\_ABCD trải qua PCR bằng cách sử dụng

cặp mồi được biểu thị bởi SEQ ID NO: 21 (mồi xuôi) và SEQ ID NO: 22 (mồi ngược) dựa trên pECCG117\_Ptrc\_Ava\_ABCD làm mạch khuôn. Khoảng 7kb đoạn PCR được nối vào vector pSKH130  $\Delta$  pinR được xử lý với các enzym giới hạn BamHI và PstI để điều chế pSKH130  $\Delta$  pinR::Ava\_ABCD bằng cách sử dụng bộ In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.). Tiếp theo, các đoạn vùng gen khởi động Ptrc và PCJ1 trải qua PCR bằng cách sử dụng cặp mồi xuôi và mồi ngược được biểu thị bởi SEQ ID NO: 23 và SEQ ID NO: 24, SEQ ID NO: 25 và SEQ ID NO: 26, và SEQ ID NO: 25 và SEQ ID NO: 27, tương ứng, để kiểm soát biểu hiện của Ava-ABCD. Sau đó, các đoạn này được nối với vector pSKH130  $\Delta$  pinR::Ava\_ABCD đã được xử lý với enzym giới hạn ScaI để điều chế pSKH130  $\Delta$  pinR::Ptrc-Ava-ABCD và pSKH130  $\Delta$  pinR::PCJ1-Ava-ABCD bằng cách sử dụng bộ In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.). Các plasmid tái tổ hợp được biến nạp vào trong các chủng W3110  $\Delta$  aroD  $\Delta$  fhuA::Pn-aroG-Pn-ppsA-Pn-tktA được điều chế trong Ví dụ 5 bằng cách xung điện, và các chủng được đưa vào trong nhiễm sắc thể bằng tái tổ hợp sơ cấp (trao đổi chéo), tiếp theo cắt bỏ vùng vector ngoại trừ gen đích ra khỏi nhiễm sắc thể bằng tái tổ hợp thứ cấp (trao đổi chéo).

Việc đưa vào của gen Ava\_ABCD được xác nhận trong các *E.coli* được biến nạp, trong đó tái tổ hợp thứ cấp đã được hoàn thành, bằng PCR sử dụng cặp mồi được biểu thị bởi SEQ ID NO: 28 (mồi xuôi) và SEQ ID NO: 29 (mồi ngược).

Ví dụ 8: Đánh giá khả năng sản sinh Shinorin của các chủng có nhiễm sắc thể được chèn ava\_ABCD vào.

Các chủng được điều chế trong ví dụ 7 được cấy trên môi trường LB đặc và sau đó được nuôi cấy qua đêm trong tủ ấm ở 37°C. Tiếp theo, lấy một vòng que cấy bạch kim của mỗi chủng được cấy chuyển vào trong 25 ml môi trường chuẩn độ của ví dụ 4, và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong bảng 6 bên dưới.

Bảng 6

| Tên các chủng                                              | OD (600nm) | Nồng độ Shinorin (mg/L) |
|------------------------------------------------------------|------------|-------------------------|
| W3110 $\Delta$ aroD $\Delta$ fhuA::Pn-aroG-Pn-ppsA-Pn-tktA | 18,5       | -                       |
| W3110 $\Delta$ aroD $\Delta$ fhuA::Pn-aroG-Pn-             | 17,8       | 1928                    |

|                                                                                                                                                                   |      |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-----|
| <i>ppsA</i> -Pn- <i>tklA</i> /<br>pECCG117_PCJ1_Ava_ABCD                                                                                                          |      |     |
| W3110 $\triangle$ <i>aroD</i> $\triangle$ <i>fhuA</i> ::Pn- <i>aroG</i> -Pn-<br><i>ppsA</i> -Pn- <i>tklA</i> $\triangle$ <i>pinR</i> ::P <sub>trc</sub> -Ava-ABCD | 18,2 | 483 |
| W3110 $\triangle$ <i>aroD</i> $\triangle$ <i>fhuA</i> ::Pn- <i>aroG</i> -Pn-<br><i>ppsA</i> -Pn- <i>tklA</i> $\triangle$ <i>pinR</i> ::PCJ1-Ava-ABCD              | 17,9 | 832 |

Như thể hiện trong bảng 6, khi Ava-ABCD được đưa vào trong nhiễm sắc thể, đã xác nhận rằng shionorin đã được sản xuất, và nồng độ của chúng được tăng lên theo mật độ của các vùng gen khởi động. Tuy nhiên, cũng đã xác nhận rằng sản xuất shionorin bị giảm so với các chủng có sinh tổng hợp shionorin được tăng cường thông qua các plasmid. Khi plasmid pECCG117\_PCJ1\_Ava\_ABCD được đưa vào thêm trong các chủng được đưa vào với Ava-ABCD trên nhiễm sắc thể, sản xuất shionorin được tăng lên khoảng 353% và 152% so với chủng chỉ được đưa vào với Ava-ABCD trên nhiễm sắc thể (dựa trên vùng gen khởi động CJ1) và chủng chỉ được đưa vào với plasmid, tương ứng.

Ví dụ 9: Xây dựng các vector biểu hiện quá mức gen MAA và đánh giá khả năng sản sinh MAA của chúng.

4-deoxygadusol và mycosporin-glyxin là những sản phẩm trung gian được sản xuất trong quá trình sinh tổng hợp shionorin và là các axit amin dạng mycosporin có tác dụng ngăn ngừa các tia UV cùng một lúc. Các vector được xây dựng để xác nhận có hay không những chất này có thể được sản xuất trong các chủng bị xóa *AroD* của *E.coli*. Các kết quả được thể hiện trong bảng 7 bên dưới.

P<sub>trc</sub>\_Ava\_AB và P<sub>trc</sub>\_Ava\_ABC trải qua PCR bằng cách sử dụng các cặp mồi được biểu thị bằng SEQ ID NO: 30 và SEQ ID NO: 31 và SEQ ID NO: 30 và SEQ ID NO: 32 dựa trên pECCG117\_P<sub>trc</sub>\_Ava\_ABCD như là mạch khuôn. pECCG117\_P<sub>trc</sub>\_Ava\_AB và pECCG117\_P<sub>trc</sub>\_Ava\_ABC được điều chế bằng cách nối các đoạn PCR với vector pECCG117 P<sub>trc</sub> GFP được xử lý với các enzym giới hạn BamHI và SpeI. Theo cách tương tự, pECCG117\_PCJ1\_Ava\_AB và pECCG117\_PCJ1\_Ava\_ABC được điều chế bằng cách nối các đoạn PCR thu được bằng cách sử dụng cặp mồi được biểu thị bởi SEQ ID NO: 30 và SEQ ID NO: 31 và SEQ ID NO: 30 và SEQ ID NO: 32 dựa trên pECCG117\_PCJ1\_Ava\_ABCD như là mạch khuôn với vector pECCG117 P<sub>trc</sub> GFP được xử lý với các enzym giới hạn

BamHI và SpeI. Trình tự nucleotit và trình tự axit amin của Ava\_AB và Ava\_ABC được biểu diễn cụ thể bởi SEQ ID NO: 85 đến SEQ ID NO: 88.

Bảng 7

| Tên các vector        | Các mạch khuôn được sử dụng | Các môi được sử dụng      |
|-----------------------|-----------------------------|---------------------------|
| pECCG117_Ptrc_Ava_AB  | pECCG117_Ptrc_Ava_ABCD      | SEQ ID NO: 30 (Môi xuôi)  |
| pECCG117_PCJ1_Ava_AB  |                             | SEQ ID NO: 31 (Môi ngược) |
| pECCG117_Ptrc_Ava_ABC | pECCG117_PCJ1_Ava_ABCD      | SEQ ID NO: 30 (Môi xuôi)  |
| pECCG117_PCJ1_Ava_ABC |                             | SEQ ID NO: 32 (Môi ngược) |

Các vector được điều chế được biến nạp vào trong các chủng W3110  $\Delta$  *aroD*  $\Delta$  *fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA* được sử dụng trong ví dụ 8 bằng cách sử dụng phương pháp xung điện thông thường, và mỗi chủng được cấy trải trên môi trường LB đặc và được nuôi cấy qua đêm trong tủ ấm ở 37°C. Các chủng được nuôi cấy qua đêm trong môi trường LB đặc được cấy chuyển vào trong 25 ml môi trường chuẩn độ của Ví dụ 4, sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Sau khi hoàn thành nuôi cấy, sản xuất MAAs được đo bằng sắc ký lỏng hiệu năng cao, và nồng độ của MAAs trong nuôi cấy của mỗi chủng đã kiểm tra được thể hiện trong bảng 8 bên dưới.

Bảng 8

| Tên các chủng                                                                                    | OD (600 nm) | Nồng độ 4-Deoxygadusol (mg/L) | Nồng độ Mycosporin glyxin (mg/L) |
|--------------------------------------------------------------------------------------------------|-------------|-------------------------------|----------------------------------|
| W3110 $\Delta$ <i>aroD</i> $\Delta$ <i>fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA</i>                         | 19,2        | -                             | -                                |
| W3110 $\Delta$ <i>aroD</i> $\Delta$ <i>fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA</i> / pECCG117_Ptrc_Ava_AB  | 18,2        | 12,0                          | -                                |
| W3110 $\Delta$ <i>aroD</i> $\Delta$ <i>fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA</i> / pECCG117_Pcjl_Ava_AB  | 17,6        | 25,3                          | -                                |
| W3110 $\Delta$ <i>aroD</i> $\Delta$ <i>fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA</i> / pECCG117_Ptrc_Ava_ABC | 18,7        | 2,0                           | 9,3                              |
| W3110 $\Delta$ <i>aroD</i> $\Delta$ <i>fhuA::Pn-aroG-Pn-ppsA-Pn-tpkA</i> / pECCG117_Pcjl_Ava_ABC | 18,1        | 2,9                           | 19,7                             |

Như thể hiện trong bảng 8, khi các gen Ava\_AB và Ava\_ABC được đưa vào, đã

xác nhận rằng 4-đeoxygadusol và mycosporin glyxin được sản xuất, và hàm lượng của chúng đã tăng lên khi mật độ của các vùng gen khởi động được tăng cường.

Xây dựng các vi sinh vật tái tổ hợp sản xuất MAA có nguồn gốc từ *Corynebacterium glutamicum* và sản xuất MAAs sử dụng vi sinh vật đó

Ví dụ 10: Đánh giá khả năng sản sinh shinorin của các chủng được đưa vào với các vector biểu hiện quá mức gen sinh tổng hợp shinorin.

Để xác nhận khả năng sản sinh MAA của *Corynebacterium glutamicum*, bốn plasmid được điều chế trong ví dụ 1 được đưa vào trong chủng *Corynebacterium glutamicum* 13032 để điều chế các chủng có sinh tổng hợp shinorin được tăng cường, và các chủng được cấy trải trên môi trường BHIS đặc chứa kanamycin và được nuôi cấy qua đêm trong tủ ấm ở 30°C. Lấy một vòng que cấy bạch kim của mỗi chủng được nuôi cấy qua đêm trong môi trường BHIS đặc được cấy chuyển vào trong 25 ml của môi trường chuẩn độ [thành phần môi trường: 40 g/L glucoza, 1 g/L KH<sub>2</sub>PO<sub>4</sub>, 10 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 g/L MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g/L NaCl, 5 g/L dịch chiết nấm men, 30 g/L canxi cacbonat, pH 7,0], và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong Bảng 9 bên dưới.

Bảng 9

| Tên các chủng                      | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|------------------------------------|-------------|-------------------------|
| c.gl 13032                         | 72,1        | -                       |
| c.gl 13032/ pECCG117_Ptrc_Ava_ABCD | 71,5        | 132                     |
| c.gl 13032/ pECCG117_PCJ1_Ava_ABCD | 69,8        | 496                     |
| c.gl 13032/ pECCG117_Ptrc_Npr_ABCD | 70,9        | 103                     |
| c.gl 13032/ pECCG117_PCJ1_Npr_ABCD | 71,4        | 421                     |

Như thể hiện trong bảng 9, đã xác nhận rằng sản xuất shinorin là có thể khi các gen sinh tổng hợp shinorin được đưa vào trong *Corynebacterium glutamicum* 13032, và năng suất của chúng có thể được tăng lên 375% tùy thuộc vào mật độ của vùng gen khởi động.

Ví dụ 11: Xây dựng các vector được đưa vào với gen sinh tổng hợp shinorin trên nhiễm sắc thể và các chủng.

Để đưa các gen sinh tổng hợp shinorin vào trong *Corynebacterium glutamicum*, các plasmid pDC ΔN1021\_Ava\_ABCD đã được điều chế. Ava\_ABCD trải qua PCR

bằng cách sử dụng cặp môi được biểu thị bởi SEQ ID NO: 33 (môi xuôi) và SEQ ID NO: 34 (môi ngược) dựa trên pECCG117\_Ptrc\_Ava\_ABCD làm mạch khuôn. Đoạn PCR khoảng 7 kb được nối với vector pDC ΔN1021 đã được xử lý với enzym giới hạn NdeI để điều chế pDC ΔN1021\_Ava\_ABCD bằng cách sử dụng bộ In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.). Tiếp theo, các đoạn vùng gen khởi động CJ7, Lysc8, và O2 trải qua PCR bằng cách sử dụng các cặp môi xuôi và môi ngược được biểu thị bởi SEQ ID NO: 35 và SEQ ID NO: 36, SEQ ID NO: 37 và SEQ ID NO: 38, và SEQ ID NO: 39 và SEQ ID NO: 40, tương ứng để kiểm soát biểu hiện của Ava-ABCD. Sau đó, các đoạn được nối với vector pDC ΔN1021\_Ava\_ABCD đã được xử lý với enzym giới hạn NdeI để điều chế pDC ΔN1021\_Pcj7\_Ava\_ABCD, pDC ΔN1021\_Plysc8\_Ava\_ABCD và pDC ΔN1021\_PO2\_Ava\_ABCD bằng cách sử dụng bộ In-Fusion® HD Cloning Kit (Clontech Laboratories, Inc.).

Các plasmid tái tổ hợp được biến nạp vào trong *Corynebacterium glutamicum* 13032 bằng xung điện (van der Rest et al. 1999), và được đưa vào trong nhiễm sắc thể bằng tái tổ hợp sơ cấp (trao đổi chéo), sau đó cắt bỏ plasmid ra khỏi nhiễm sắc thể bằng tái tổ hợp thứ cấp (trao đổi chéo).

Việc đưa vào của gen Ava\_ABCD đã được xác nhận trong các chủng được biến nạp *Corynebacterium glutamicum*, trong đó tái tổ hợp thứ cấp đã được hoàn thành, bằng PCR sử dụng cặp môi đặc hiệu gen được biểu thị bởi SEQ ID NO: 33 (môi xuôi) và SEQ ID NO: 34 (môi ngược).

Ví dụ 12: Đánh giá khả năng sản sinh shinorin của các chủng được đưa vào với gen sinh tổng hợp shinorin trên nhiễm sắc thể.

Tất cả các chủng được cấy trải trên môi trường BHIS đặc và được nuôi cấy qua đêm trong tủ ấm ở 30°C để xác nhận khả năng sản sinh shinorin. Lấy một vòng que cấy bạch kim của mỗi chủng được nuôi cấy qua đêm trong môi trường BHIS đặc được cấy chuyển vào trong 25 ml môi trường chuẩn độ của ví dụ 11, và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong bảng 10 bên dưới.

Bảng 10

| Tên các chủng | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|---------------|-------------|-------------------------|
| c.gl 13032    | 70,2        | -                       |

|                                   |      |     |
|-----------------------------------|------|-----|
| c.gl 13032 ΔN1021_PCJ7_Ava_ABCD   | 76,1 | 36  |
| c.gl 13032 ΔN1021_Plysc8_Ava_ABCD | 79,8 | 75  |
| c.gl 13032 ΔN1021_PO2_Ava_ABCD    | 72,5 | 173 |

Như được thể hiện trong bảng 10, đã được xác nhận rằng, khi bản sao đơn của các gen sinh tổng hợp shinorin được đưa vào trong *Corynebacterium glutamicum* kiểu đại, shinorin có thể được sản xuất với số lượng từ 36 mg đến 173 mg.

Ví dụ 13: Xây dựng *Corynebacterium aroD* (3-đehydroquinat dehydrataza)-các vector bị xóa và các chủng.

Như đã đề cập trong ví dụ 3, các chủng bị xóa gen *aroD* được điều chế để xác nhận có sản xuất shinorin hay không có thể được tăng lên thông qua xóa *aroD* (3-đehydroquinat dehydrataza). Để điều chế các chủng bị xóa gen *aroD* (SEQ ID NO: 79 và SEQ ID NO: 80) ở vị trí cụ thể của *Corynebacterium glutamicum*, các plasmid pDC-Δ*aroD*, trong đó khung đọc mở của *aroD* bị xóa nội sinh, đã được xây dựng. Gen nội sinh bị xóa của pDC-Δ*aroD* được tạo ra bằng cách đưa vào các đoạn gen được tạo ra bằng trao đổi chéo-PCR sử dụng cặp mồi xuôi và mồi ngược được biểu thị bởi SEQ ID NO: 41 và SEQ ID NO: 42, và SEQ ID NO: 43 và SEQ ID NO: 44, dựa trên bộ gen DNA của *Corynebacterium glutamicum* ATCC 13032 làm mạch khuôn vào trong vector pDC. Các plasmid tái tổ hợp được biến nạp vào trong *Corynebacterium glutamicum* 13032 ΔN1021\_PO2\_Ava\_ABCD bằng xung điện (van der Rest và cộng sự. 1999), và các plasmid được đưa vào trong nhiễm sắc thể bằng cách tái tổ hợp sơ cấp (trao đổi chéo), sau đó cắt bỏ plasmid ra khỏi nhiễm sắc thể bằng tái tổ hợp thứ cấp (trao đổi chéo).

Việc xóa gen *aroDI* đã được xác nhận trong các chủng *Corynebacterium glutamicum* được biến nạp, trong đó tái tổ hợp thứ cấp đã được hoàn thành, bằng PCR sử dụng cặp mồi đặc hiệu gen được biểu thị bởi SEQ ID NO: 41 và SEQ ID NO: 44.

Ví dụ 14: Đánh giá của *aroD* (3-đehydroquinat dehydrataza) bị xóa

Các chủng được kỳ vọng có tích lũy DHQ do xóa 3-đehydroquinat dehydrataza trong chủng *Corynebacterium glutamicum* 13032 ΔN1021\_PO2\_Ava\_ABCD được cấy trên môi trường BHIS đặc và được nuôi cấy qua đêm trong tủ ẩm ở 30°C. Lấy một vòng que cấy bạch kim của mỗi chủng được nuôi cấy qua đêm trong môi trường BHIS đặc được cấy chuyển vào trong 25 ml môi trường chuẩn độ [thành phần môi

trường: 40 g/L glucoza, 1 g/L  $\text{KH}_2\text{PO}_4$ , 10 g/L  $(\text{NH}_4)_2\text{SO}_4$ , 5 g/L  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 5 g/L NaCl, 5 g/L dịch chiết nấm men, 30 g/L canxi cacbonat, độ pH 7,0], và sau đó được nuôi cấy trong tủ ấm ở 37°C ở tốc độ lắc 200 vòng/phút trong 48 giờ. Các kết quả được thể hiện trong bảng 11 bên dưới.

Bảng 11

| Tên các chủng                                                                                 | OD (600 nm) | Nồng độ Shinorin (mg/L) |
|-----------------------------------------------------------------------------------------------|-------------|-------------------------|
| c.gl 13032 $\Delta\text{N1021\_PO2\_Ava\_ABCD}$                                               | 71,3        | 182                     |
| c.gl 13032 $\Delta\text{N1021\_PO2\_Ava\_ABCD\_}\Delta\text{aroD}$                            | 74,1        | 435                     |
| c.gl 13032 $\Delta\text{N1021\_PO2\_Ava\_ABCD\_}\Delta\text{aroD}$<br>/pECCG117_PCJ1_Ava_ABCD | 73,2        | 1162                    |

Như thể hiện trong bảng 11, đã được xác nhận rằng gen *aroD* bị xóa, nồng độ của shinorin được cải thiện tới 239% so với nhóm đối chứng, do sinh tổng hợp shinorin còn được tăng cường thông qua pECCG117\_PCJ1\_Ava\_ABCD, nồng độ của shinorin được tăng lên. Theo đó, chủng bị xóa *aroD*, c.gl 13032  $\Delta\text{N1021\_PO2\_Ava\_ABCD\_}\Delta\text{aroD}$ , được định danh là CB06-0019 và được ký gửi tại Bảng tàng giống chuẩn Vi sinh vật Hàn Quốc theo Hiệp ước Budapest vào ngày 26 tháng 06 năm 2017, với số định danh KCCM12046P.

Xây dựng các vi sinh vật tái tổ hợp sản xuất MAA có nguồn gốc từ nấm men và sản xuất MAAs sử dụng vi sinh vật đó.

Ví dụ 15: Xây dựng các vector nấm men biểu hiện quá mức gen sinh tổng hợp shinorin có nguồn gốc từ vi tảo.

Các vector *S.cerevisiae*, trong đó các gen sinh tổng hợp shinorin được đưa vào dựa trên bộ gen DNA của *A. variabilis* ATCC29413 và *N. punctiforme* ATCC29133, đã được xây dựng. Các vector được điều chế bằng cách sử dụng các vùng gen khởi động ADH, TEF, và GPD của *S. cerevisiae*, và các mạch khuôn và các môi được sử dụng để điều chế tổng hai mươi tư vector biểu hiện gen sinh tổng hợp shinorin được thể hiện trong bảng 12 bên dưới. Các trình tự nucleotit và các trình tự axit amin của Ava\_A, Ava\_B, Ava\_C, Ava\_D, Npr\_A, Npr\_B, Npr\_C, và Npr\_D được biểu diễn cụ thể bởi SEQ ID NO: 89 đến SEQ ID NO: 104 theo thứ tự của các enzym.

Bảng 12

| Tên các vector  | Các mạch khuôn được sử dụng                   | Các môi được sử dụng (môi xuôi, môi ngược) |
|-----------------|-----------------------------------------------|--------------------------------------------|
| p413-pADH-Ava_A | Bộ gen DNA của <i>A. variabilis</i> ATCC29413 | SEQ ID NO: 45, SEQ ID NO: 46               |
| p413-pADH-Ava_B |                                               | SEQ ID NO: 47, SEQ ID NO: 48               |
| p413-pADH-Ava_C |                                               | SEQ ID NO: 49, SEQ ID NO: 50               |
| p413-pADH-Ava_D |                                               | SEQ ID NO: 51, SEQ ID NO: 52               |
| p413-pTEF-Ava_A |                                               | SEQ ID NO: 45, SEQ ID NO: 46               |
| p413-pTEF-Ava_B |                                               | SEQ ID NO: 47, SEQ ID NO: 48               |
| p413-pTEF-Ava_C |                                               | SEQ ID NO: 49, SEQ ID NO: 50               |
| p413-pTEF-Ava_D |                                               | SEQ ID NO: 51, SEQ ID NO: 52               |
| p413-pGPD-Ava_A |                                               | SEQ ID NO: 45, SEQ ID NO: 46               |
| p413-pGPD-Ava_B |                                               | SEQ ID NO: 47, SEQ ID NO: 50               |
| p413-pGPD-Ava_C |                                               | SEQ ID NO: 49, SEQ ID NO: 50               |
| p413-pGPD-Ava_D |                                               | SEQ ID NO: 51, SEQ ID NO: 52               |
| p413-pADH-Npr_A | Bộ gen DNA của <i>N.punctiforme</i> ATCC29133 | SEQ ID NO: 53, SEQ ID NO: 54               |
| p413-pADH-Npr_B |                                               | SEQ ID NO: 55, SEQ ID NO: 56               |
| p413-pADH-Npr_C |                                               | SEQ ID NO: 57, SEQ ID NO: 58               |
| p413-pADH-Npr_D |                                               | SEQ ID NO: 59, SEQ ID NO: 60               |
| p413-pTEF-Npr_A |                                               | SEQ ID NO: 53, SEQ ID NO: 54               |
| p413-pTEF-Npr_B |                                               | SEQ ID NO: 55, SEQ ID NO: 56               |
| p413-pTEF-Npr_C |                                               | SEQ ID NO: 57, SEQ ID NO: 58               |
| p413-pTEF-Npr_D |                                               | SEQ ID NO: 59, SEQ ID NO: 60               |
| p413-pGPD-Npr_A |                                               | SEQ ID NO: 53, SEQ ID NO: 54               |
| p413-pGPD-Npr_B |                                               | SEQ ID NO: 55, SEQ ID NO: 56               |
| p413-pGPD-Npr_C |                                               | SEQ ID NO: 57, SEQ ID NO: 58               |
| p413-pGPD-Npr_D |                                               | SEQ ID NO: 59, SEQ ID NO: 60               |

Các đoạn gen thu được bằng PCR sử dụng kết hợp các mạch khuôn và cặp môi ở trên được nối với vector p413/414/415/416-pADH/pTEF/pGPD-CYC1\_terminator đã được xử lý với enzym giới hạn BamH1/XhoI sử dụng enzym T4 ligaza, nhờ đó điều chế hai mươi tư vector p413/414/415/416-pADH/pTEF/pGPD-A, B, C, D. Việc xây dựng mỗi vector biểu hiện và trình tự gen của chúng được xác nhận bằng các kỹ thuật giải trình tự. Các vector biểu hiện được điều chế như vậy được đưa vào trong chủng *S. cerevisiae* CEN.PK-1D kiểu đại để điều chế các chủng có khả năng sản sinh shinorin.

Ví dụ 16: Đánh giá các chủng có khả năng sản sinh shinorin được đưa vào với các vector biểu hiện quá mức gen sinh tổng hợp shinorin.

Để xác nhận khả năng sản sinh MAA của các nấm men, hai mươi tư plasmid được điều chế trong Ví dụ 15 được đưa vào trong chủng *Saccharomyces cerevisiae* CEN.PK-1D (*S. cerevisiae* CEN.PK-1D) để điều chế các chủng có sinh tổng hợp shinorin được tăng cường, và các chủng được cấy trải trên môi trường SC đặc (synthetic complete-tổng hợp hoàn thành) không chứa Leu, Trp, Ura, và His, và được nuôi cấy qua đêm trong tủ ấm ở 30°C. Lấy một vòng que cấy bạch kim của mỗi chủng được nuôi cấy qua đêm được cấy chuyển vào trong 25 ml môi trường chuẩn độ trong bảng 13 và sau đó được nuôi cấy trong tủ ấm ở 30°C ở tốc độ lắc 150 vòng/phút trong 24 giờ. Các kết quả được thể hiện trong Bảng 14 bên dưới.

Bảng 13

| Thành phần                                                             | Nồng độ (g/L) |
|------------------------------------------------------------------------|---------------|
| YNB (Yeast nitrogen base) Không chứa các axit amin                     | 6,7           |
| Hỗn hợp axit amin (không bao gồm Leuxin, Tryptophan, Histidin, Uracil) | 2             |
| Glucosa                                                                | 20            |

Bảng 14

| Plasmid          | 24 giờ            |          | Nồng độ Shinorin (mg/L) |
|------------------|-------------------|----------|-------------------------|
|                  | OD <sub>600</sub> | Đường dư |                         |
| pADH-Ava_A,B,C,D | 11,0              | 0,0      | 107                     |
| pTEF-Ava_A,B,C,D | 11,1              | 0,0      | 215                     |
| pGPD-Ava_A,B,C,D | 11,5              | 0,0      | 302                     |
| pADH-Npr_A,B,C,D | 20,1              | 0,0      | 234                     |
| pTEF-Npr_A,B,C,D | 20,4              | 0,0      | 387                     |
| pGPD-Npr_A,B,C,D | 20,5              | 0,0      | 521                     |

Dựa trên các kết quả ở trên, đã xác nhận rằng *S. cerevisiae* CEN.PK-1D (tức là, chủng nấm men) được thể hiện hoạt động cao của các gen Npr A, B, C, D so với các gen Ava A, B, C, D. Ngoài ra, đã xác nhận rằng mức độ biểu hiện của các gen được kiểm soát theo mật độ của các vùng gen khởi động, và do đó, sản xuất shinorin bị thay đổi. Cụ thể, đã xác nhận rằng sản xuất shinorin là cao nhất (521 mg/L) khi vùng gen khởi động GPD (Glyceraldehyde-3-phosphat dehydrogenase)-vector dựa trên Npr A, B, C, D được đưa vào.

Ví dụ 17: Tăng sản xuất shinorin bằng cách xóa *ARO1* trong *S. cerevisiae* và việc

đưa vào của *E. coli aroB*.

Để xác định có hay không sự bất hoạt của dehydroquinate dehydratase cải thiện sản xuất shinorin trong các nấm men, gen *ARO1* bị xóa trong *S. cerevisiae* CEN.PK-1D. Gen *ARO1* của *S. cerevisiae* là gen đảm nhiệm năm chức năng, và chức năng của 3-dehydroquinate synthase tương ứng với *E. coli aroB* bị mất khi xóa gen *ARO1*, làm cho không thể tổng hợp 3-DHQ. Do đó, sau khi xóa *S. cerevisiae ARO1* (SEQ ID NO: 81 và SEQ ID NO: 82), là tương đồng *E. coli aroB* trên nhiễm sắc thể, gen *E. coli aroB* (SEQ ID NO: 83 và SEQ ID NO: 84) được chèn tại cùng vị trí dựa trên vùng gen khởi động GDP. Các mạch khuôn và các môi sử dụng được thể hiện trong bảng 15. Hai mươi tư plasmid được điều chế trong ví dụ 15 được đưa vào trong chủng *S. cerevisiae* CEN.PK-1D, trong đó các gen *ARO1* bị xóa và gen *E. coli aroB* được đưa vào, và các chủng được cấy trên môi trường SC đặc (hoàn thành tổng hợp) không chứa Leu, Trp, Ura, và His và được nuôi cấy qua đêm trong tủ ấm ở 30°C. Lấy một vòng que cấy bạch kim của mỗi chủng được nuôi cấy qua đêm được cấy chuyển vào trong 25 ml môi trường chuẩn độ được thể hiện trong bảng 13 và sau đó được nuôi cấy trong tủ ấm ở 30°C ở tốc độ lắc 150 vòng/phút trong 24 giờ. Các kết quả được thể hiện trong bảng 16 bên dưới.

Bảng 15

| Tên catxet                            | DNA được khuếch đại | Mạch khuôn                   | Trình tự môi (môi xuôi, môi ngược) |
|---------------------------------------|---------------------|------------------------------|------------------------------------|
| Catxet <i>ARO1</i> ΔpGPD- <i>aroB</i> | pGPD                | gDNA của <i>S.cerevisiae</i> | SEQ ID NO: 61, SEQ ID NO: 62       |
|                                       | LoxP(Ura)           | vectơ pUG28                  | SEQ ID NO: 63, SEQ ID NO: 64       |
|                                       | <i>aroB</i>         | gDNA của W3110               | SEQ ID NO: 65, SEQ ID NO: 66       |
|                                       | <i>ARO1</i> đoạn 1  | gDNA của <i>S.cerevisiae</i> | SEQ ID NO: 67, SEQ ID NO: 68       |
|                                       | <i>ARO1</i> đoạn 2  | gDNA của <i>S.cerevisiae</i> | SEQ ID NO: 69, SEQ ID NO: 70       |

Bảng 16

| Các plasmid         | 24 giờ            |          | Nồng độ Shinorin (mg/L) |
|---------------------|-------------------|----------|-------------------------|
|                     | OD <sub>600</sub> | Đường dư |                         |
| WT pADH-Ava_A,B,C,D | 11,0              | 0,0      | 103                     |
| WT pTEF-Ava_A,B,C,D | 11,1              | 0,0      | 235                     |

|                                                                          |      |     |       |
|--------------------------------------------------------------------------|------|-----|-------|
| WT pGPD-Ava_A,B,C,D                                                      | 11,5 | 0,0 | 323   |
| WT pADH-Npr_A,B,C,D                                                      | 20,1 | 0,0 | 230   |
| WT pTEF-Npr_A,B,C,D                                                      | 20,4 | 0,0 | 390   |
| WT pGPD-Npr_A,B,C,D                                                      | 20,5 | 0,0 | 545   |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pADH-Ava_A,B,C,D | 13,0 | 0,0 | 305   |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pTEF-Ava_A,B,C,D | 13,1 | 0,0 | 635   |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pGPD-Ava_A,B,C,D | 13,3 | 0,0 | 912   |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pADH-Npr_A,B,C,D | 24,1 | 0,0 | 670   |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pTEF-Npr_A,B,C,D | 25,3 | 0,0 | 1.230 |
| <i>S. cerevisiae</i> <i>ARO1</i> $\Delta$ pGPD-aroB,<br>pGPD-Npr_A,B,C,D | 24,8 | 0,0 | 1.600 |

Dựa trên các kết quả được thể hiện trong bảng 16, đã xác nhận rằng sản xuất shinorin đã tăng lên gấp ba lần trong các chủng, trong đó khả năng sản xuất DHQ được tăng cường bằng cách xóa *ARO1* và đưa vào *E. coli* *aroB* so với các chủng WT. Ngoài ra, đã xác nhận rằng các chủng cho thấy hoạt động cao của các gen Npr A, B, C, D, và sản xuất shinorin như vậy được tăng lên khi mật độ của các vùng gen khởi động tăng lên. Cụ thể, đã xác nhận rằng sản xuất shinorin là cao nhất (1,6 g/L) khi vùng gen khởi động GPD-vector dựa trên Npr A, B, C, D được đưa vào.

Người có hiểu biết trung bình trong cùng lĩnh vực kỹ thuật sẽ nhận ra rằng sáng chế có thể được thực hiện trong các hình thức cụ thể khác nhau mà không tách rời khỏi nguyên lý hoặc các đặc điểm thiết yếu của sáng chế. Các phương án được mô tả phải được xem xét trong tất cả các khía cạnh chỉ mang tính minh họa và không hạn chế. Phạm vi của sáng chế được chỉ định bằng các yêu cầu bảo hộ kèm theo thay vì mô tả ở trên. Tất cả các thay đổi trong ý nghĩa và phạm vi tương đương của các yêu cầu bảo hộ được chấp nhận trong phạm vi của sáng chế.

## YÊU CẦU BẢO HỘ

1. Vi sinh vật sản sinh các axit amin dạng mycosporin, trong đó hoạt động 3-đehydroquinat dehydrataza bị bất hoạt so với vi sinh vật không biến đổi.
2. Vi sinh vật theo điểm 1, trong đó vi sinh vật chứa gen mã hóa cho ít nhất một protein được chọn từ nhóm bao gồm 2-đemetyl 4-đeoxygadusol synthaza, O-methyltransferaza, và C-N ligaza.
3. Vi sinh vật theo điểm 1, trong đó vi sinh vật chứa gen mã hóa cho ít nhất một protein được chọn từ nhóm bao gồm non-ribosomal peptit synthetaza, enzym dạng non-ribosomal peptit synthetaza (NRPS-dạng enzym), và D-Ala D-Ala ligaza.
4. Vi sinh vật theo điểm 1, trong đó hoạt tính của ít nhất một protein được chọn từ nhóm bao gồm 2-đehydro-3-đeoxyphosphoheptonat aldolaza, phosphoenolpyruvat synthetaza, transketolaza I/II, và 3-đehydroquinat synthaza được tăng cường hơn so với vi sinh vật không biến đổi.
5. Vi sinh vật theo điểm 1, trong đó vi sinh vật là vi sinh vật thuộc chi *Corynebacterium*, vi sinh vật thuộc chi *Escherichia*, hoặc nấm men.
6. Vi sinh vật theo điểm 5, trong đó nấm men được đưa vào với gen mã hóa 3-đehydroquinat synthaza.
7. Vi sinh vật theo điểm 1, trong đó các axit amin dạng mycosporin được chọn ít nhất một từ nhóm bao gồm mycosporin-2-glyxin, palythinol, axit palythenic, deoxygadusol, mycosporin-metylamin-threonin, mycosporin-glyxin-valin, palythin, asterina-330, shinorin, porphyra-334, euhalothece-362, mycosporin-glyxin, mycosporin-ornithin, mycosporin-lysin, mycosporin-glutamic axit-glyxin, mycosporin-metylamin-serin, mycosporin-taurin, palythen, palythin-serin, palythin-serin-sulfat, palythinol, và usujiren.
8. Phương pháp để sản xuất các axit amin dạng mycosporin bao gồm:
  - nuôi cấy vi sinh vật của điểm bất kỳ từ 1 đến 7; và
  - thu hồi các axit amin dạng mycosporin từ vi sinh vật được nuôi cấy hoặc môi trường.

## DANH MỤC TRÌNH TỰ GEN

<110> CJ CHEILJEDANG CORPORATION

<120> VI SINH VẬT SẢN SINH CÁC AXIT AMIN DẠNG MYCOSPORIN VÀ PHƯƠNG  
PHÁP SẢN XUẤT CÁC AXIT AMIN DẠNG MYCOSPORIN BẰNG VI SINH VẬT NÀY

<130> OPA18235-PCT

<150> KR10-2017-0103795

<151> 2017-08-16

<160> 104

<170> KopatentIn 2.0

<210> 1

<211> 6461

<212> DNA

<213> Anabaena variabilis ATCC29413

<400> 1

```
atgagtatcg tccaagcaaa gtttgaagct aaggaaacat cttttcatgt agaaggttac 60
gaaaagattg agtatgattt ggtgtatgta gatggtattt ttgaaatcca gaattctgca 120
ctagcagatg tatatcaagg ttttggacga tgcttggcga ttgtagatgc taacgtcagt 180
cggttgtatg gtaatcaaatt tcaggcatat ttccagtatt atggtataga actgaggcta 240
tttcctatta ccattactga accagataag actattcaaa ctttcgagag agttatagat 300
gtctttgcag atttcaaatt agtccgcaaa gaaccagtat tagtcgtggg tggcggttta 360
attacagatg ttgtcggctt tgcttgttct acatatcgtc gcagcagcaa ttacatccgc 420
attcctacta cattgattgg attaatgat gccagtgtag caattaagggt agcagttaat 480
catcgcaaac tgaaaaaccg tttgggtgct tatcatgctt ctcgcaaagt attttttagat 540
ttctccttgt tgcgtactct ccctacagac caagtacgta acgggatggc ggaattggta 600
aaaatcgctg tagtagcgca tcaagaagtt tttgaattgt tggagaagta cggcgaagaa 660
ttactacgta ctcatTTTgg caatatagat gcaactccag agattaaaga aatagcccat 720
cgTTTgactt acaaagctat ccataagatg ttggaattgg aagttcccaa cctgcatgag 780
ttagacctag atagggatgat tgcttacggg cacacttgga gtccacacct ggaacttgcg 840
cctcgtctac ccatgttcca cggacacgcc gttaatgtag atatggcttt ctcggaacg 900
atcgccgcc gtagaggata tattacaatt gcagaacgcg atcgatTTTT aggattaatg 960
agtcgcgttg gtctatccct cgaccatccc atgttgata tagatatTTT gtggcgtggg 1020
actgaatcta tcacattaac tcgtgatggg ttgttaagag ctgctatgcc aaaaccatt 1080
```

ggtgattgtg tcttcgtcaa tgacctgaca agagaagaat tagcagccgc attagctgac 1140  
 cacaagaac tttgtaccag ttatccccgt ggtggtgaag gtgtggatgt gtatcccgtt 1200  
 tatcaaaaag aattaatcgg gagtgttaaa taatgacttt tttgaattca aaatgcaaaa 1260  
 tactccacgg atacactgcg cgagcgcggt agcatttctg ttcgcggagc gtcccgtagg 1320  
 gaaagagaag gctacgcaaa taatcggaca ctaattgtct ttaattttga attttgaatt 1380  
 ttgaattttg aattggagcg aagcgacttg acaaagtga ttgtccaacc aacagctaga 1440  
 cctgttacac cattgggaat ttttaaccaag cagttagaag ccatagtcca agaggttaag 1500  
 caacatccag atttacctgg ggaattgata gcaaaccatcc atcaggcttg gcgtttagcc 1560  
 gcaggatatag acccttattt ggaagaatgc accactccag aatctcctga actcgctgca 1620  
 ttggcaaaaa ccacagccac cgaagcctgg ggagaacact tccacggagg tacaaccgtc 1680  
 cgtcctctag aacaagagat gctttctggt catatcgaag gacaaacctt aaagatgttt 1740  
 gttcacatga ccaaagctaa aaaagtctta gaaattggga tgtttaccgg ttattcggcg 1800  
 ctggcgatgg cggaagcatt accagaggat ggactgcttg tggcttgtga agttgaccct 1860  
 tacgcggcgg aaattggaca gaaagccttt caacaatctc cccacggtagg aaagattcgt 1920  
 gtggaattgg atgcagcctt agcaactctt gataagttag cagaagctgg ggagtctttt 1980  
 gacttggtat ttatcgacgc agataaaaaa gagtatgtag cctattttca caagttgcta 2040  
 ggtagcagtt tgtagcacc agatggcttt atttgtgtag ataacacctt attacaaggg 2100  
 gaagtttatc taccagcaga ggaacgtagc gtcaatggtg aagcgatcgc gcaatttaat 2160  
 catacagtag ctatagacc ccgtgtagaa caggttttgt tgccgttgcg agatggttta 2220  
 acaattatcc gcagaatata accttaattg tccaatcgac tatggcaciaa tcccttcccc 2280  
 tttcttcgcg acctgctaca ccgtctcttc cttcccagac gaaaatagcc gcaattatcc 2340  
 aaaatatctg cactttggct ttgttattac tagcattgcc cattaatgcc accattgttt 2400  
 ttatatcctt gttagtcttc cgaccgcaaa aggtcaaagc agcaaaccoc caaaccattc 2460  
 ttatcagtg cggtgaagatg accaaagctt tacaactagc aaggtcattc cacgcggctg 2520  
 gacatagagt tgtcttggtg gaaaccata aatactgggt gactgggtcat cgtttttccc 2580  
 aagcagtgga taagttttac acagtccccg caccacagga caatcccaa gcttacattc 2640  
 aggctttgggt agatatcgtc aaacaagaaa acatcgatgt ttatatccc gtcaccagtc 2700  
 cagtgggtag ctactacgac tcattagcca aaccagagtt atcccattat tgcgaagtgt 2760  
 ttcactttga cgcagatatt acccaaatgt tggatgataa atttgcgttg acacaaaaag 2820

cgcgatcgct tggtttatca gtacccaaat cctttaaaat tacctcacca gaacaagtca 2880  
 tcaacttcga tttttctgga gagacacgta aatacatcct caaaagcatt ccctacgact 2940  
 cagtgcggcg gttggactta accaaactcc cctgtgctac tccagaggaa acagcagcat 3000  
 tcgtcagaag tttgccaaatt actcccgaaa aaccgtggat tatgcaggaa tttatccccg 3060  
 gtaaggaatt ctgcacccat agcaccgttc ggaatgggga actcagactg cattgctggt 3120  
 gcgaatcttc agccttccaa gttaattatg agaatgtaaa taaccgcaa attaccgaat 3180  
 gggtagagca ttttgtcaag gaactgaaac tgacaggaca gatttccttt gactttatcc 3240  
 aagccgaaga cggaacagtt tacgccatcg agtgtaacct ccgcacacat tcagcaatta 3300  
 ccacatttta cgaccacccc caggtagcag aagcgtactt gagtcaagca ccgacgactg 3360  
 aaaccataca accactaacg acaagcaagc ctacctattg gacttatcac gaagtttggc 3420  
 gtttaactgg tatccgttct ttcaccagc tgcaaagatg gctggggaat atttggcgcg 3480  
 ggactgatgc gatttatcag ccagatgacc ccttaccgtt tttgatggta catcattggc 3540  
 aaattcccct actgttattg aataatttgc gtcgtcttaa aggttggacg cggatagatt 3600  
 tcaatattgg gaagttggtg gaattggggg gagattagtt tttaaacgca gagggacgct 3660  
 gaggttagcg cagcgaaaag ttctggagga gggtttcct ccgtaggaaa cttttcaaga 3720  
 gagagggacg cggagtgtgt tttctctgcg tctctgcgtg agaaattttt tattattgag 3780  
 caaagttaga agatatgcag actatagatt ttaatatctg taagttactt gtagagtgga 3840  
 acgcgacca cagagattat gatctttccc agagtttaca tgaactaatt gtagctcaag 3900  
 tagaacgaac acctgaggcg atcgctgtca cctttgacaa gcaacaacta acttatcaag 3960  
 aactaaatca taaagcaaac cagctaggac attatttaca aacattagga gtccagccag 4020  
 aaaccctggt aggcgtttgt ttagaacgtt ccttagaaat ggttatctgt cttttaggaa 4080  
 tcctcaaagc tgggggtgct tatgttccta ttgaccctga atatcctcaa gaacgcatag 4140  
 cttatatgct agaagattct caggtgaagg tactactaac tcaagaaaaa ttactcaatc 4200  
 aaattcccca ccatcaagca caaactatct gtgtagatag ggaatgggag aaaatttcca 4260  
 cacaagctaa taccaatccc aaaagtaata taaaaacgga taatcttgct tatgtaattt 4320  
 acacctctgg ttccactggt aaaccaaag gtgcaatgaa caccacaaa ggtatctgta 4380  
 atcgcttatt gtggatgcag gaagcttatt aaatcgattc cacagatagc attttacaaa 4440  
 aaaccccctt tagttttgat gtttccgttt gggagttctt ttggacttta ttaactggcg 4500  
 cacgtttggt aatagccaaa ccaggcggac ataaagatag tgcttacctc atcgatttaa 4560

ttactcaaga acaaactact acgttgcat tttgtccctc aatgctgcaa gtgtttttac 4620  
 aaaatcgcca tgtaagcaaa tgcagctctc taaaaagagt tatttgtagc ggtgaagctt 4680  
 tatctataga tttacaaaat agatTTTTcc agcatttgca atgtgaatta cataacctct 4740  
 atggcccgac agaagcagca attgatgtca cattttggca atgtagaaaa gatagtaatt 4800  
 taaagagtgt acctattggt cgtcccatg ctaatactca aatttatatt cttgatgccc 4860  
 atttacaacc agtaaataatt ggtgtcactg gtgaaattta tattggtggt gtaggggttg 4920  
 ctctgtggtta tttgaataaa gaagaattga ccaaagaaaa atttattatt aatccctttc 4980  
 ccaattctga gtttaagcga ctttataaaa caggtgat t agctcgttat ttacccgatg 5040  
 gaaatattga atatcttgggt agaacagatt atcaagtaaa aattcgggggt tatagaattg 5100  
 aaattggcga gattgaaaat gttttatctt cacacccaca agtcagagaa gctgtagtca 5160  
 tagcgcgggga tgataacgct caagaaaaac aaatcatcgc ttatattacc tataactcca 5220  
 tcaaacctca gcttgataat ctgctgatt tcctaaaagc aaggctacct gattttatga 5280  
 ttccagccgc ttttgtgatg ctggagcatc ttcctttaac tcccagtgggt aaagtagacc 5340  
 gtaaggcatt acctaacct gatttattta attatagtga acataattcc tatgtagcgc 5400  
 ctcggaatga agttgaagaa aaattagtac aaatctggtc gaatattctg catttaccta 5460  
 aagtaggtgt gacagaaaac ttttctgcta ttgggtgtaa ttcctcaaa gctctacatt 5520  
 taatttctca aattgaagag ttatttgcta aagagatct cttagcaaca cttttaacaa 5580  
 atccagtaat tgcagattta gccagggtta ttcaagcaaa caaccaaact cataattcac 5640  
 ccctagtctc aattcaacca caaggtaagc agcagccttt cttttgtata catcctgctg 5700  
 gtggtcatgt tttatgctat tttaaactcg cacaatatat aggaactgac caaccatttt 5760  
 atggcttaca agctcaagga ttttatggag atgaagcacc cttgacgcga gttgaagata 5820  
 tggctagtct ctacgtcaaa actattagag aatttcaacc ccaagggcct tatcgtgtcg 5880  
 ggggggtggtc atttggtgga gtcgtagctt atgaagtagc acagcagtta catagacaag 5940  
 gacaagaagt atctttacta gcaatattag attcttacgt accgattctg ctggataaac 6000  
 aaaaacccat tgatgacgtt tatttagttg gtgttctctc cagagttttt ggcggtatgt 6060  
 ttggtcaaga taatctagtc acacctgaag aaatagaaaa tttaactgta gaagaaaaaa 6120  
 ttaattacat cattgataaa gcacggagcg ctagaatatt cccgcctgggt gtagaacgtc 6180  
 aaaataatcg ccgtattctt gatgttttgg tgggaacttt aaaagcaact tattcctata 6240  
 taagacaacc atatccagga aaagtactg tatttcgagc cagggaaaaa catattatgg 6300

ctcctgaccc gaccttagtt tgggtagaat tattttctgt aatggcggct caagaaatta 6360  
 agattattga tgtccctgga aaccattatt cgtttgttct agaaccatcat gtacagggtt 6420  
 tagcacagcg tttacaagat tgtctggaaa ataattcata a

6461

<210> 2  
 <211> 2035  
 <212> PRT  
 <213> Anabaena variabilis ATCC29413

<400> 2  
 Met Ser Ile Val Gln Ala Lys Phe Glu Ala Lys Glu Thr Ser Phe His  
 1 5 10  
 15

Val Glu Gly Tyr Glu Lys Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly  
 20 25  
 30

Ile Phe Glu Ile Gln Asn Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe  
 35 40  
 45

Gly Arg Cys Leu Ala Ile Val Asp Ala Asn Val Ser Arg Leu Tyr Gly  
 50 55 60

Asn Gln Ile Gln Ala Tyr Phe Gln Tyr Tyr Gly Ile Glu Leu Arg Leu  
 65 70 75  
 80

Phe Pro Ile Thr Ile Thr Glu Pro Asp Lys Thr Ile Gln Thr Phe Glu  
 85 90  
 95

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

Val Leu Val Val Gly Gly Gly Leu Ile Thr Asp Val Val Gly Phe Ala  
 115 120  
 125

Cys Ser Thr Tyr Arg Arg Ser Ser Asn Tyr Ile Arg Ile Pro Thr Thr  
 130 135 140

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

His Arg Lys Leu Lys Asn Arg Leu Gly Ala Tyr His Ala Ser Arg Lys  
 165 170  
 175

Val Phe Leu Asp Phe Ser Leu Leu Arg Thr Leu Pro Thr Asp Gln Val

[illegible]

Leu Glu Ala Ile Val Gln Glu Val Lys Gln His Pro Asp Leu Pro Gly  
 435 440  
 445

Glu Leu Ile Ala Asn Ile His Gln Ala Trp Arg Leu Ala Ala Gly Ile  
 450 455 460

Asp Pro Tyr Leu Glu Glu Cys Thr Thr Pro Glu Ser Pro Glu Leu Ala  
 465 470 475  
 480

Ala Leu Ala Lys Thr Thr Ala Thr Glu Ala Trp Gly Glu His Phe His  
 485 490 495

Gly Gly Thr Thr Val Arg Pro Leu Glu Gln Glu Met Leu Ser Gly His  
 500 505  
 510

Ile Glu Gly Gln Thr Leu Lys Met Phe Val His Met Thr Lys Ala Lys  
 515 520  
 525

Lys Val Leu Glu Ile Gly Met Phe Thr Gly Tyr Ser Ala Leu Ala Met  
 530 535 540

Ala Glu Ala Leu Pro Glu Asp Gly Leu Leu Val Ala Cys Glu Val Asp  
 545 550 555  
 560

Pro Tyr Ala Ala Glu Ile Gly Gln Lys Ala Phe Gln Gln Ser Pro His  
 565 570 575

Gly Gly Lys Ile Arg Val Glu Leu Asp Ala Ala Leu Ala Thr Leu Asp  
 580 585  
 590

Lys Leu Ala Glu Ala Gly Glu Ser Phe Asp Leu Val Phe Ile Asp Ala  
 595 600  
 605

Asp Lys Lys Glu Tyr Val Ala Tyr Phe His Lys Leu Leu Gly Ser Ser  
 610 615 620

Leu Leu Ala Pro Asp Gly Phe Ile Cys Val Asp Asn Thr Leu Leu Gln  
 625 630 635  
 640

Gly Glu Val Tyr Leu Pro Ala Glu Glu Arg Ser Val Asn Gly Glu Ala  
 645 650  
 655

Ile Ala Gln Phe Asn His Thr Val Ala Ile Asp Pro Arg Val Glu Gln  
 660 665  
 670

Val Leu Leu Pro Leu Arg Asp Gly Leu Thr Ile Ile Arg Arg Ile Gln

## 41505

[illegible]

```

Thr Pro Glu Glu Thr Ala Ala Phe Val Arg Ser Leu Pro Ile Thr Pro      940
  930                                935

Glu Lys Pro Trp Ile Met Gln Glu Phe Ile Pro Gly Lys Glu Phe Cys
945                                950                                955
                                960

Thr His Ser Thr Val Arg Asn Gly Glu Leu Arg Leu His Cys Cys Cys
                                965                                970
                                975

Glu Ser Ser Ala Phe Gln Val Asn Tyr Glu Asn Val Asn Asn Pro Gln
                                980                                985
                                990

Ile Thr Glu Trp Val Gln His Phe Val Lys Glu Leu Lys Leu Thr Gly
  995                                1000                                1005

Gln Ile Ser Phe Asp Phe Ile Gln Ala Glu Asp Gly Thr Val Tyr Ala
  1010                                1015                                1020

Ile Glu Cys Asn Pro Arg Thr His Ser Ala Ile Thr Thr Phe Tyr Asp
1025                                1030                                1035                                1040

His Pro Gln Val Ala Glu Ala Tyr Leu Ser Gln Ala Pro Thr Thr Glu
                                1045                                1050                                1055

Thr Ile Gln Pro Leu Thr Thr Ser Lys Pro Thr Tyr Trp Thr Tyr His
  1060                                1065
  1070

Glu Val Trp Arg Leu Thr Gly Ile Arg Ser Phe Thr Gln Leu Gln Arg
                                1075                                1080
  1085

Trp Leu Gly Asn Ile Trp Arg Gly Thr Asp Ala Ile Tyr Gln Pro Asp
  1090                                1095                                1100

Asp Pro Leu Pro Phe Leu Met Val His His Trp Gln Ile Pro Leu Leu
1105                                1110                                1115                                1120

Leu Leu Asn Asn Leu Arg Arg Leu Lys Gly Trp Thr Arg Ile Asp Phe
                                1125                                1130                                1135

Asn Ile Gly Lys Leu Val Glu Leu Gly Gly Asp Met Gln Thr Ile Asp
  1140                                1145
  1150

Phe Asn Ile Arg Lys Leu Leu Val Glu Trp Asn Ala Thr His Arg Asp
  1155                                1160
  1165

Tyr Asp Leu Ser Gln Ser Leu His Glu Leu Ile Val Ala Gln Val Glu
  1170                                1175                                1180

Arg Thr Pro Glu Ala Ile Ala Val Thr Phe Asp Lys Gln Gln Leu Thr
1185                                1190                                1195                                1200

```

Tyr Gln Glu Leu Asn His Lys Ala Asn Gln Leu Gly His Tyr Leu Gln  
 1205 1210 1215  
 Thr Leu Gly Val Gln Pro Glu Thr Leu Val Gly Val Cys Leu Glu Arg  
 1220 1225  
 1230  
 Ser Leu Glu Met Val Ile Cys Leu Leu Gly Ile Leu Lys Ala Gly Gly  
 1235 1240 1245  
 Ala Tyr Val Pro Ile Asp Pro Glu Tyr Pro Gln Glu Arg Ile Ala Tyr  
 1250 1255 1260  
 Met Leu Glu Asp Ser Gln Val Lys Val Leu Leu Thr Gln Glu Lys Leu  
 1265 1270 1275 1280  
 Leu Asn Gln Ile Pro His His Gln Ala Gln Thr Ile Cys Val Asp Arg  
 1285 1290  
 1295  
 Glu Trp Glu Lys Ile Ser Thr Gln Ala Asn Thr Asn Pro Lys Ser Asn  
 1300 1305  
 1310  
 Ile Lys Thr Asp Asn Leu Ala Tyr Val Ile Tyr Thr Ser Gly Ser Thr  
 1315 1320 1325  
 Gly Lys Pro Lys Gly Ala Met Asn Thr His Lys Gly Ile Cys Asn Arg  
 1330 1335 1340  
 Leu Leu Trp Met Gln Glu Ala Tyr Gln Ile Asp Ser Thr Asp Ser Ile  
 1345 1350 1355 1360  
 Leu Gln Lys Thr Pro Phe Ser Phe Asp Val Ser Val Trp Glu Phe Phe  
 1365 1370  
 1375  
 Trp Thr Leu Leu Thr Gly Ala Arg Leu Val Ile Ala Lys Pro Gly Gly  
 1380 1385 1390  
 His Lys Asp Ser Ala Tyr Leu Ile Asp Leu Ile Thr Gln Glu Gln Ile  
 1395 1400 1405  
 Thr Thr Leu His Phe Val Pro Ser Met Leu Gln Val Phe Leu Gln Asn  
 1410 1415 1420  
 Arg His Val Ser Lys Cys Ser Ser Leu Lys Arg Val Ile Cys Ser Gly  
 1425 1430 1435  
 1440  
 Glu Ala Leu Ser Ile Asp Leu Gln Asn Arg Phe Phe Gln His Leu Gln  
 1445 1450  
 1455  
 Cys Glu Leu His Asn Leu Tyr Gly Pro Thr Glu Ala Ala Ile Asp Val  
 1460 1465 1470

Thr Phe Trp Gln Cys Arg Lys Asp Ser Asn Leu Lys Ser Val Pro Ile  
 1475 1480  
 1485

Gly Arg Pro Ile Ala Asn Thr Gln Ile Tyr Ile Leu Asp Ala Asp Leu  
 1490 1495 1500

Gln Pro Val Asn Ile Gly Val Thr Gly Glu Ile Tyr Ile Gly Gly Val  
 1505 1510 1515 1520

Gly Val Ala Arg Gly Tyr Leu Asn Lys Glu Glu Leu Thr Lys Glu Lys  
 1525 1530  
 1535

Phe Ile Ile Asn Pro Phe Pro Asn Ser Glu Phe Lys Arg Leu Tyr Lys  
 1540 1545  
 1550

Thr Gly Asp Leu Ala Arg Tyr Leu Pro Asp Gly Asn Ile Glu Tyr Leu  
 1555 1560  
 1565

Gly Arg Thr Asp Tyr Gln Val Lys Ile Arg Gly Tyr Arg Ile Glu Ile  
 1570 1575 1580

Gly Glu Ile Glu Asn Val Leu Ser Ser His Pro Gln Val Arg Glu Ala  
 1585 1590 1595 1600

Val Val Ile Ala Arg Asp Asp Asn Ala Gln Glu Lys Gln Ile Ile Ala  
 1605 1610 1615

Tyr Ile Thr Tyr Asn Ser Ile Lys Pro Gln Leu Asp Asn Leu Arg Asp  
 1620 1625  
 1630

Phe Leu Lys Ala Arg Leu Pro Asp Phe Met Ile Pro Ala Ala Phe Val  
 1635 1640  
 1645

Met Leu Glu His Leu Pro Leu Thr Pro Ser Gly Lys Val Asp Arg Lys  
 1650 1655 1660

Ala Leu Pro Lys Pro Asp Leu Phe Asn Tyr Ser Glu His Asn Ser Tyr  
 1665 1670 1675 1680

Val Ala Pro Arg Asn Glu Val Glu Glu Lys Leu Val Gln Ile Trp Ser  
 1685 1690 1695

Asn Ile Leu His Leu Pro Lys Val Gly Val Thr Glu Asn Phe Phe Ala  
 1700 1705  
 1710

Ile Gly Gly Asn Ser Leu Lys Ala Leu His Leu Ile Ser Gln Ile Glu  
 1715 1720 1725

Glu Leu Phe Ala Lys Glu Ile Ser Leu Ala Thr Leu Leu Thr Asn Pro  
 1730 1735 1740

Val Ile Ala Asp Leu Ala Lys Val Ile Gln Ala Asn Asn Gln Ile His  
 1745 1750 1755 1760  
 Asn Ser Pro Leu Val Pro Ile Gln Pro Gln Gly Lys Gln Gln Pro Phe  
 1765 1770  
 1775  
 Phe Cys Ile His Pro Ala Gly Gly His Val Leu Cys Tyr Phe Lys Leu  
 1780 1785 1790  
 Ala Gln Tyr Ile Gly Thr Asp Gln Pro Phe Tyr Gly Leu Gln Ala Gln  
 1795 1800 1805  
 Gly Phe Tyr Gly Asp Glu Ala Pro Leu Thr Arg Val Glu Asp Met Ala  
 1810 1815 1820  
 Ser Leu Tyr Val Lys Thr Ile Arg Glu Phe Gln Pro Gln Gly Pro Tyr  
 1825 1830 1835 1840  
 Arg Val Gly Gly Trp Ser Phe Gly Gly Val Val Ala Tyr Glu Val Ala  
 1845 1850 1855  
 Gln Gln Leu His Arg Gln Gly Gln Glu Val Ser Leu Leu Ala Ile Leu  
 1860 1865  
 1870  
 Asp Ser Tyr Val Pro Ile Leu Leu Asp Lys Gln Lys Pro Ile Asp Asp  
 1875 1880 1885  
 Val Tyr Leu Val Gly Val Leu Ser Arg Val Phe Gly Gly Met Phe Gly  
 1890 1895 1900  
 Gln Asp Asn Leu Val Thr Pro Glu Glu Ile Glu Asn Leu Thr Val Glu  
 1905 1910 1915 1920  
 Glu Lys Ile Asn Tyr Ile Ile Asp Lys Ala Arg Ser Ala Arg Ile Phe  
 1925 1930 1935  
 Pro Pro Gly Val Glu Arg Gln Asn Asn Arg Arg Ile Leu Asp Val Leu  
 1940 1945  
 1950  
 Val Gly Thr Leu Lys Ala Thr Tyr Ser Tyr Ile Arg Gln Pro Tyr Pro  
 1955 1960 1965  
 Gly Lys Val Thr Val Phe Arg Ala Arg Glu Lys His Ile Met Ala Pro  
 1970 1975 1980  
 Asp Pro Thr Leu Val Trp Val Glu Leu Phe Ser Val Met Ala Ala Gln  
 1985 1990 1995  
 2000  
 Glu Ile Lys Ile Ile Asp Val Pro Gly Asn His Tyr Ser Phe Val Leu  
 2005 2010  
 2015  
 Glu Pro His Val Gln Val Leu Ala Gln Arg Leu Gln Asp Cys Leu Glu

2020  
2030

2025

Asn Asn Ser  
2035

<210> 3  
<211> 4528  
<212> DNA  
<213> Nostoc punctiforme ATCC29133

<400> 3

```

atgagtaatg ttcaagcatc gtttgaagca acggaagctg aattccgcgt ggaagggttac 60
gaaaaaattg agtttagtct tgtctatgta aatggtgcat ttgatatcag taacagagaa 120
attgcagaca gctatgagaa gtttggccgt tgtctgactg tgattgatgc taatgtcaac 180
agattatatg gcaagcaa at caagtcatat tttagacact atggtattga tctgaccgta 240
gttcccattg tgattactga gcctactaaa acccttgcaa cctttgagaa aattgttgat 300
gctttttctg acttttggtt aatccgcaag gaaccagtat tagtagtggg tgggtggttta 360
accactgatg tagctgggtt tgcgtgtgct gcttaccgtc gtaagagtaa ctatattcgg 420
gttccgacaa cgctgattgg tctgattgat gcagggtgtag cgattaaggt tgcagtcaac 480
catcgcaagt taaaaaatcg cttgggtgca tatcatgctc ctttgaaagt catcctcgat 540
ttctccttct tgcaaacatt accaacggct caagttcgta atgggatggc agagttggtc 600
aaaattgctg ttgtggcgaa ctctgaagtc tttgaattgt tgtatgaata tggcgaagag 660
ttgctttcca ctacttttg atattgtgaat ggtacaaagg aactgaaagc gatcgacat 720
aaactcaatt acgaggcaat aaaaactatg ctggagttgg aaactccaaa cttgcatgag 780
ttagacctcg atcgcgtc at tgcctacggc cacacttgga gtccgacctt agaattagct 840
ccgatgatac cgttgtttca cggtcatgcc gtcaatatag acatggcttt gtctgcaact 900
attgcggcaa gacgtggcta cattacatca ggagaacgcg atcgcathtt gagcttgatg 960
agtcgtatag gtttatcaat cgatcatcct ctactagatg gcgatttgct ctggtatgct 1020
acccaatcta ttagcttgac gcgagacggg aaacaacgcg cagctatgcc taaacccatt 1080
ggtagagtgc tctttgtcaa cgatttcacc cgtgaagaac tagatgcagc tttagctgaa 1140
cacaaacgtc tgtgtgtctac atacctcgt ggtggagatg gcattgacgc ttacatagaa 1200
actcaagaag aatccaaact attgggagtg tgaaaacatg accagtatht taggacgaga 1260
taccgcaaga ccaataacgc cacatagcat tctggtagca cagctacaaa aaaccctcag 1320
aatggcagag gaaagtaata ttccttcaga gatactgact tctctgcgcc aagggttgca 1380

```

attagcagca ggttttagatc cctatctgga tgattgcact acccctgaat cgaccgcatt 1440  
 gacagcacta gcgcagaaga caagcattga agactggagt aaacgcttca gtgatgggtga 1500  
 aacagtgcgt caattagagc aagaaatgct ctcaggacat cttgaaggac aaacactaaa 1560  
 gatgtttgtg catatcacta aggctaaaag catcctagaa gtgggaatgt tcacaggata 1620  
 ttcagctttg gcaatggcag aggcggtacc agatgatggg cgactgattg cttgtgaagt 1680  
 agactcctat gtggccgagt ttgctcaaac ttgctttcaa gagtctcccc acggccgcaa 1740  
 gattgttgta gaagtggcac ctgcactaga gacgctgcac aagctgggtg ctaaaaaaga 1800  
 atcctttgat ctgatcttca ttgatgcgga taaaaaggag tatatagaat acttccaaat 1860  
 tatcttggat agccatttac tagctcccg cggattaatc tgtgtagata atactttgtt 1920  
 gcaaggacaa gtttacctgc catcagaaca gcgtacagcc aatgggtgaag cgatcgcctca 1980  
 attcaaccgc attgtcgccg cagatcctcg tgtagagcaa gttctgttac ccatacgaga 2040  
 tgggtataacc ctgattagac gcttgggtata agcggccgcg agctcctcga gatggcacia 2100  
 tcaatctctt tatctcttcc tcaatccaca acgcatcaa aggggtgtgag gctaaaaata 2160  
 gcagctctac tgaagactat cgggactcta atactactgc tgatagcctt gccgcttaat 2220  
 gctttgatag tattgatatc tctgatgtgt aggccgttta caaaaaaacc tgccgtagcc 2280  
 actcatcccc agaatatctt ggtcagtggc ggcaaaatga ccaaagcatt gcaacttgcc 2340  
 cgctccttcc atgcagccgg acacagagtt attctgattg aaggtcataa atactgggtta 2400  
 tcagggcata gattctcaaa ttctgtgagt cgtttttata cagttcctgc accacaagac 2460  
 gaccagaag gctataccca agcgcatttg gaaattgtca aacgagagaa gattgacgtt 2520  
 tatgtaccgg tatgcagccc tgtagctagt tattacgact ctttggcaaa gtctgcacta 2580  
 tcagaatatt gtgaggtttt tcactttgat gctgatataa ccaagatgct ggatgataaa 2640  
 tttgccttta ccgatcgggc gcgatcgctt ggtttatcag ccccgaaatc ttttaaaatt 2700  
 accgatcctg aacaagttat caacttcgat ttagtaaaag agacgcgcaa atatattctt 2760  
 aagagtatth cttacgactc agttcgccgc ttaaatttaa ccaaacttcc ttgtgatacc 2820  
 ccagaagaga cagcagcgtt tgtcaagagt ttacccatca gccagaaaa accttggatt 2880  
 atgcaagaat ttattcctgg gaaagaatta tgcaccata gcacagtccg agacggcgaa 2940  
 ttaagggtgc attgctgttc aaattcttca gcgtttcaga ttaactatga aaatgtcgaa 3000  
 aatccccaaa ttcaagaatg ggtacaacat ttctgcaaaa gtttacggct gactggacaa 3060  
 atatctcttg actttatcca agctgaagat ggtacagctt atgccattga atgtaatcct 3120

cgcacccatt cggcgatcac aatgttctac aatcaccag gtgttgaga agcctatctt 3180  
 ggtaaaactc ctctagctgc acctttggaa cctcttgag atagcaagcc cacttactgg 3240  
 atatatcacg aaatctggcg attgactggg attcgctctg gacaacaatt acaaacttgg 3300  
 tttgggagat tagtcagagg tacagatgcc atttatcgcc tggacgatcc gataccattht 3360  
 ttaactttgc accattggca gattacttta cttttgctac aaaatttgca acgactcaaa 3420  
 gggtgggtaa agatcgattht taatatcggt aaactcgtgg aattaggggg cgactaaaaa 3480  
 catgccagta cttaatatcc ttcatttagt tgggtctgca cacgataagt tttactgtga 3540  
 tttatcacgt ctttatgccc aagactgttht agctgcaaca gcagatccat cgctttataa 3600  
 ctttcaaatt gcatatatca caccgatcg gcagtggcga tttcctgact ctctcagtcg 3660  
 agaagatatt gctcttacca aaccgattht tgtgtttgat gccatacaat ttctaacagg 3720  
 ccaaaacatt gacatgatgt taccacaaat gttttgtatt cctggaatga ctcagtaccg 3780  
 tgccctattc gatctgctca agatccctta tataggaaat accccagata ttatggcgat 3840  
 cgcggccac aaagccagag ccaaagcaat tgtcgaagca gcaggggtaa aagtgcctcg 3900  
 tggagaattg cttcgccaag gagatattcc aacaattaca cctccagcag tcgtcaaacc 3960  
 tgtaagttct gacaactctt taggagtagt cttagttaaa gatgtgactg aatatgatgc 4020  
 tgccctaaag aaagcatttg aatatgcttc ggaggtcatc gtagaagcat tcatcgaact 4080  
 tggtcgagaa gtcagatgcg gcatcattgt aaaagacggt gagctaatag gtttaccctt 4140  
 tgaagagtat ctggtagacc cacacgataa acctatccgt aactatgctg ataaactcca 4200  
 acaaactgac gatggcgact tgcatttgac tgctaaagat aatatcaagg cttggatttht 4260  
 agaccctaac gacccaatca ccaaaaagg ttagcaagtg gctaaaagg gtcacaggc 4320  
 tttgggttgt cgccactaca gtttatthtga cttccgaatc gatccaaagg gacaaccttg 4380  
 gttcttagaa gctggattat attgttcttht tgccccaaa agtgtgattt cttctatggc 4440  
 gaaagcagcc ggaatccctc taaatgattt attaataacc gctattaatg aaacattggg 4500  
 tagtaataaa aaggtgttac aaaattga 4528

<210> 4  
 <211> 1496  
 <212> PRT  
 <213> Nostoc punctiforme ATCC29133

<400> 4  
 Met Ser Asn Val Gln Ala Ser Phe Glu Ala Thr Glu Ala Glu Phe Arg

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

|     |     |     |     |
|-----|-----|-----|-----|
|     | 245 |     | 250 |
|     | 255 |     |     |
| Asn | Leu | His | Glu |
| Leu | Asp | Leu | Asp |
| Arg | Val | Ile | Ala |
| Tyr | Gly | His | Thr |
|     | 260 |     | 265 |
|     | 270 |     |     |
| Trp | Ser | Pro | Thr |
| Leu | Glu | Leu | Ala |
| Pro | Met | Ile | Pro |
| Leu | Phe | His | Gly |
|     | 275 |     | 280 |
|     | 285 |     |     |
| His | Ala | Val | Asn |
| Ile | Asp | Met | Ala |
| Leu | Ser | Ala | Thr |
| Ile | Ala | Ala | Arg |
|     | 290 |     | 295 |
|     |     |     | 300 |
| Arg | Gly | Tyr | Ile |
| Thr | Ser | Gly | Glu |
| Arg | Asp | Arg | Ile |
| Leu | Ser | Leu | Met |
|     | 305 |     | 310 |
|     |     |     | 315 |
|     |     |     | 320 |
| Ser | Arg | Ile | Gly |
| Leu | Ser | Ile | Asp |
| His | Pro | Leu | Leu |
| Asp | Gly | Asp | Leu |
|     | 325 |     | 330 |
|     |     |     | 335 |
| Leu | Trp | Tyr | Ala |
| Thr | Gln | Ser | Ile |
| Ser | Leu | Thr | Arg |
| Asp | Gly | Lys | Gln |
|     | 340 |     | 345 |
|     | 350 |     |     |
| Arg | Ala | Ala | Met |
| Pro | Lys | Pro | Ile |
| Gly | Glu | Cys | Phe |
| Phe | Val | Asn | Asp |
|     | 355 |     | 360 |
|     | 365 |     |     |
| Phe | Thr | Arg | Glu |
| Glu | Leu | Asp | Ala |
| Ala | Leu | Ala | Glu |
| His | Lys | Arg | Leu |
|     | 370 |     | 375 |
|     |     |     | 380 |
| Cys | Ala | Thr | Tyr |
| Pro | Arg | Gly | Gly |
| Asp | Gly | Ile | Asp |
| Ala | Tyr | Ile | Glu |
|     | 385 |     | 390 |
|     |     |     | 395 |
|     |     |     | 400 |
| Thr | Gln | Glu | Glu |
| Ser | Lys | Leu | Leu |
| Gly | Val | Met | Thr |
| Ser | Ile | Leu | Gly |
|     | 405 |     | 410 |
|     |     |     | 415 |
| Arg | Asp | Thr | Ala |
| Arg | Pro | Ile | Thr |
| Pro | His | Ser | Ile |
| Leu | Val | Ala | Gln |
|     | 420 |     | 425 |
|     | 430 |     |     |
| Leu | Gln | Lys | Thr |
| Leu | Arg | Met | Ala |
| Glu | Glu | Ser | Asn |
| Ile | Pro | Ser | Glu |
|     | 435 |     | 440 |
|     | 445 |     |     |
| Ile | Leu | Thr | Ser |
| Leu | Arg | Gln | Gly |
| Leu | Gln | Leu | Ala |
| Ala | Gly | Leu | Asp |
|     | 450 |     | 455 |
|     |     |     | 460 |
| Pro | Tyr | Leu | Asp |
| Asp | Cys | Thr | Thr |
| Pro | Glu | Ser | Thr |
| Ala | Leu | Thr | Ala |
|     | 465 |     | 470 |
|     |     |     | 475 |
|     |     |     | 480 |
| Leu | Ala | Gln | Lys |
| Thr | Ser | Ile | Glu |
| Asp | Trp | Ser | Lys |
| Arg | Phe | Ser | Asp |
|     | 485 |     | 490 |
|     | 495 |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|-----|
| Gly | Glu | Thr | Val | Arg | Gln | Leu | Glu | Gln | Glu | Met | Leu | Ser | Gly | His | Leu |  |     |
|     |     |     |     |     |     |     | 500 |     |     |     |     |     | 505 |     |     |  |     |
|     |     |     |     |     |     |     | 510 |     |     |     |     |     |     |     |     |  |     |
| Glu | Gly | Gln | Thr | Leu | Lys | Met | Phe | Val | His | Ile | Thr | Lys | Ala | Lys | Ser |  |     |
|     |     |     |     |     |     |     | 515 |     |     |     |     |     | 520 |     |     |  |     |
|     |     |     |     |     |     |     | 525 |     |     |     |     |     |     |     |     |  |     |
| Ile | Leu | Glu | Val | Gly | Met | Phe | Thr | Gly | Tyr | Ser | Ala | Leu | Ala | Met | Ala |  |     |
|     | 530 |     |     |     |     |     |     |     |     |     | 535 |     |     |     |     |  | 540 |
| Glu | Ala | Leu | Pro | Asp | Asp | Gly | Arg | Leu | Ile | Ala | Cys | Glu | Val | Asp | Ser |  |     |
| 545 |     |     |     |     |     |     |     |     |     | 550 |     |     |     |     | 555 |  |     |
|     |     |     |     |     |     |     | 560 |     |     |     |     |     |     |     |     |  |     |
| Tyr | Val | Ala | Glu | Phe | Ala | Gln | Thr | Cys | Phe | Gln | Glu | Ser | Pro | His | Gly |  |     |
|     |     |     |     | 565 |     |     |     |     |     |     |     |     |     |     | 570 |  |     |
|     |     |     |     | 575 |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Arg | Lys | Ile | Val | Val | Glu | Val | Ala | Pro | Ala | Leu | Glu | Thr | Leu | His | Lys |  |     |
|     |     |     |     |     |     |     |     | 580 |     |     |     |     |     | 585 |     |  |     |
|     |     |     |     |     |     |     | 590 |     |     |     |     |     |     |     |     |  |     |
| Leu | Val | Ala | Lys | Lys | Glu | Ser | Phe | Asp | Leu | Ile | Phe | Ile | Asp | Ala | Asp |  |     |
|     |     |     |     |     |     |     | 595 |     |     |     |     |     | 600 |     |     |  |     |
|     |     |     |     |     |     |     | 605 |     |     |     |     |     |     |     |     |  |     |
| Lys | Lys | Glu | Tyr | Ile | Glu | Tyr | Phe | Gln | Ile | Ile | Leu | Asp | Ser | His | Leu |  |     |
|     | 610 |     |     |     |     |     |     |     |     |     | 615 |     |     |     |     |  | 620 |
| Leu | Ala | Pro | Asp | Gly | Leu | Ile | Cys | Val | Asp | Asn | Thr | Leu | Leu | Gln | Gly |  |     |
| 625 |     |     |     |     |     |     |     |     |     | 630 |     |     |     |     | 635 |  |     |
|     |     |     |     |     |     |     | 640 |     |     |     |     |     |     |     |     |  |     |
| Gln | Val | Tyr | Leu | Pro | Ser | Glu | Gln | Arg | Thr | Ala | Asn | Gly | Glu | Ala | Ile |  |     |
|     |     |     |     | 645 |     |     |     |     |     |     |     |     |     |     | 650 |  |     |
|     |     |     |     | 655 |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Ala | Gln | Phe | Asn | Arg | Ile | Val | Ala | Ala | Asp | Pro | Arg | Val | Glu | Gln | Val |  |     |
|     |     |     |     |     |     |     |     | 660 |     |     |     |     |     | 665 |     |  |     |
|     |     |     |     |     |     |     | 670 |     |     |     |     |     |     |     |     |  |     |
| Leu | Leu | Pro | Ile | Arg | Asp | Gly | Ile | Thr | Leu | Ile | Arg | Arg | Leu | Val | Met |  |     |
|     |     |     |     |     |     |     | 675 |     |     |     |     | 680 |     |     |     |  |     |
|     |     |     |     |     |     |     | 685 |     |     |     |     |     |     |     |     |  |     |
| Ala | Gln | Ser | Ile | Ser | Leu | Ser | Leu | Pro | Gln | Ser | Thr | Thr | Pro | Ser | Lys |  |     |
|     | 690 |     |     |     |     |     |     |     |     |     | 695 |     |     |     |     |  | 700 |
| Gly | Val | Arg | Leu | Lys | Ile | Ala | Ala | Leu | Leu | Lys | Thr | Ile | Gly | Thr | Leu |  |     |
| 705 |     |     |     |     | 710 |     |     |     |     |     | 715 |     |     |     |     |  | 720 |
| Ile | Leu | Leu | Leu | Ile | Ala | Leu | Pro | Leu | Asn | Ala | Leu | Ile | Val | Leu | Ile |  |     |
|     |     |     |     | 725 |     |     |     |     |     |     |     |     |     | 730 |     |  |     |
|     |     |     |     | 735 |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Ser | Leu | Met | Cys | Arg | Pro | Phe | Thr | Lys | Lys | Pro | Ala | Val | Ala | Thr | His |  |     |

[illegible]

## 41505

[illegible]

Ala Leu Phe Asp Leu Leu Lys Ile Pro Tyr Ile Gly Asn Thr Pro Asp  
 1250 1255 1260  
 Ile Met Ala Ile Ala Ala His Lys Ala Arg Ala Lys Ala Ile Val Glu  
 1265 1270 1275 1280  
 Ala Ala Gly Val Lys Val Pro Arg Gly Glu Leu Leu Arg Gln Gly Asp  
 1285 1290 1295  
 Ile Pro Thr Ile Thr Pro Pro Ala Val Val Lys Pro Val Ser Ser Asp  
 1300 1305  
 1310  
 Asn Ser Leu Gly Val Val Leu Val Lys Asp Val Thr Glu Tyr Asp Ala  
 1315 1320 1325  
 Ala Leu Lys Lys Ala Phe Glu Tyr Ala Ser Glu Val Ile Val Glu Ala  
 1330 1335 1340  
 Phe Ile Glu Leu Gly Arg Glu Val Arg Cys Gly Ile Ile Val Lys Asp  
 1345 1350 1355 1360  
 Gly Glu Leu Ile Gly Leu Pro Leu Glu Glu Tyr Leu Val Asp Pro His  
 1365 1370 1375  
 Asp Lys Pro Ile Arg Asn Tyr Ala Asp Lys Leu Gln Gln Thr Asp Asp  
 1380 1385  
 1390  
 Gly Asp Leu His Leu Thr Ala Lys Asp Asn Ile Lys Ala Trp Ile Leu  
 1395 1400  
 1405  
 Asp Pro Asn Asp Pro Ile Thr Gln Lys Val Gln Gln Val Ala Lys Arg  
 1410 1415 1420  
 Cys His Gln Ala Leu Gly Cys Arg His Tyr Ser Leu Phe Asp Phe Arg  
 1425 1430 1435 1440  
 Ile Asp Pro Lys Gly Gln Pro Trp Phe Leu Glu Ala Gly Leu Tyr Cys  
 1445 1450 1455  
 Ser Phe Ala Pro Lys Ser Val Ile Ser Ser Met Ala Lys Ala Ala Gly  
 1460 1465  
 1470  
 Ile Pro Leu Asn Asp Leu Leu Ile Thr Ala Ile Asn Glu Thr Leu Gly  
 1475 1480 1485  
 Ser Asn Lys Lys Val Leu Gln Asn  
 1490 1495

<210> 5  
 <211> 45  
 <212> DNA  
 <213> Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi xuôi cho Ava\_ABCD

&lt;400&gt;      5

acaatttcac acaggaaaga tatcatgagt atcgtccaag caaag  
45

&lt;210&gt;      6

&lt;211&gt;      45

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi ngược cho Ava\_ABCD

&lt;400&gt;      6

ctcatccgcc aaaacagctc tagattatga attattttcc agaca  
45

&lt;210&gt;      7

&lt;211&gt;      31

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi xuôi cho Npr\_ABCD

&lt;400&gt;      7

gatcgatatc atgagtaatg ttcaagcatc g  
31

&lt;210&gt;      8

&lt;211&gt;      36

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi ngược cho Npr\_ABCD

&lt;400&gt;      8

gatctctaga tcaattttgt aacacctttt tattac  
36

&lt;210&gt;      9

&lt;211&gt;      71

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

<223> mỗi xuôi cho catxet xóa aroD

<400> 9  
tcatgggggtt cggcgcctga caggctgacc gcgtgcagaa agggtaaaaa gctggagctg 60  
cttcgaagtt c 71

<210> 10  
<211> 79  
<212> DNA  
<213> Trình tự nhân tạo

<220>  
<223> mỗi ngược cho catxet xóa aroD

<400> 10  
atatattttt tagttcggcg gggaggggtgt tcccgccgaa atattattgc gccatgggtcc 60  
atatgaatat cctccttag 79

<210> 11  
<211> 23  
<212> DNA  
<213> Trình tự nhân tạo

<220>  
<223> mỗi xuôi cho xóa aroD

<400> 11  
caaagatttc cctctggaat atg 23

<210> 12  
<211> 21  
<212> DNA  
<213> Trình tự nhân tạo

<220>  
<223> mỗi ngược cho xóa aroD

<400> 12  
cagatgtgat tttccctacg c 21

<210> 13  
<211> 40  
<212> DNA  
<213> Trình tự nhân tạo

<220>  
<223> mỗi xuôi cho aroG

<400> 13  
 aaggagctca ctagtggtac ggatgctcct gttatggctcg 40

<210> 14  
 <211> 41  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho aroG

<400> 14  
 ctctcttcga caatttcac ttacccgcga cgcgctttta c 41

<210> 15  
 <211> 41  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho ppsA

<400> 15  
 gtaaaagcgc gtcgcgggta agtggaaatt gtcgaagaga g 41

<210> 16  
 <211> 42  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho ppsA

<400> 16  
 gccctagtagg atctgatggg ttatttcttc agttcagcca gg 42

<210> 17  
 <211> 42  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho tktA

<400> 17  
 cctggctgaa ctgaagaaat aacccatcag atccactagg gc 42

<210> 18

|       |                                                                                 |    |
|-------|---------------------------------------------------------------------------------|----|
| <211> | 42                                                                              |    |
| <212> | DNA                                                                             |    |
| <213> | Trình tự nhân tạo                                                               |    |
| <220> |                                                                                 |    |
| <223> | mỗi ngược cho tktA                                                              |    |
| <400> | 18                                                                              |    |
|       | cgacaccggt cagcaggggtg ttacagcagt tcttttgctt tc                                 | 42 |
| <210> | 19                                                                              |    |
| <211> | 20                                                                              |    |
| <212> | DNA                                                                             |    |
| <213> | Trình tự nhân tạo                                                               |    |
| <220> |                                                                                 |    |
| <223> | mỗi xuôi cho xác nhận của<br>pSKH130-deltafhuA-Pn-aroG-Pn-ppsA-Pn-tktA plasmid  |    |
| <400> | 19                                                                              |    |
|       | gaaaatgccg atgggtaccg                                                           | 20 |
| <210> | 20                                                                              |    |
| <211> | 26                                                                              |    |
| <212> | DNA                                                                             |    |
| <213> | Trình tự nhân tạo                                                               |    |
| <220> |                                                                                 |    |
| <223> | mỗi ngược cho xác nhận của<br>pSKH130-deltafhuA-Pn-aroG-Pn-ppsA-Pn-tktA plasmid |    |
| <400> | 20                                                                              |    |
|       | cgttagaacg cggctacaat taatac                                                    | 26 |
| <210> | 21                                                                              |    |
| <211> | 49                                                                              |    |
| <212> | DNA                                                                             |    |
| <213> | Trình tự nhân tạo                                                               |    |
| <220> |                                                                                 |    |
| <223> | mỗi xuôi cho Ava_ABCD                                                           |    |
| <400> | 21                                                                              |    |
|       | catctcacac cgagattatt ttagtactat gagtatcgtc caagcaaag                           | 49 |
| <210> | 22                                                                              |    |
| <211> | 46                                                                              |    |
| <212> | DNA                                                                             |    |
| <213> | Trình tự nhân tạo                                                               |    |

&lt;220&gt;

&lt;223&gt;      mỗi ngược cho Ava\_ABCD

&lt;400&gt;      22

caatttcgta tagagtctca ctagtaccgc aaaaaggcca tccgtc

46

&lt;210&gt;      23

&lt;211&gt;      43

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi xuôi cho đoạn gen khởi động Ptrc

&lt;400&gt;      23

tctcacaccg agattatctt agtcgactgc acggtgcacc aat

43

&lt;210&gt;      24

&lt;211&gt;      50

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi ngược cho đoạn gen khởi động Ptrc

&lt;400&gt;      24

tttgcttgga cgatactcat agtctgtttc ctgtgtgaaa ttgttatccg

50

&lt;210&gt;      25

&lt;211&gt;      43

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi xuôi cho đoạn gen khởi động PCJ1 hoặc PCJ2.2

&lt;400&gt;      25

tctcacaccg agattatctt agtcaccgcg ggcttattcc att

43

&lt;210&gt;      26

&lt;211&gt;      49

&lt;212&gt;      DNA

&lt;213&gt;      Trình tự nhân tạo

&lt;220&gt;

&lt;223&gt;      mỗi ngược cho đoạn gen khởi động PCJ1 hoặc PCJ2.2

&lt;400&gt;      26

ttgcttggaac gataactcata gtatcttaat ctcctagatt gggtttcac 49

<210> 27  
 <211> 50  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho đoạn gen khởi động PCJ2.2

<400> 27  
 ttgcttgga cgataactcat agtatcttaa tctcctagat ttggtttcac 50

<210> 28  
 <211> 22  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho xác nhận của tích hợp Ava-ABCD

<400> 28  
 cagaatgagt gaacaaccac gg 22

<210> 29  
 <211> 25  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho xác nhận của tích hợp Ava-ABCD

<400> 29  
 catatccaat aaccagaca aaacc 25

<210> 30  
 <211> 31  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho P<sub>trc</sub>/PCJ1-Ava\_AB hoặc P<sub>trc</sub>/PCJ1-Ava\_ABC

<400> 30  
 gatcggatcc atgagtatcg tccaagcaaa g 31

<210> 31  
 <211> 29  
 <212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi ngược cho Ptrc/PCJ1-Ava\_AB

<400> 31

gatcactagt ttaaggttgt attctgcgg

29

<210> 32

<211> 29

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi ngược cho Ptrc/PCJ1-Ava\_ABC

<400> 32

gatcactagt ctaatctccc cccaattcc

29

<210> 33

<211> 60

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi xuôi cho Ava\_ABCD

<400> 33

ccaaacacca acaaaaggct ctacccatat gatgagtatc gtccaagcaa agtttgaagc

60

<210> 34

<211> 60

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi ngược cho Ava\_ABCD

<400> 34

agagggtgac gctagtcaga actagtttat gaattatttt ccagacaatc ttgttaaagc

60

<210> 35

<211> 58

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi xuôi cho đoạn gen khởi động CJ7

<400> 35  
 ccaaacacca acaaaaggct ctacccatat gagaaacatc ccagcgctac taataggg 58

<210> 36  
 <211> 48  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho đoạn gen khởi động CJ7

<400> 36  
 caaactttgc ttggacgata ctcatagtgt ttccctttcgt tgggtacg 48

<210> 37  
 <211> 47  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho đoạn gen khởi động lysc8

<400> 37  
 caccaacaaa aggctctacc catatgaaca ctccctctggc taggtag 47

<210> 38  
 <211> 48  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho đoạn gen khởi động lysc8

<400> 38  
 caaactttgc ttggacgata ctcacttttg tgcacctttc gatctacg 48

<210> 39  
 <211> 54  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho đoạn gen khởi động O2

<400> 39  
 ccaaacacca acaaaaggct ctacccatat gcaataatcg tgaattttgg cagc 54

<210> 40  
 <211> 53

<212> DNA  
 <213> Trình tự nhân tạo  
  
 <220>  
 <223> mỗi ngược cho đoạn gen khởi động O2  
  
 <400> 40  
 caaactttgc ttggacgata ctcattgttt tgatctcctc caataatcta tgc 53  
  
 <210> 41  
 <211> 32  
 <212> DNA  
 <213> Trình tự nhân tạo  
  
 <220>  
 <223> mỗi xuôi cho các gen aroD để cấu trúc pDC delta aroD  
  
 <400> 41  
 gatcggatcc cactggacgt ttgggtgaga cc 32  
  
 <210> 42  
 <211> 37  
 <212> DNA  
 <213> Trình tự nhân tạo  
  
 <220>  
 <223> mỗi ngược cho các gen aroD để cấu trúc pDC delta aroD  
  
 <400> 42  
 gatcggatcc actagtttta ggttccattt ctaattg 37  
  
 <210> 43  
 <211> 29  
 <212> DNA  
 <213> Trình tự nhân tạo  
  
 <220>  
 <223> mỗi xuôi cho các gen aroD để cấu trúc pDC delta aroD  
  
 <400> 43  
 gatcactagt atgaacgaca gtattctcc 29  
  
 <210> 44  
 <211> 30  
 <212> DNA  
 <213> Trình tự nhân tạo  
  
 <220>  
 <223> mỗi ngược cho các gen aroD để cấu trúc pDC delta aroD

<400> 44  
 gatcaagctt gttacatcct gacgttgtgg 30

<210> 45  
 <211> 30  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho Ava\_A

<400> 45  
 cgcggatcca tgagtatcgt ccaagcaaag 30

<210> 46  
 <211> 33  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Ava\_A

<400> 46  
 cgcctcgagt tatttaacac tcccgattaa ttc 33

<210> 47  
 <211> 31  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho Ava\_B

<400> 47  
 cgcggatcca tgacaaatgt gattgtccaa c 31

<210> 48  
 <211> 35  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Ava\_B

<400> 48  
 cgcctcgagt taagggttgta ttctgcggat aattg 35

<210> 49

<211> 32  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho Ava\_C

<400> 49  
 cgcggatcca tggcacaatc ccttcccctt tc

32

<210> 50  
 <211> 31  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Ava\_C

<400> 50  
 cgccctcgagc taatctcccc ccaattccac c

31

<210> 51  
 <211> 34  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho Ava\_D

<400> 51  
 cgcggatcca tgcagactat agattttaat attc

34

<210> 52  
 <211> 33  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Ava\_D

<400> 52  
 cgccctcgagt tatgaattat tttccagaca atc  
 33

<210> 53  
 <211> 30  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>

<223> mỗi xuôi cho Npr\_A

<400> 53

cgcggatcca tgagtaatgt tcaagcatcg

30

<210> 54

<211> 29

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi ngược cho Npr\_A

<400> 54

cgccctcgagt cacactccca atagtttgg

29

<210> 55

<211> 31

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi xuôi cho Npr\_B

<400> 55

cgcggatcca tgaccagtat tttaggacga g

31

<210> 56

<211> 28

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi ngược cho Npr\_B

<400> 56

cgccctcgagt tataccaagc gtctaatac

28

<210> 57

<211> 32

<212> DNA

<213> Trình tự nhân tạo

<220>

<223> mỗi xuôi cho Npr\_C

<400> 57

cgcggatcca tggcacaatc aatctcttta tc

32

<210> 58  
 <211> 31  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Npr\_C

<400> 58  
 cgcctcgagt tagtcgcccc ctaattccac g

31

<210> 59  
 <211> 31  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho Npr\_D

<400> 59  
 cgcggatcca tgccagtact taatatacctt c

31

<210> 60  
 <211> 35  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho Npr\_D

<400> 60  
 cgcctcgagt caattttgta acaccttttt attac

35

<210> 61  
 <211> 48  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho pGPD

<400> 61  
 acgtaagata attgtatatt acgcagttta tcattatcaa tactcgcc

48

<210> 62  
 <211> 40  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223>      mỗi ngược cho pGPD

<400>      62  
 gtaacgacaa tcctctccat gtggtcttga atcaaagctg 40

<210>      63  
 <211>      39  
 <212>      DNA  
 <213>      Trình tự nhân tạo

<220>  
 <223>      mỗi xuôi cho LoxP(Ura)

<400>      63  
 ccgattgtca atcagcgtaa cagctgaagc ttcgtacgc 39

<210>      64  
 <211>      44  
 <212>      DNA  
 <213>      Trình tự nhân tạo

<220>  
 <223>      mỗi ngược cho LoxP(Ura)

<400>      64  
 ttgttttggc cagtcggcaa gggcataggc cactagtgga tctg 44

<210>      65  
 <211>      40  
 <212>      DNA  
 <213>      Trình tự nhân tạo

<220>  
 <223>      mỗi xuôi cho aroB

<400>      65  
 cagctttgat tcaagaccac atggagagga ttgtcgttac 40

<210>      66  
 <211>      39  
 <212>      DNA  
 <213>      Trình tự nhân tạo

<220>  
 <223>      mỗi ngược cho aroB

<400>      66  
 gcgtacgaag cttcagctgt tacgctgatt gacaatcgg 39

<210> 67  
 <211> 32  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho ARO1 đoạn 1

<400> 67  
 accattcaag agtatatgag agaaggaaaa ct

32

<210> 68  
 <211> 47  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho ARO1 đoạn 1

<400> 68  
 ggcgagtatt gataatgata aactgcgtaa tataacaatta tcttacg

47

<210> 69  
 <211> 44  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi xuôi cho ARO1 đoạn 2

<400> 69  
 cagatccact agtggcctat gcccttgccg actggcctaaa acaa

44

<210> 70  
 <211> 20  
 <212> DNA  
 <213> Trình tự nhân tạo

<220>  
 <223> mỗi ngược cho ARO1 đoạn 2

<400> 70  
 caaatatcaa tggcacgttg

20

<210> 71  
 <211> 759  
 <212> DNA  
 <213> Escherichia coli W3110

```

<400>      71
atgaaaaccg taactgtaaa agatctcgtc attggtacgg gcgcacctaa aatcatcgtc   60
tcgctgatgg cgaaagatat cgccagcgtg aaatccgaag ctctcgcta tcgtgaagcg  120
gactttgata ttctggaatg gcgtgtggac cactatgccg acctctccaa tgtggagtct  180
gtcatggcgg cagcaaaaat tctccgtgag accatgccag aaaaaccgct gctgtttacc  240
ttccgcagtg ccaaagaagg cggcgagcag gcgatttcca ccgaggctta tattgcactc  300
aatcgtgcag ccatcgacag cggcctggtt gatatgatcg atctggagtt atttaccggt  360
gatgatcagg ttaaagaaac cgtcgcttac gccacgcgc atgatgtgaa agtagtcatg  420
tccaaccatg acttcataa aacgccgga gccgaagaaa tcattgcccg tctgcgcaaa  480
atgcaatcct tcgacgccga tattcctaag attgcgctga tgccgcaaag taccagcgat  540
gtgctgacgt tgcttgccgc gaccctggag atgcaggagc agtatgccga tcgtccaatt  600
atcacgatgt cgatggcaaa aactggcgta atttctcgtc tggctggtga agtatttggc  660
tcggcggcaa cttttggtgc ggtaaaaaaa gcgtctgcgc cagggcaa atctcggtaaat  720
gatttgcgca cggtattaac tattttacac caggcataa                               759

```

```

<210>      72
<211>      252
<212>      PRT
<213>      Escherichia coli W3110

```

```

<400>      72
Met Lys Thr Val Thr Val Lys Asp Leu Val Ile Gly Thr Gly Ala Pro
  1                               5                               10
                               15
Lys Ile Ile Val Ser Leu Met Ala Lys Asp Ile Ala Ser Val Lys Ser
                               20                               25
                               30
Glu Ala Leu Ala Tyr Arg Glu Ala Asp Phe Asp Ile Leu Glu Trp Arg
                               35                               40
                               45
Val Asp His Tyr Ala Asp Leu Ser Asn Val Glu Ser Val Met Ala Ala
  50                               55                               60
Ala Lys Ile Leu Arg Glu Thr Met Pro Glu Lys Pro Leu Leu Phe Thr
  65                               70                               75
                               80
Phe Arg Ser Ala Lys Glu Gly Gly Glu Gln Ala Ile Ser Thr Glu Ala
                               85                               90
                               95

```

Tyr Ile Ala Leu Asn Arg Ala Ala Ile Asp Ser Gly Leu Val Asp Met  
 100 105

110

Ile Asp Leu Glu Leu Phe Thr Gly Asp Asp Gln Val Lys Glu Thr Val  
 115 120

125

Ala Tyr Ala His Ala His Asp Val Lys Val Val Met Ser Asn His Asp  
 130 135 140

Phe His Lys Thr Pro Glu Ala Glu Glu Ile Ile Ala Arg Leu Arg Lys  
 145 150 155

160

Met Gln Ser Phe Asp Ala Asp Ile Pro Lys Ile Ala Leu Met Pro Gln  
 165 170

175

Ser Thr Ser Asp Val Leu Thr Leu Leu Ala Ala Thr Leu Glu Met Gln  
 180 185

190

Glu Gln Tyr Ala Asp Arg Pro Ile Ile Thr Met Ser Met Ala Lys Thr  
 195 200

205

Gly Val Ile Ser Arg Leu Ala Gly Glu Val Phe Gly Ser Ala Ala Thr  
 210 215 220

Phe Gly Ala Val Lys Lys Ala Ser Ala Pro Gly Gln Ile Ser Val Asn  
 225 230 235

240

Asp Leu Arg Thr Val Leu Thr Ile Leu His Gln Ala  
 245 250

<210> 73  
 <211> 1053  
 <212> DNA  
 <213> Escherichia coli W3110

<400> 73  
 atgaattatc agaacgacga ttacgcac aaagaaatca aagagttact tcctcctgtc 60  
 gcattgctgg aaaaattccc cgctactgaa atgccgcga atacggttgc ccatgcccga 120  
 aaagcgatcc ataagatcct gaaaggtaat gatgatcgcc tgttggttgt gattggccca 180  
 tgctcaattc atgatcctgt cgcggcaaaa gagtatgcca ctgccttgct ggcgctgcgt 240  
 gaagagctga aagatgagct ggaaatcgta atgcgcgtct attttgaaaa gccgcgtacc 300  
 acggtgggct ggaaagggt gattaacgat ccgcatatgg ataatagctt ccagatcaac 360  
 gacggtctgc gtatagcccg taaattgctg cttgatatta acgacagcgg tctgccagcg 420

gcaggtgagt ttctcgatat gatcacccca caatatctcg ctgacctgat gagctggggc 480  
gcaattggcg cacgtaccac cgaatcgag gtgcaccgag aactggcatc agggctttct 540  
tgtccggtcg gcttcaaaaa tggcaccgac ggtacgatta aagtggctat cgatgccatt 600  
aatgccgccg gtgcgccgca ctgcttcctg tccgtaacga aatgggggca ttcggcgatt 660  
gtgaatacca gcggtaacgg cgattgccat atcattctgc gcggcggtaa agagcctaac 720  
tacagcgcgag agcacgttgc tgaagtgaag gaagggctga acaaagcagg cctgccagca 780  
caggtgatga tcgatttcag ccatgctaac tcgtccaaac aattcaaaaa gcagatggat 840  
gtttgtgctg acgtttgcca gcagattgcc ggtggcgaaa aggccattat tggcgtgatg 900  
gtggaaagcc atctggtgga aggcaatcag agcctcgaga gcggggagcc gctggcctac 960  
ggtaagagca tcaccgatgc ctgcatcggc tgggaagata ccgatgctct gttacgtcaa 1020  
ctggcgaatg cagtaaaagc gcgtcgcggg taa  
1053

<210> 74  
<211> 350  
<212> PRT  
<213> Escherichia coli W3110

<400> 74  
Met Asn Tyr Gln Asn Asp Asp Leu Arg Ile Lys Glu Ile Lys Glu Leu  
1 5 10  
15

Leu Pro Pro Val Ala Leu Leu Glu Lys Phe Pro Ala Thr Glu Asn Ala  
20 25  
30

Ala Asn Thr Val Ala His Ala Arg Lys Ala Ile His Lys Ile Leu Lys  
35 40  
45

Gly Asn Asp Asp Arg Leu Leu Val Val Ile Gly Pro Cys Ser Ile His  
50 55 60

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

Glu Glu Leu Lys Asp Glu Leu Glu Ile Val Met Arg Val Tyr Phe Glu  
85 90  
95

Lys Pro Arg Thr Thr Val Gly Trp Lys Gly Leu Ile Asn Asp Pro His  
100 105  
110

Met Asp Asn Ser Phe Gln Ile Asn Asp Gly Leu Arg Ile Ala Arg Lys

|                                                                 |     |     |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|-----|-----|
| 125                                                             |     | 115 |     | 120 |     |
| Leu Leu Leu Asp Ile Asn Asp Ser Gly Leu Pro Ala Ala Gly Glu Phe |     |     |     |     |     |
| 130                                                             |     |     | 135 |     | 140 |
| Leu Asp Met Ile Thr Pro Gln Tyr Leu Ala Asp Leu Met Ser Trp Gly |     |     |     |     |     |
| 145                                                             |     |     | 150 |     | 155 |
|                                                                 | 160 |     |     |     |     |
| Ala Ile Gly Ala Arg Thr Thr Glu Ser Gln Val His Arg Glu Leu Ala |     |     |     |     |     |
|                                                                 | 165 |     |     |     | 170 |
|                                                                 | 175 |     |     |     |     |
| Ser Gly Leu Ser Cys Pro Val Gly Phe Lys Asn Gly Thr Asp Gly Thr |     |     |     |     |     |
|                                                                 |     | 180 |     | 185 |     |
|                                                                 | 190 |     |     |     |     |
| Ile Lys Val Ala Ile Asp Ala Ile Asn Ala Ala Gly Ala Pro His Cys |     |     |     |     |     |
|                                                                 |     | 195 |     | 200 |     |
|                                                                 | 205 |     |     |     |     |
| Phe Leu Ser Val Thr Lys Trp Gly His Ser Ala Ile Val Asn Thr Ser |     |     |     |     |     |
| 210                                                             |     |     | 215 |     | 220 |
| Gly Asn Gly Asp Cys His Ile Ile Leu Arg Gly Gly Lys Glu Pro Asn |     |     |     |     |     |
| 225                                                             |     |     | 230 |     | 235 |
|                                                                 | 240 |     |     |     |     |
| Tyr Ser Ala Lys His Val Ala Glu Val Lys Glu Gly Leu Asn Lys Ala |     |     |     |     |     |
|                                                                 | 245 |     |     |     | 250 |
|                                                                 | 255 |     |     |     |     |
| Gly Leu Pro Ala Gln Val Met Ile Asp Phe Ser His Ala Asn Ser Ser |     |     |     |     |     |
|                                                                 |     | 260 |     | 265 |     |
|                                                                 | 270 |     |     |     |     |
| Lys Gln Phe Lys Lys Gln Met Asp Val Cys Ala Asp Val Cys Gln Gln |     |     |     |     |     |
|                                                                 |     | 275 |     | 280 |     |
|                                                                 | 285 |     |     |     |     |
| Ile Ala Gly Gly Glu Lys Ala Ile Ile Gly Val Met Val Glu Ser His |     |     |     |     |     |
| 290                                                             |     |     | 295 |     | 300 |
| Leu Val Glu Gly Asn Gln Ser Leu Glu Ser Gly Glu Pro Leu Ala Tyr |     |     |     |     |     |
| 305                                                             |     |     | 310 |     | 315 |
|                                                                 | 320 |     |     |     |     |
| Gly Lys Ser Ile Thr Asp Ala Cys Ile Gly Trp Glu Asp Thr Asp Ala |     |     |     |     |     |
|                                                                 | 325 |     |     |     | 330 |
|                                                                 | 335 |     |     |     |     |
| Leu Leu Arg Gln Leu Ala Asn Ala Val Lys Ala Arg Arg Gly         |     |     |     |     |     |
|                                                                 |     | 340 |     | 345 |     |
|                                                                 | 350 |     |     |     |     |
| <210>                                                           | 75  |     |     |     |     |

<211> 2379  
 <212> DNA  
 <213> Escherichia coli W3110

<400> 75  
 atgtccaaca atggctcgtc accgctggtg ctttgggtata accaactcgg catgaatgat 60  
 gtagacaggg ttgggggcaa aaatgcctcc ctgggtgaaa tgattactaa tctttccgga 120  
 atgggtgttt ccgttccgaa tggtttcgcc acaaccgccg acgcgtttaa ccagtttctg 180  
 gaccaaagcg gcgtaaacca gcgcatttat gaactgctgg ataaaacgga tattgacgat 240  
 gttactcagc ttgcgaaagc gggcgcgcaa atccgccagt ggattatcga cactcccttc 300  
 cagcctgagc tggaaaacgc catccgcgaa gcctatgcac agctttccgc cgatgacgaa 360  
 aacgcctctt ttgcggtgcg ctctccgcc accgcagaag atatgccgga cgcttctttt 420  
 gccggtcagc aggaaacctt cctcaacgtt cagggttttg acgccgttct cgtggcagtg 480  
 aaacatgtat ttgcttctct gtttaacgat cgcgccatct cttatcgtgt gcaccaggg 540  
 tacgatcacc gtggtgtggc gctctccgcc ggtgttcaac ggatggtgcg ctctgacctc 600  
 gcatcatctg gcgtgatgtt ctccattgat accgaatccg gctttgacca ggtggtgttt 660  
 atcacttccg catggggcct tggtgagatg gtcgtgcagg gtgcggttaa cccggatgag 720  
 ttttacgtgc ataaaccgac actggcggcg aatcgcccgg ctatcgtgcg ccgcaccatg 780  
 gggtcgaaaa aaatccgcat ggtttacgcg ccgaccagc agcacggcaa gcagggttaa 840  
 atcgaagacg taccgcagga acagcgtgac atcttctcgc tgaccaacga agaagtgcag 900  
 gaactggcaa aacaggccgt acaaattgag aaacactacg gtcgcccgat ggatattgag 960  
 tgggcgaaaag atggccacac cggtaaactg ttcatgtgac aggcgcgtcc ggaaaccgtg 1020  
 cgctcacgcg gtcaggatcat ggagcgttat acgctgcatt cacagggtaa gattatcgcc 1080  
 gaaggccgtg ctatcgttca tcgcatcggg gcgggtccgg tgaaagtcat ccatgacatc 1140  
 agcgaaatga accgcatcga acctggcgac gtgctgggta ctgacatgac cgaccgggac 1200  
 tgggaaccga tcatgaagaa agcatctgcc atcgtcacca accgtggcgg tcgtacctgt 1260  
 cacgcggcga tcatcgctcg tgaactgggc attccggcgg tagtgggctg tggagatgca 1320  
 acagaacgga tgaaagacgg tgagaacgtc actgtttctt gtgccgaagg tgataccggt 1380  
 tacgtctatg cggagttgct ggaatttagc gtgaaaagct ccagcgtaga aacgatgccg 1440  
 gatctgccgt tgaaagtgat gatgaacgtc ggtaaccgg accgtgcttt cgacttcgcc 1500  
 tgccctaccga acgaaggcgt gggccttgcg cgtctggaat ttatcatcaa ccgtatgatt 1560  
 ggcgtocacc cacgcgcact gcttgagttt gacgatcagg aaccgcagtt gcaaaacgaa 1620

```

atccgcgaga tgatgaaagg ttttgattct ccgcgtgaat ttacgttg tcgtctgact 1680
gaagggatcg cgacgctggg tgccgcgttt tatccgaagc gcgtcattgt ccgtctctct 1740
gattttaaat cgaacgaata tgccaacctg gtcgggtggg agcgttacga gccagatgaa 1800
gagaacccga tgctcggcct ccgtggcgcg ggccgctatg tttccgacag cttccgcgac 1860
tgtttcgcgc tggagtgtga agcagtgaaa cgtgtgcgca acgacatggg actgaccaac 1920
gttgagatca tgatcccgtt cgtgcgtacc gtagatcagg cgaaagcggg ggttgaagaa 1980
ctggcgcgctc aggggctgaa acgtggcgag aacgggctga aaatcatcat gatgtgtgaa 2040
atcccgtcca acgccttgct ggccgagcag ttctctgaat atttcgacgg cttctcaatt 2100
ggctcaaacg atatgacgca gctggcgctc ggtctggacc gtgactccgg cgtggtgtct 2160
gaattgttcg atgagcgcaa cgatgcggtg aaagcactgc tgcgatggc tatccgtgcc 2220
gcgaagaaac agggcaaata tgcgggatt tgcggtcagg gtccgtccga ccacgaagac 2280
tttgccgcat ggttgatgga agaggggatc gatagcctgt ctctgaaccc ggacaccgtg 2340
gtgcaaact ggtaagcct ggctgaactg aagaaataa 2379

```

```

<210>      76
<211>      792
<212>      PRT
<213>      Escherichia coli W3110

```

```

<400>      76
Met Ser Asn Asn Gly Ser Ser Pro Leu Val Leu Trp Tyr Asn Gln Leu
  1                               5                      10
                        15

```

```

Gly Met Asn Asp Val Asp Arg Val Gly Gly Lys Asn Ala Ser Leu Gly
                        20                      25
                30

```

```

Glu Met Ile Thr Asn Leu Ser Gly Met Gly Val Ser Val Pro Asn Gly
                        35                      40
                45

```

```

Phe Ala Thr Thr Ala Asp Ala Phe Asn Gln Phe Leu Asp Gln Ser Gly
  50                               55                      60

```

```

Val Asn Gln Arg Ile Tyr Glu Leu Leu Asp Lys Thr Asp Ile Asp Asp
  65                               70                      75
                80

```

```

Val Thr Gln Leu Ala Lys Ala Gly Ala Gln Ile Arg Gln Trp Ile Ile
                        85                      90
                95

```

```

Asp Thr Pro Phe Gln Pro Glu Leu Glu Asn Ala Ile Arg Glu Ala Tyr

```

[illegible]

## 41505

[illegible]

590 580 585  
 Gly Glu Arg Tyr Glu Pro Asp Glu Glu Asn Pro Met Leu Gly Phe Arg  
 595 600  
 605  
 Gly Ala Gly Arg Tyr Val Ser Asp Ser Phe Arg Asp Cys Phe Ala Leu  
 610 615 620  
 Glu Cys Glu Ala Val Lys Arg Val Arg Asn Asp Met Gly Leu Thr Asn  
 625 630 635  
 640  
 Val Glu Ile Met Ile Pro Phe Val Arg Thr Val Asp Gln Ala Lys Ala  
 645 650  
 655  
 Val Val Glu Glu Leu Ala Arg Gln Gly Leu Lys Arg Gly Glu Asn Gly  
 660 665  
 670  
 Leu Lys Ile Ile Met Met Cys Glu Ile Pro Ser Asn Ala Leu Leu Ala  
 675 680  
 685  
 Glu Gln Phe Leu Glu Tyr Phe Asp Gly Phe Ser Ile Gly Ser Asn Asp  
 690 695 700  
 Met Thr Gln Leu Ala Leu Gly Leu Asp Arg Asp Ser Gly Val Val Ser  
 705 710 715  
 720  
 Glu Leu Phe Asp Glu Arg Asn Asp Ala Val Lys Ala Leu Leu Ser Met  
 725 730  
 735  
 Ala Ile Arg Ala Ala Lys Lys Gln Gly Lys Tyr Val Gly Ile Cys Gly  
 740 745  
 750  
 Gln Gly Pro Ser Asp His Glu Asp Phe Ala Ala Trp Leu Met Glu Glu  
 755 760  
 765  
 Gly Ile Asp Ser Leu Ser Leu Asn Pro Asp Thr Val Val Gln Thr Trp  
 770 775 780  
 Leu Ser Leu Ala Glu Leu Lys Lys  
 785 790

<210> 77  
 <211> 1992  
 <212> DNA  
 <213> Escherichia coli W3110  
 <400> 77

atgtcctcac gtaaagagct tgccaatgct attcgtgcgc tgagcatgga cgcagtacag 60  
 aaagccaaat ccggtcacco ggggtgcccct atgggtatgg ctgacattgc cgaagtccctg 120  
 tggcgtgatt tcctgaaaca caaccgcag aatccgtcct gggctgaccg tgaccgcttc 180  
 gtgctgtcca acggccacgg ctccatgctg atctacagcc tgctgcacct caccggttac 240  
 gatctgccga tggaagaact gaaaaacttc cgtcagctgc actctaaaac tccgggtcac 300  
 ccggaagtgg gttacaccgc tgggtgtggaa accaccaccg gtccgctggg tcaggggtatt 360  
 gccaacgcag tcggtatggc gattgcagaa aaaacgctgg cggcgcagtt taaccgtccg 420  
 ggccacgaca ttgtcgacca ctacacctac gccttcattg gcgacggctg catgatggaa 480  
 ggcattctccc acgaagtttg ctctctggcg ggtacgctga agctgggtaa actgattgca 540  
 ttctacgatg acaacggtat ttctatcgat ggtcacgttg aaggctggtt caccgacgac 600  
 accgcaatgc gtttcgaagc ttacggctgg cacgttattc gcgacatcga cggtcatgac 660  
 gcggcatcta tcaaacgcgc agtagaagaa gcgcgcgcag tgactgacaa accttcctctg 720  
 ctgatgtgca aaaccatcat cggtttcggt tccccgaaca aagccggtac ccacgactcc 780  
 cacggtgcgc cgctgggcga cgctgaaatt gccctgacct gcgaacaact gggctggaaa 840  
 tatgcgcctg tcgaaatccc gtctgaaatc tatgctcagt gggatgcgaa agaagcaggc 900  
 caggcgaaag aatccgcatg gaacgagaaa ttcgctgctt acgcgaaagc ttatccgcag 960  
 gaagccgctg aatttaccg ccgtatgaaa ggcgaaatgc cgtctgactt cgacgctaaa 1020  
 gcgaaagagt tcatcgctaa actgcaggct aatccggcga aaatcgccag ccgtaaagcg 1080  
 tctcagaatg ctatcgaagc gttcgggtccg ctggtgccgg aattcctcgg cggttctgct 1140  
 gacctggcgc cgtctaacct gacctgtgg tctggttcta aagcaatcaa cgaagatgct 1200  
 gcgggtaact acatccacta cgggtgttcgc gagttcggta tgaccgcgat tgctaacggg 1260  
 atctccctgc acggtggctt cctgcggtac acctccacct tcctgatggt cgtggaatac 1320  
 gcacgtaacg ccgtacgtat ggctgcgctg atgaaacagc gtcaggatgat ggtttacacc 1380  
 cagcactcca tcggtctggg cgaagacggc ccgactcacc agccggttga gcaggctcgt 1440  
 tctctgcgcg taaccccgaa catgtctaca tggcgtccgt gtgaccagggt tgaatccgcg 1500  
 gtcgcgtgga aatacgggtg tgagcgtcag gacggcccga ccgcactgat cctctcccgt 1560  
 cagaacctgg cgcagcagga acgaactgaa gagcaactgg caaacatcgc gcgcgggtgg 1620  
 tatgtgctga aagactgcgc cggtcagccg gaactgattt tcatcgctac cggttcagaa 1680  
 gttgaactgg ctggtgctgc ctacgaaaaa ctgactgccg aaggcgtgaa agcgcgcgtg 1740

gtgtccatgc cgtctaccga cgcatttgac aagcaggatg ctgcttaccg tgaatccgta 1800  
 ctgccgaaag cggttactgc acgcgttgct gtagaagcgg gtattgctga ctactggtac 1860  
 aagtatgttg gcctgaacgg tgctatcgtc ggtatgacca ccttcggtga atctgctccg 1920  
 gcagagctgc tgtttgaaga gttcggcttc actgttgata acgttggtgc gaaagcaaaa 1980  
 gaactgctgt aa 1992

<210> 78  
 <211> 663  
 <212> PRT  
 <213> Escherichia coli W3110

<400> 78  
 Met Ser Ser Arg Lys Glu Leu Ala Asn Ala Ile Arg Ala Leu Ser Met  
 1 5 10  
 15

Asp Ala Val Gln Lys Ala Lys Ser Gly His Pro Gly Ala Pro Met Gly  
 20 25  
 30

Met Ala Asp Ile Ala Glu Val Leu Trp Arg Asp Phe Leu Lys His Asn  
 35 40  
 45

Pro Gln Asn Pro Ser Trp Ala Asp Arg Asp Arg Phe Val Leu Ser Asn  
 50 55 60

Gly His Gly Ser Met Leu Ile Tyr Ser Leu Leu His Leu Thr Gly Tyr  
 65 70 75  
 80

Asp Leu Pro Met Glu Glu Leu Lys Asn Phe Arg Gln Leu His Ser Lys  
 85 90  
 95

Thr Pro Gly His Pro Glu Val Gly Tyr Thr Ala Gly Val Glu Thr Thr  
 100 105  
 110

Thr Gly Pro Leu Gly Gln Gly Ile Ala Asn Ala Val Gly Met Ala Ile  
 115 120  
 125

Ala Glu Lys Thr Leu Ala Ala Gln Phe Asn Arg Pro Gly His Asp Ile  
 130 135 140

Val Asp His Tyr Thr Tyr Ala Phe Met Gly Asp Gly Cys Met Met Glu  
 145 150 155  
 160

Gly Ile Ser His Glu Val Cys Ser Leu Ala Gly Thr Leu Lys Leu Gly  
 165 170  
 175

```

Lys Leu Ile Ala Phe Tyr Asp Asp Asn Gly Ile Ser Ile Asp Gly His
      180                               185
      190

Val Glu Gly Trp Phe Thr Asp Asp Thr Ala Met Arg Phe Glu Ala Tyr
      195                               200
      205

Gly Trp His Val Ile Arg Asp Ile Asp Gly His Asp Ala Ala Ser Ile
      210                               215
      220

Lys Arg Ala Val Glu Glu Ala Arg Ala Val Thr Asp Lys Pro Ser Leu
      225                               230
      235
      240

Leu Met Cys Lys Thr Ile Ile Gly Phe Gly Ser Pro Asn Lys Ala Gly
      245                               250
      255

Thr His Asp Ser His Gly Ala Pro Leu Gly Asp Ala Glu Ile Ala Leu
      260                               265
      270

Thr Arg Glu Gln Leu Gly Trp Lys Tyr Ala Pro Phe Glu Ile Pro Ser
      275                               280
      285

Glu Ile Tyr Ala Gln Trp Asp Ala Lys Glu Ala Gly Gln Ala Lys Glu
      290                               295
      300

Ser Ala Trp Asn Glu Lys Phe Ala Ala Tyr Ala Lys Ala Tyr Pro Gln
      305                               310
      315
      320

Glu Ala Ala Glu Phe Thr Arg Arg Met Lys Gly Glu Met Pro Ser Asp
      325                               330
      335

Phe Asp Ala Lys Ala Lys Glu Phe Ile Ala Lys Leu Gln Ala Asn Pro
      340                               345
      350

Ala Lys Ile Ala Ser Arg Lys Ala Ser Gln Asn Ala Ile Glu Ala Phe
      355                               360
      365

Gly Pro Leu Leu Pro Glu Phe Leu Gly Gly Ser Ala Asp Leu Ala Pro
      370                               375
      380

Ser Asn Leu Thr Leu Trp Ser Gly Ser Lys Ala Ile Asn Glu Asp Ala
      385                               390
      395
      400

Ala Gly Asn Tyr Ile His Tyr Gly Val Arg Glu Phe Gly Met Thr Ala
      405                               410
      415

```

Ile Ala Asn Gly Ile Ser Leu His Gly Gly Phe Leu Pro Tyr Thr Ser  
 420 425  
 430  
 Thr Phe Leu Met Phe Val Glu Tyr Ala Arg Asn Ala Val Arg Met Ala  
 435 440  
 445  
 Ala Leu Met Lys Gln Arg Gln Val Met Val Tyr Thr His Asp Ser Ile  
 450 455 460  
 Gly Leu Gly Glu Asp Gly Pro Thr His Gln Pro Val Glu Gln Val Ala  
 465 470 475  
 480  
 Ser Leu Arg Val Thr Pro Asn Met Ser Thr Trp Arg Pro Cys Asp Gln  
 485 490  
 495  
 Val Glu Ser Ala Val Ala Trp Lys Tyr Gly Val Glu Arg Gln Asp Gly  
 500 505  
 510  
 Pro Thr Ala Leu Ile Leu Ser Arg Gln Asn Leu Ala Gln Gln Glu Arg  
 515 520  
 525  
 Thr Glu Glu Gln Leu Ala Asn Ile Ala Arg Gly Gly Tyr Val Leu Lys  
 530 535 540  
 Asp Cys Ala Gly Gln Pro Glu Leu Ile Phe Ile Ala Thr Gly Ser Glu  
 545 550 555  
 560  
 Val Glu Leu Ala Val Ala Ala Tyr Glu Lys Leu Thr Ala Glu Gly Val  
 565 570  
 575  
 Lys Ala Arg Val Val Ser Met Pro Ser Thr Asp Ala Phe Asp Lys Gln  
 580 585  
 590  
 Asp Ala Ala Tyr Arg Glu Ser Val Leu Pro Lys Ala Val Thr Ala Arg  
 595 600  
 605  
 Val Ala Val Glu Ala Gly Ile Ala Asp Tyr Trp Tyr Lys Tyr Val Gly  
 610 615 620  
 Leu Asn Gly Ala Ile Val Gly Met Thr Thr Phe Gly Glu Ser Ala Pro  
 625 630 635  
 640  
 Ala Glu Leu Leu Phe Glu Glu Phe Gly Phe Thr Val Asp Asn Val Val  
 645 650  
 655  
 Ala Lys Ala Lys Glu Leu Leu

660

<210> 79  
 <211> 438  
 <212> DNA  
 <213> Corynebacterium glutamicum ATCC 13032

<400> 79  
 atgcctggaa aaattctcct cctcaacggc ccaaacctga acatgctggg caaacgcgag 60  
 cctgacattt acggacacga caccttgga gacgtcgtcg cgctggcaac cgctgaggct 120  
 gcgaagcacg gccttgaggt tgaggcgctg cagagcaatc acgaaggtga gctaatacgat 180  
 gcgctgcaca acgctcgcgg caccacatc ggttgctgta ttaaccccg cggcctgact 240  
 cacacttcgg tggcgctttt ggatgctgtg aaggcgctctg agcttcctac cgctgagggtg 300  
 cacatttcca atccgcatgc ccgtgaagag ttccgccacc attcttacat ttccctcgcc 360  
 gcggtctccg ttatcgctgg cgctggcatc cagggttacc gtttcgcggt cgatatcctg 420  
 gcaaattctca aaaagtag 438

<210> 80  
 <211> 145  
 <212> PRT  
 <213> Corynebacterium glutamicum ATCC 13032

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

His His Ser Tyr Ile Ser Leu Ala Ala Val Ser Val Ile Ala Gly Ala  
 115 120  
 125

Gly Ile Gln Gly Tyr Arg Phe Ala Val Asp Ile Leu Ala Asn Leu Lys  
 130 135 140

Lys  
 145

<210> 81  
 <211> 4767  
 <212> DNA  
 <213> *Saccharomyces cerevisiae*

<400> 81  
 atggtgcagt tagccaaagt cccaattcta ggaaatgata ttatccacgt tgggtataac 60  
 attcatgacc atttggttga aaccataatt aaacattgtc cttcttcgac atacgttatt 120  
 tgcaatgata cgaacttgag taaagttcca tactaccagc aattagtcct ggaattcaag 180  
 gcttctttgc cagaaggctc tcgtttactt acttatgttg ttaaaccagg tgagacaagt 240  
 aaaagtagag aaaccaaagc gcagctagaa gattatcttt tagtggaagg atgtactcgt 300  
 gatacggtta tggtagcgat cggtggtggt gttattggtg acatgattgg gttcgttgca 360  
 tctacattta tgagaggtgt tcgtgttgtc caagtaccaa catccttatt ggcaatggtc 420  
 gattcctcca ttggtggtaa aactgctatt gacactcctc taggtaaaaa ctttattggt 480  
 gcattttggc aaccaaatt tgtccttgta gatattaaat ggctagaaac gttagccaag 540  
 agagagttha tcaatgggat ggcagaagtt atcaagactg cttgtatttg gaacgctgac 600  
 gaatttacta gattagaatc aaacgcttcg ttgttcttaa atgttggtta tggggcaaaa 660  
 aatgtcaagg ttaccaatca attgacaaac gagattgacg agatatcgaa tacagatatt 720  
 gaagctatgt tggatcatac atataagtta gttcttgaga gtattaaggt caaagcggaa 780  
 gttgtctctt cggatgaacg tgaatccagt ctaagaaacc ttttgaactt cggacattct 840  
 attggtcatg cttatgaagc tatactaacc ccacaagcat tacatggtga atgtgtgtcc 900  
 attggtatgg ttaaagaggc ggaattatcc cgttatttcg gtattctctc ccctacccaa 960  
 gttgcacgtc tatccaagat tttggttgcc tacgggttgcc ctgtttcgcc tgatgagaaa 1020  
 tggtttaaag agctaacctt acataagaaa acaccattgg atatcttatt gaagaaaatg 1080  
 agtattgaca agaaaaacga gggttccaaa aagaagggtg tcattttaga aagtattggt 1140  
 aagtgctatg gtgactccgc tcaatttggt agcgatgaag acctgagatt tattctaaca 1200

gatgaaaccc tcgtttaccc cttcaaggac atccctgctg atcaacagaa agttgttata 1260  
ccccctggtt ctaagtccat ctccaatcgt gctttaattc ttgctgccct cggtgaagggt 1320  
caatgtaaaa tcaagaactt attacattct gatgatacta aacatatggt aaccgctggt 1380  
catgaattga aagggtgctac gatatcatgg gaagataatg gtgagacggt agtggtggaa 1440  
ggacatggtg gttccacatt gtcagcttgt gctgaccctt tatatctagg taatgcagggt 1500  
actgcatcta gatttttgac ttccttgggt gccttgggtca attctacttc aagccaaaag 1560  
tatatcggtt taactggtaa cgcaagaatg caacaaagac caattgctcc tttggtcgat 1620  
tctttgcgtg ctaatggtac taaaattgag tacttgaata atgaagggtc cctgccaatc 1680  
aaagtttata ctgattcgggt attcaaagggt ggtagaattg aattagctgc tacagtttct 1740  
tctcagtacg tatcctctat cttgatgtgt gcccatacgt ctgaagaacc tgtaactttg 1800  
gctcttggtg gtggttaagcc aatctctaaa ttgtacgtcg atatgacaat aaaaatgatg 1860  
gaaaaattcg gtatcaatgt tgaaacttct actacagaac cttacactta ttatattcca 1920  
aaggagacatt atattaaccc atcagaatac gtcattgaaa gtgatgcctc aagtgtaca 1980  
taccattggt ccttcgccgc aatgactgggt actaccgtaa cggttccaaa cattggtttt 2040  
gagtcgttac aagggtgatgc cagatttgca agagatgtct tgaaacctat ggggtgtaaa 2100  
ataactcaaa cggcaacttc aactactggt tcgggtcctc ctgtagggtac tttaaagcca 2160  
ttaaaacatg ttgatatgga gccaatgact gatgcgttct taactgcatg tggtgttgcc 2220  
gctatttcgc acgacagtga tccaaattct gcaaatacaa ccaccattga aggtattgca 2280  
aaccagcgtg tcaaagagtg taacagaatt ttggccatgg ctacagagct cgccaaattt 2340  
ggcgtcaaaa ctacagaatt accagatgggt attcaagtcc atgggtttaaa ctcgataaaa 2400  
gatttgaagg ttccttccga ctcttctgga cctgtcgggtg tatgcacata tgatgatcat 2460  
cgtgtggcca tgagtttctc gcttcttgca ggaatggtaa attctcaaaa tgaacgtgac 2520  
gaagttgcta atcctgtaag aatacttgaa agacattgta ctggtaaaac ctggcctggc 2580  
tggtgggatg tgttacattc cgaactagggt gccaaattag atgggtgcaga acctttagag 2640  
tgacatcca aaaagaactc aaagaaaagc gttgtcatta ttggcatgag agcagctggc 2700  
aaaactacta taagtaaatt gtgcgcatcc gctctggggt acaaattagt tgacctagac 2760  
gagctgtttg agcaacagca taacaatcaa agtggttaaac aatttggtgt ggagaacggt 2820  
tgggagaagt tccgtgagga agaaacaaga attttcaagg aagttattca aaattacggc 2880  
gatgatggat atgttttctc aacagggtggc ggtattgttg aaagcgtgta gtctagaaaa 2940

gccttaaaag attttgctc atcaggtgga tacgttttac acttacatag ggatattgag 3000  
 gagacaattg tctttttaca aagtgatcct tcaagacctg cctatgtgga agaaattcgt 3060  
 gaagtttgga acagaaggga ggggtggtat aaagaatgct caaattttctc tttctttgct 3120  
 cctcattgct cgcagaagc tgagttccaa gctctaagaa gatcgtttag taagtacatt 3180  
 gcaaccatta caggtgtcag agaaatagaa attccaagcg gaagatctgc ctttgtgtgt 3240  
 ttaacctttg atgacttaac tgaacaaact gagaatttga ctccaatctg ttatggttgt 3300  
 gaggtgtag aggtcagagt agaccatttg gctaattact ctgctgattt cgtgagtaaa 3360  
 cagttatcta tattgcgtaa agccactgac agtattccta tcatttttac tgtgcgaacc 3420  
 atgaagcaag gtggcaactt tcctgatgaa gagttcaaaa ccttgagaga gctatacgat 3480  
 attgccttga agaatggtgt tgaattcctt gacttagaac taactttacc tactgatatc 3540  
 caatatgagg ttattaacaa aaggggcaac accaagatca ttggttccca tcatgacttc 3600  
 caaggattat actcctggga cgacgctgaa tgggaaaaca gattcaatca agcgttaact 3660  
 cttgatgtgg atgttgtaaa atttgtgggt acggctgtta atttcgaaga taatttgaga 3720  
 ctggaacact ttagggatac acacaagaat aagcctttaa ttgcagttaa tatgacttct 3780  
 aaaggtagca tttctcgtgt tttgaataat gttttaacac ctgtgacatc agatttattg 3840  
 cctaactccg ctgcccctgg ccaattgaca gtagcacaaa ttaacaagat gtatacatct 3900  
 atgggaggta tcgagcctaa ggaactgttt gttgttggaag agccaattgg ccactctaga 3960  
 tcgccaattt tacataacac tggctatgaa attttaggtt tacctcaciaa gttcgataaa 4020  
 tttgaaactg aatccgcaca attggtgaaa gaaaaacttt tggacggaaa caagaacttt 4080  
 ggcggtgctg cagtcacaat tcctctgaaa ttagatataa tgcagtacat ggatgaattg 4140  
 actgatgctg ctaaagttat tgggtgctgta aacacagtta taccattggg taacaagaag 4200  
 ttttaagggtg ataataccga ctgggttaggt atccgtaatg ccttaattaa caatggcggt 4260  
 cccgaatatg ttggtcatac cgctgggtttg gttatcggtg caggtggcac ttctagagcc 4320  
 gccctttacg ccttgcacag tttaggttgc aaaaagatct tcataatcaa caggacaact 4380  
 tcgaaattga agccattaat agagtcactt ccatctgaat tcaacattat tggaatagag 4440  
 tccactaaat ctatagaaga gattaaggaa cacgttggcg ttgctgtcag ctgtgtacca 4500  
 gccgacaaac cattagatga cgaactttta agtaagctgg agagattcct tgtgaaagggt 4560  
 gcccatgctg cttttgtacc aaccttattg gaagccgcat acaaaccaag cgttactccc 4620  
 gttatgacaa tttcacaaga caaatatcaa tggcacgttg tccctggatc acaaatgtta 4680

gtacaccaag gtgtagctca gtttgaaaag tggacaggat tcaagggccc tttcaaggcc 4740

atTTTTgatg ccggttacgaa agagtag 4767

<210> 82  
 <211> 1588  
 <212> PRT  
 <213> *Saccharomyces cerevisiae*

<400> 82  
 Met Val Gln Leu Ala Lys Val Pro Ile Leu Gly Asn Asp Ile Ile His  
 1 5 10

15

Val Gly Tyr Asn Ile His Asp His Leu Val Glu Thr Ile Ile Lys His  
 20 25

30

Cys Pro Ser Ser Thr Tyr Val Ile Cys Asn Asp Thr Asn Leu Ser Lys  
 35 40

45

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

Glu Gly Ser Arg Leu Leu Thr Tyr Val Val Lys Pro Gly Glu Thr Ser  
 65 70 75

80

Lys Ser Arg Glu Thr Lys Ala Gln Leu Glu Asp Tyr Leu Leu Val Glu  
 85 90

95

Gly Cys Thr Arg Asp Thr Val Met Val Ala Ile Gly Gly Gly Val Ile  
 100 105

110

Gly Asp Met Ile Gly Phe Val Ala Ser Thr Phe Met Arg Gly Val Arg  
 115 120

125

Val Val Gln Val Pro Thr Ser Leu Leu Ala Met Val Asp Ser Ser Ile 140  
 130 135

Gly Gly Lys Thr Ala Ile Asp Thr Pro Leu Gly Lys Asn Phe Ile Gly  
 145 150 155

160

Ala Phe Trp Gln Pro Lys Phe Val Leu Val Asp Ile Lys Trp Leu Glu  
 165 170 175

Thr Leu Ala Lys Arg Glu Phe Ile Asn Gly Met Ala Glu Val Ile Lys  
 180 185

190

Thr Ala Cys Ile Trp Asn Ala Asp Glu Phe Thr Arg Leu Glu Ser Asn

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

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|-----|
| His | Ser | Asp | Asp | Thr | Lys | His | Met | Leu | Thr | Ala | Val | His | Glu | Leu | Lys |  |     |
| 450 |     |     |     |     |     |     |     |     |     |     | 455 |     |     |     |     |  | 460 |
| Gly | Ala | Thr | Ile | Ser | Trp | Glu | Asp | Asn | Gly | Glu | Thr | Val | Val | Val | Glu |  |     |
| 465 |     |     |     |     | 470 |     |     |     |     |     |     |     |     |     | 475 |  |     |
|     |     |     |     |     | 480 |     |     |     |     |     |     |     |     |     |     |  |     |
| Gly | His | Gly | Gly | Ser | Thr | Leu | Ser | Ala | Cys | Ala | Asp | Pro | Leu | Tyr | Leu |  |     |
|     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     |     | 495 |  |     |
| Gly | Asn | Ala | Gly | Thr | Ala | Ser | Arg | Phe | Leu | Thr | Ser | Leu | Ala | Ala | Leu |  |     |
|     |     |     |     |     |     |     |     | 500 |     |     |     |     |     | 505 |     |  |     |
|     |     |     |     |     | 510 |     |     |     |     |     |     |     |     |     |     |  |     |
| Val | Asn | Ser | Thr | Ser | Ser | Gln | Lys | Tyr | Ile | Val | Leu | Thr | Gly | Asn | Ala |  |     |
|     |     |     |     |     |     |     | 515 |     |     |     |     | 520 |     |     |     |  |     |
|     |     |     |     |     | 525 |     |     |     |     |     |     |     |     |     |     |  |     |
| Arg | Met | Gln | Gln | Arg | Pro | Ile | Ala | Pro | Leu | Val | Asp | Ser | Leu | Arg | Ala |  |     |
|     | 530 |     |     |     |     |     |     |     |     |     | 535 |     |     |     |     |  | 540 |
| Asn | Gly | Thr | Lys | Ile | Glu | Tyr | Leu | Asn | Asn | Glu | Gly | Ser | Leu | Pro | Ile |  |     |
| 545 |     |     |     |     | 550 |     |     |     |     |     |     |     |     |     | 555 |  |     |
|     |     |     |     |     | 560 |     |     |     |     |     |     |     |     |     |     |  |     |
| Lys | Val | Tyr | Thr | Asp | Ser | Val | Phe | Lys | Gly | Gly | Arg | Ile | Glu | Leu | Ala |  |     |
|     |     |     |     | 565 |     |     |     |     |     |     |     |     |     | 570 |     |  |     |
|     |     |     |     | 575 |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Ala | Thr | Val | Ser | Ser | Gln | Tyr | Val | Ser | Ser | Ile | Leu | Met | Cys | Ala | Pro |  |     |
|     |     |     |     |     |     |     |     | 580 |     |     |     |     |     | 585 |     |  |     |
|     |     |     |     |     | 590 |     |     |     |     |     |     |     |     |     |     |  |     |
| Tyr | Ala | Glu | Glu | Pro | Val | Thr | Leu | Ala | Leu | Val | Gly | Gly | Lys | Pro | Ile |  |     |
|     |     |     |     |     |     |     | 595 |     |     |     |     | 600 |     |     |     |  |     |
|     |     |     |     |     | 605 |     |     |     |     |     |     |     |     |     |     |  |     |
| Ser | Lys | Leu | Tyr | Val | Asp | Met | Thr | Ile | Lys | Met | Met | Glu | Lys | Phe | Gly |  |     |
|     | 610 |     |     |     |     |     |     |     |     |     | 615 |     |     |     |     |  | 620 |
| Ile | Asn | Val | Glu | Thr | Ser | Thr | Thr | Glu | Pro | Tyr | Thr | Tyr | Tyr | Ile | Pro |  |     |
| 625 |     |     |     |     | 630 |     |     |     |     |     |     |     |     |     | 635 |  |     |
|     |     |     |     |     | 640 |     |     |     |     |     |     |     |     |     |     |  |     |
| Lys | Gly | His | Tyr | Ile | Asn | Pro | Ser | Glu | Tyr | Val | Ile | Glu | Ser | Asp | Ala |  |     |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     |     | 655 |  |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Ser | Ser | Ala | Thr | Tyr | Pro | Leu | Ala | Phe | Ala | Ala | Met | Thr | Gly | Thr | Thr |  |     |
|     |     |     |     |     |     |     |     | 660 |     |     |     |     |     | 665 |     |  |     |
|     |     |     |     |     | 670 |     |     |     |     |     |     |     |     |     |     |  |     |
| Val | Thr | Val | Pro | Asn | Ile | Gly | Phe | Glu | Ser | Leu | Gln | Gly | Asp | Ala | Arg |  |     |
|     |     | 675 |     |     |     |     |     |     |     |     |     | 680 |     |     |     |  |     |
|     |     | 685 |     |     |     |     |     |     |     |     |     |     |     |     |     |  |     |
| Phe | Ala | Arg | Asp | Val | Leu | Lys | Pro | Met | Gly | Cys | Lys | Ile | Thr | Gln | Thr |  |     |
| 690 |     |     |     |     |     |     |     |     |     |     | 695 |     |     |     |     |  | 700 |

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Ala Thr Ser Thr Thr Val Ser Gly Pro Pro Val Gly Thr Leu Lys Pro |     |
| 705                                                             | 715 |
|                                                                 | 710 |
|                                                                 | 720 |
| Leu Lys His Val Asp Met Glu Pro Met Thr Asp Ala Phe Leu Thr Ala |     |
|                                                                 | 735 |
|                                                                 | 725 |
|                                                                 | 730 |
| Cys Val Val Ala Ala Ile Ser His Asp Ser Asp Pro Asn Ser Ala Asn |     |
|                                                                 | 745 |
|                                                                 | 740 |
|                                                                 | 750 |
| Thr Thr Thr Ile Glu Gly Ile Ala Asn Gln Arg Val Lys Glu Cys Asn |     |
|                                                                 | 760 |
|                                                                 | 755 |
|                                                                 | 765 |
| Arg Ile Leu Ala Met Ala Thr Glu Leu Ala Lys Phe Gly Val Lys Thr |     |
|                                                                 | 780 |
|                                                                 | 770 |
|                                                                 | 775 |
| Thr Glu Leu Pro Asp Gly Ile Gln Val His Gly Leu Asn Ser Ile Lys |     |
|                                                                 | 800 |
|                                                                 | 790 |
|                                                                 | 795 |
| Asp Leu Lys Val Pro Ser Asp Ser Ser Gly Pro Val Gly Val Cys Thr |     |
|                                                                 | 815 |
|                                                                 | 805 |
|                                                                 | 810 |
| Tyr Asp Asp His Arg Val Ala Met Ser Phe Ser Leu Leu Ala Gly Met |     |
|                                                                 | 830 |
|                                                                 | 820 |
|                                                                 | 825 |
| Val Asn Ser Gln Asn Glu Arg Asp Glu Val Ala Asn Pro Val Arg Ile |     |
|                                                                 | 845 |
|                                                                 | 835 |
|                                                                 | 840 |
| Leu Glu Arg His Cys Thr Gly Lys Thr Trp Pro Gly Trp Trp Asp Val |     |
|                                                                 | 860 |
|                                                                 | 850 |
|                                                                 | 855 |
| Leu His Ser Glu Leu Gly Ala Lys Leu Asp Gly Ala Glu Pro Leu Glu |     |
|                                                                 | 880 |
|                                                                 | 870 |
|                                                                 | 875 |
| Cys Thr Ser Lys Lys Asn Ser Lys Lys Ser Val Val Ile Ile Gly Met |     |
|                                                                 | 895 |
|                                                                 | 885 |
|                                                                 | 890 |
| Arg Ala Ala Gly Lys Thr Thr Ile Ser Lys Trp Cys Ala Ser Ala Leu |     |
|                                                                 | 910 |
|                                                                 | 900 |
|                                                                 | 905 |
| Gly Tyr Lys Leu Val Asp Leu Asp Glu Leu Phe Glu Gln Gln His Asn |     |
|                                                                 | 925 |
|                                                                 | 915 |
|                                                                 | 920 |
| Asn Gln Ser Val Lys Gln Phe Val Val Glu Asn Gly Trp Glu Lys Phe |     |
|                                                                 | 940 |
|                                                                 | 930 |
|                                                                 | 935 |
| Arg Glu Glu Glu Thr Arg Ile Phe Lys Glu Val Ile Gln Asn Tyr Gly |     |
|                                                                 | 960 |
|                                                                 | 950 |
|                                                                 | 955 |

Asp Asp Gly Tyr Val Phe Ser Thr Gly Gly Gly Ile Val Glu Ser Ala  
 965 970  
 975  
 Glu Ser Arg Lys Ala Leu Lys Asp Phe Ala Ser Ser Gly Gly Tyr Val  
 980 985  
 990  
 Leu His Leu His Arg Asp Ile Glu Glu Thr Ile Val Phe Leu Gln Ser  
 995 1000  
 1005  
 Asp Pro Ser Arg Pro Ala Tyr Val Glu Glu Ile Arg Glu Val Trp Asn  
 1010 1015 1020  
 Arg Arg Glu Gly Trp Tyr Lys Glu Cys Ser Asn Phe Ser Phe Phe Ala  
 1025 1030 1035 1040  
 Pro His Cys Ser Ala Glu Ala Glu Phe Gln Ala Leu Arg Arg Ser Phe  
 1045 1050 1055  
 Ser Lys Tyr Ile Ala Thr Ile Thr Gly Val Arg Glu Ile Glu Ile Pro  
 1060 1065 1070  
 Ser Gly Arg Ser Ala Phe Val Cys Leu Thr Phe Asp Asp Leu Thr Glu  
 1075 1080 1085  
 Gln Thr Glu Asn Leu Thr Pro Ile Cys Tyr Gly Cys Glu Ala Val Glu  
 1090 1095 1100  
 Val Arg Val Asp His Leu Ala Asn Tyr Ser Ala Asp Phe Val Ser Lys  
 1105 1110 1115 1120  
 Gln Leu Ser Ile Leu Arg Lys Ala Thr Asp Ser Ile Pro Ile Ile Phe  
 1125 1130 1135  
 Thr Val Arg Thr Met Lys Gln Gly Gly Asn Phe Pro Asp Glu Glu Phe  
 1140 1145  
 1150  
 Lys Thr Leu Arg Glu Leu Tyr Asp Ile Ala Leu Lys Asn Gly Val Glu  
 1155 1160 1165  
 Phe Leu Asp Leu Glu Leu Thr Leu Pro Thr Asp Ile Gln Tyr Glu Val  
 1170 1175 1180  
 Ile Asn Lys Arg Gly Asn Thr Lys Ile Ile Gly Ser His His Asp Phe  
 1185 1190 1195 1200  
 Gln Gly Leu Tyr Ser Trp Asp Asp Ala Glu Trp Glu Asn Arg Phe Asn  
 1205 1210 1215  
 Gln Ala Leu Thr Leu Asp Val Asp Val Val Lys Phe Val Gly Thr Ala  
 1220 1225  
 1230  
 Val Asn Phe Glu Asp Asn Leu Arg Leu Glu His Phe Arg Asp Thr His

|                                                                 |      |           |
|-----------------------------------------------------------------|------|-----------|
| 1235                                                            |      | 1240      |
| 1245                                                            |      |           |
| Lys Asn Lys Pro Leu Ile Ala Val Asn Met Thr Ser Lys Gly Ser Ile |      |           |
| 1250                                                            | 1255 | 1260      |
| Ser Arg Val Leu Asn Asn Val Leu Thr Pro Val Thr Ser Asp Leu Leu |      |           |
| 1265                                                            | 1270 | 1275 1280 |
| Pro Asn Ser Ala Ala Pro Gly Gln Leu Thr Val Ala Gln Ile Asn Lys |      |           |
|                                                                 | 1285 | 1290 1295 |
| Met Tyr Thr Ser Met Gly Gly Ile Glu Pro Lys Glu Leu Phe Val Val |      |           |
|                                                                 | 1300 | 1305      |
| 1310                                                            |      |           |
| Gly Lys Pro Ile Gly His Ser Arg Ser Pro Ile Leu His Asn Thr Gly |      |           |
| 1315                                                            |      | 1320      |
| 1325                                                            |      |           |
| Tyr Glu Ile Leu Gly Leu Pro His Lys Phe Asp Lys Phe Glu Thr Glu |      |           |
| 1330                                                            | 1335 | 1340      |
| Ser Ala Gln Leu Val Lys Glu Lys Leu Leu Asp Gly Asn Lys Asn Phe |      |           |
| 1345                                                            | 1350 | 1355 1360 |
| Gly Gly Ala Ala Val Thr Ile Pro Leu Lys Leu Asp Ile Met Gln Tyr |      |           |
|                                                                 | 1365 | 1370 1375 |
| Met Asp Glu Leu Thr Asp Ala Ala Lys Val Ile Gly Ala Val Asn Thr |      |           |
|                                                                 | 1380 | 1385 1390 |
| Val Ile Pro Leu Gly Asn Lys Lys Phe Lys Gly Asp Asn Thr Asp Trp |      |           |
| 1395                                                            | 1400 | 1405      |
| Leu Gly Ile Arg Asn Ala Leu Ile Asn Asn Gly Val Pro Glu Tyr Val |      |           |
| 1410                                                            | 1415 | 1420      |
| Gly His Thr Ala Gly Leu Val Ile Gly Ala Gly Gly Thr Ser Arg Ala |      |           |
| 1425                                                            | 1430 | 1435 1440 |
| Ala Leu Tyr Ala Leu His Ser Leu Gly Cys Lys Lys Ile Phe Ile Ile |      |           |
|                                                                 | 1445 | 1450 1455 |
| Asn Arg Thr Thr Ser Lys Leu Lys Pro Leu Ile Glu Ser Leu Pro Ser |      |           |
|                                                                 | 1460 | 1465 1470 |
| Glu Phe Asn Ile Ile Gly Ile Glu Ser Thr Lys Ser Ile Glu Glu Ile |      |           |
| 1475                                                            | 1480 | 1485      |
| Lys Glu His Val Gly Val Ala Val Ser Cys Val Pro Ala Asp Lys Pro |      |           |
| 1490                                                            | 1495 | 1500      |
| Leu Asp Asp Glu Leu Leu Ser Lys Leu Glu Arg Phe Leu Val Lys Gly |      |           |
| 1505                                                            | 1510 | 1515 1520 |
| Ala His Ala Ala Phe Val Pro Thr Leu Leu Glu Ala Ala Tyr Lys Pro |      |           |
|                                                                 | 1525 | 1530 1535 |

Ser Val Thr Pro Val Met Thr Ile Ser Gln Asp Lys Tyr Gln Trp His  
 1540 1545 1550

Val Val Pro Gly Ser Gln Met Leu Val His Gln Gly Val Ala Gln Phe  
 1555 1560 1565

Glu Lys Trp Thr Gly Phe Lys Gly Pro Phe Lys Ala Ile Phe Asp Ala  
 1570 1575 1580

Val Thr Lys Glu  
 1585

<210> 83  
 <211> 1089  
 <212> DNA  
 <213> Escherichia coli W3110

<400> 83  
 atggagagga ttgtcgttac tctcggggaa cgtagttacc caattacat cgcacctggt 60  
 ttgtttaatg aaccagcttc attcttaccg ctgaaatcgg gcgagcaggt catgttggtc 120  
 accaacgaaa ccttggtctc tctgtatctc gataaggtcc gcggcgact tgaacaggcg 180  
 ggtgttaacg tcgatagcgt taccctccct gacggcgagc agtataaaag cctggctgta 240  
 ctcgataccg tctttacggc gttgtttaca aaaccgcatg gtcgcgatac tacgctgggtg 300  
 gcgcttggtg gcggcgtagt gggcgatctg accggcttcg cggcggcgag ttatcagcgc 360  
 ggtgtccgtt tcattcaagt cccgacgacg ttactgtcgc aggtcgattc ctccgttggtc 420  
 ggcaaaactg cgggtcaacca tcccctcgtt aaaaacatga ttggcgcggt ctaccaacct 480  
 gcttcagtgg tgggtgatct cgactgtctg aaaacgcttc ccccgcgta gttagcgtcg 540  
 gggctggcag aagtcacaa atacggcatt attcttgacg gtgcgttttt taactggctg 600  
 gaagagaatc tggatgcgtt gttgcgtctg gacggtccgg caatggcgta ctgtattcgc 660  
 cgttggtgtg aactgaaggc agaagttgtc gccgcgcgac agcgcgaaac cgggttacgt 720  
 gctttactga atctgggaca cacctttggt catgccattg aagctgaaat ggggtatggc 780  
 aattggttac atggtgaagc ggtcgtcgcg ggtatggtga tggcggcgcg gacgtcggaa 840  
 cgtctcgggc agtttagttc tgccgaaacg cagcgtatta taacctgct caagcgggct 900  
 gggttaccgg tcaatgggccc gcgcgaaatg tccgcgcagg cgtatttacc gcatatgctg 960  
 cgtgacaaga aagtccttgc gggagagatg cgcttaattc ttccgttggtc aattggtaag 1020  
 agtgaagttc gcagcggcgt ttgcgacgag cttgttctta acgccattgc cgattgtcaa 1080  
 tcagcgtaa 1089

```

<210>      84
<211>      362
<212>      PRT
<213>      Escherichia coli W3110

<400>      84
Met Glu Arg Ile Val Val Thr Leu Gly Glu Arg Ser Tyr Pro Ile Thr
  1                               5                      10
                        15

Ile Ala Ser Gly Leu Phe Asn Glu Pro Ala Ser Phe Leu Pro Leu Lys
                        20                      25
                    30

Ser Gly Glu Gln Val Met Leu Val Thr Asn Glu Thr Leu Ala Pro Leu
                        35                      40
                    45

Tyr Leu Asp Lys Val Arg Gly Val Leu Glu Gln Ala Gly Val Asn Val
  50                               55                      60

Asp Ser Val Ile Leu Pro Asp Gly Glu Gln Tyr Lys Ser Leu Ala Val
  65                               70                      75
                        80

Leu Asp Thr Val Phe Thr Ala Leu Leu Gln Lys Pro His Gly Arg Asp
                        85                      90
                    95

Thr Thr Leu Val Ala Leu Gly Gly Gly Val Val Gly Asp Leu Thr Gly
                        100                     105
                    110

Phe Ala Ala Ala Ser Tyr Gln Arg Gly Val Arg Phe Ile Gln Val Pro
                        115                     120
                    125

Thr Thr Leu Leu Ser Gln Val Asp Ser Ser Val Gly Gly Lys Thr Ala
  130                               135                      140

Val Asn His Pro Leu Gly Lys Asn Met Ile Gly Ala Phe Tyr Gln Pro
  145                               150                     155                      160

Ala Ser Val Val Val Asp Leu Asp Cys Leu Lys Thr Leu Pro Pro Arg
                        165                     170                      175

Glu Leu Ala Ser Gly Leu Ala Glu Val Ile Lys Tyr Gly Ile Ile Leu
                        180                     185
                    190

Asp Gly Ala Phe Phe Asn Trp Leu Glu Glu Asn Leu Asp Ala Leu Leu
                        195                     200
                    205

Arg Leu Asp Gly Pro Ala Met Ala Tyr Cys Ile Arg Arg Cys Cys Glu
  210                               215                      220

```

Leu Lys Ala Glu Val Val Ala Ala Asp Glu Arg Glu Thr Gly Leu Arg  
 225 230 235

240

Ala Leu Leu Asn Leu Gly His Thr Phe Gly His Ala Ile Glu Ala Glu  
 245 250 255

Met Gly Tyr Gly Asn Trp Leu His Gly Glu Ala Val Ala Ala Gly Met  
 260 265

270

Val Met Ala Ala Arg Thr Ser Glu Arg Leu Gly Gln Phe Ser Ser Ala  
 275 280

285

Glu Thr Gln Arg Ile Ile Thr Leu Leu Lys Arg Ala Gly Leu Pro Val  
 290 295

300

Asn Gly Pro Arg Glu Met Ser Ala Gln Ala Tyr Leu Pro His Met Leu  
 305 310 315

320

Arg Asp Lys Lys Val Leu Ala Gly Glu Met Arg Leu Ile Leu Pro Leu  
 325 330

335

Ala Ile Gly Lys Ser Glu Val Arg Ser Gly Val Ser His Glu Leu Val  
 340 345

350

Leu Asn Ala Ile Ala Asp Cys Gln Ser Ala  
 355

360

<210> 85  
 <211> 2247  
 <212> DNA  
 <213> Anabaena variabilis ATCC29413

<400> 85  
 atgagtatcg tccaagcaaa gtttgaagct aaggaaacat cttttcatgt agaagggttac 60  
 gaaaagattg agtatgattt ggtgtatgta gatggtattt ttgaaatcca gaattctgca 120  
 ctagcagatg tatatcaagg ttttggacga tgcttggcga ttgtagatgc taacgtcagt 180  
 cggttgtatg gtaatcaaatt tcaggcatat ttccagtatt atggtataga actgaggcta 240  
 tttcctatta ccattactga accagataag actattcaaa ctttcgagag agttatagat 300  
 gtctttgcag atttcaaatt agtccgcaaa gaaccagtat tagtcgtggg tggcgggtta 360  
 attacagatg ttgtcggctt tgcttgttct acatatcgtc gcagcagcaa ttacatccgc 420  
 attcctacta cattgattgg attaattgat gccagtgtag caattaagggt agcagttaat 480  
 catcgcaaac tgaaaaaccg tttgggtgct tatcatgctt ctcgcaaagt attttttagat 540

ttctccttgt tgcgtactct ccctacagac caagtacgta acgggatggc ggaattggta 600  
 aaaatcgctg tagtagcgca tcaagaagtt tttgaattgt tggagaagta cggcgaagaa 660  
 ttactacgta ctcatTTtTg caatatagat gcaactccag agattaaaga aatagcccat 720  
 cgtttgactt acaaagctat ccataagatg ttggaattgg aagttcccaa cctgcatgag 780  
 ttagacctag ataggggtgat tgcttacggT cacacttgga gtcccacctt ggaacttgcg 840  
 cctcgtctac ccatgttcca cggacacgcc gttaatgtag atatggcttt ctcggaacg 900  
 atcgccgccc gtagaggata tattacaatt gcagaacgcg atcgtatTTT aggattaatg 960  
 agtcgcgttg gtctatccct cgaccatccc atgttgata tagatatTTT gtggcgtggT 1020  
 actgaatcta tcacattaac tcgtgatggT ttgttaagag ctgctatgcc aaaaccatt 1080  
 ggtgatttgT tcttcgtcaa tgacctgaca agagaagaat tagcagccgc attagctgac 1140  
 cacaaagaac tttgtaccag ttatccccgt ggtggtgaag gtgtggatgt gtatcccgTt 1200  
 tatcaaaaag aattaatcgT gagtgttaaa taatgacttt tttgaattca aaatgcaaaa 1260  
 tactccacgg atacactgcg cgagcgcggT agcatttctg ttcgcggagc gtcccgtagg 1320  
 gaaagagaag gctacgcaa taatcggaca ctaattgtct ttaattttga attttgaatt 1380  
 ttgaattttg aattggagcg aagcgacttg acaaatgtga ttgtccaacc aacagctaga 1440  
 cctgttacac cattgggaat tttaaccaag cagttagaag ccatagtcca agaggTTaag 1500  
 caacatccag atttacctgg ggaattgata gcaaacatcc atcaggcttg gcgttttagcc 1560  
 gcaggTatag acccttattt ggaagaatgc accactccag aatctcctga actcgttgca 1620  
 ttggcaaaaa ccacagccac cgaagcctgg ggagaacact tccacggagg tacaaccgtc 1680  
 cgtcctctag aacaagagat gctttctggT catatcgaag gacaaacctt aaagatgTtt 1740  
 gttcacatga ccaaagctaa aaaagtotta gaaattggga tgtttaccgg ttattcggcg 1800  
 ctggcgatgg cggaagcatt accagaggat ggactgcttg tggcttTga agttgacct 1860  
 tacgcggcgg aaattggaca gaaagccttt caacaatctc cccacggTgg aaagattcgt 1920  
 gtggaattgg atgcagcctt agcaactctt gataagttag cagaagctgg ggagtctTTT 1980  
 gacttggtat ttatcgacgc agataaaaaa gagtatgtag cctatTTTca caagttgcta 2040  
 ggtagcagtt tgttagcacc agatggcttt atttgtgtag ataacacctt attacaaggg 2100  
 gaagtttatc taccagcaga ggaacgtagc gtcaatggTg aagcgatcg ccaatttaat 2160  
 catacagtag ctatagaccc ccgtgtagaa caggTTTTgt tgccgttgcg agatggTTta 2220  
 acaattatcc gcagaataca accttaa 2247

<210> 86  
<211> 820  
<212> PRT  
<213> Anabaena variabilis ATCC29413

<400> 86  
Met Ser Ile Val Gln Ala Lys Phe Glu Ala Lys Glu Thr Ser Phe His  
1 5 10  
15  
Val Glu Gly Tyr Glu Lys Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly  
20 25  
30  
Ile Phe Glu Ile Gln Asn Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe  
35 40  
45  
Gly Arg Cys Leu Ala Ile Val Asp Ala Asn Val Ser Arg Leu Tyr Gly  
50 55 60  
Asn Gln Ile Gln Ala Tyr Phe Gln Tyr Tyr Gly Ile Glu Leu Arg Leu  
65 70 75  
80  
Phe Pro Ile Thr Ile Thr Glu Pro Asp Lys Thr Ile Gln Thr Phe Glu  
85 90  
95  
Arg Val Ile Asp Val Phe Ala Asp Phe Lys Leu Val Arg Lys Glu Pro  
100 105  
110  
Val Leu Val Val Gly Gly Gly Leu Ile Thr Asp Val Val Gly Phe Ala  
115 120  
125  
Cys Ser Thr Tyr Arg Arg Ser Ser Asn Tyr Ile Arg Ile Pro Thr Thr  
130 135 140  
Leu Ile Gly Leu Ile Asp Ala Ser Val Ala Ile Lys Val Ala Val Asn  
145 150 155  
160  
His Arg Lys Leu Lys Asn Arg Leu Gly Ala Tyr His Ala Ser Arg Lys  
165 170 175  
Val Phe Leu Asp Phe Ser Leu Leu Arg Thr Leu Pro Thr Asp Gln Val  
180 185  
190  
Arg Asn Gly Met Ala Glu Leu Val Lys Ile Ala Val Val Ala His Gln  
195 200  
205  
Glu Val Phe Glu Leu Leu Glu Lys Tyr Gly Glu Glu Leu Leu Arg Thr  
210 215 220

His Phe Gly Asn Ile Asp Ala Thr Pro Glu Ile Lys Glu Ile Ala His  
 225 230 235  
 240

Arg Leu Thr Tyr Lys Ala Ile His Lys Met Leu Glu Leu Glu Val Pro  
 245 250 255

Asn Leu His Glu Leu Asp Leu Asp Arg Val Ile Ala Tyr Gly His Thr  
 260 265  
 270

Trp Ser Pro Thr Leu Glu Leu Ala Pro Arg Leu Pro Met Phe His Gly  
 275 280  
 285

His Ala Val Asn Val Asp Met Ala Phe Ser Ala Thr Ile Ala Ala Arg  
 290 295 300

Arg Gly Tyr Ile Thr Ile Ala Glu Arg Asp Arg Ile Leu Gly Leu Met  
 305 310 315  
 320

Ser Arg Val Gly Leu Ser Leu Asp His Pro Met Leu Asp Ile Asp Ile  
 325 330  
 335

Leu Trp Arg Gly Thr Glu Ser Ile Thr Leu Thr Arg Asp Gly Leu Leu  
 340 345  
 350

Arg Ala Ala Met Pro Lys Pro Ile Gly Asp Cys Val Phe Val Asn Asp  
 355 360  
 365

Leu Thr Arg Glu Glu Leu Ala Ala Ala Leu Ala Asp His Lys Glu Leu  
 370 375 380

Cys Thr Ser Tyr Pro Arg Gly Gly Glu Gly Val Asp Val Tyr Pro Val  
 385 390 395  
 400

Tyr Gln Lys Glu Leu Ile Gly Ser Val Lys Met Ser Ile Val Gln Ala  
 405 410  
 415

Lys Phe Glu Ala Lys Glu Thr Ser Phe His Val Glu Gly Tyr Glu Lys  
 420 425  
 430

Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly Ile Phe Glu Ile Gln Asn  
 435 440  
 445

Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe Gly Arg Cys Leu Ala Ile  
 450 455 460

Val Asp Ala Asn Val Ser Arg Leu Tyr Gly Asn Gln Ile Gln Ala Tyr

465 470 475  
 480  
 Phe Gln Tyr Tyr Gly Ile Glu Leu Arg Leu Phe Pro Ile Thr Ile Thr  
 485 490 495  
 Glu Pro Asp Lys Thr Ile Gln Thr Phe Glu Arg Val Ile Asp Val Phe  
 500 505  
 510  
 Ala Asp Phe Lys Leu Val Arg Lys Glu Pro Val Leu Val Val Gly Gly  
 515 520  
 525  
 Gly Leu Ile Thr Asp Val Val Gly Phe Ala Cys Ser Thr Tyr Arg Arg  
 530 535 540  
 Ser Ser Asn Tyr Ile Arg Ile Pro Thr Thr Leu Ile Gly Leu Ile Asp  
 545 550 555  
 560  
 Ala Ser Val Ala Ile Lys Val Ala Val Asn His Arg Lys Leu Lys Asn  
 565 570  
 575  
 Arg Leu Gly Ala Tyr His Ala Ser Arg Lys Val Phe Leu Asp Phe Ser  
 580 585  
 590  
 Leu Leu Arg Thr Leu Pro Thr Asp Gln Val Arg Asn Gly Met Ala Glu  
 595 600  
 605  
 Leu Val Lys Ile Ala Val Val Ala His Gln Glu Val Phe Glu Leu Leu  
 610 615 620  
 Glu Lys Tyr Gly Glu Glu Leu Leu Arg Thr His Phe Gly Asn Ile Asp  
 625 630 635  
 640  
 Ala Thr Pro Glu Ile Lys Glu Ile Ala His Arg Leu Thr Tyr Lys Ala  
 645 650  
 655  
 Ile His Lys Met Leu Glu Leu Glu Val Pro Asn Leu His Glu Leu Asp  
 660 665  
 670  
 Leu Asp Arg Val Ile Ala Tyr Gly His Thr Trp Ser Pro Thr Leu Glu  
 675 680  
 685  
 Leu Ala Pro Arg Leu Pro Met Phe His Gly His Ala Val Asn Val Asp  
 690 695 700  
 Met Ala Phe Ser Ala Thr Ile Ala Ala Arg Arg Gly Tyr Ile Thr Ile  
 705 710 715  
 720

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

Leu Asp His Pro Met Leu Asp Ile Asp Ile Leu Trp Arg Gly Thr Glu  
 740 745  
 750

Ser Ile Thr Leu Thr Arg Asp Gly Leu Leu Arg Ala Ala Met Pro Lys  
 755 760  
 765

Pro Ile Gly Asp Cys Val Phe Val Asn Asp Leu Thr Arg Glu Glu Leu  
 770 775 780

Ala Ala Ala Leu Ala Asp His Lys Glu Leu Cys Thr Ser Tyr Pro Arg  
 785 790 795  
 800

Gly Gly Glu Gly Val Asp Val Tyr Pro Val Tyr Gln Lys Glu Leu Ile  
 805 810 815

Gly Ser Val Lys  
 820

<210> 87  
 <211> 3638  
 <212> DNA  
 <213> Anabaena variabilis ATCC29413

<400> 87  
 atgagtatcg tccaagcaaa gtttgaagct aaggaaacat cttttcatgt agaaggttac 60  
 gaaaagattg agtatgattt ggtgtatgta gatggatatt ttgaaatcca gaattctgca 120  
 ctagcagatg tatatcaagg ttttggacga tgcttggcga ttgtagatgc taacgtcagt 180  
 cggttgtatg gtaatcaaat tcaggcatat ttccagtatt atggtataga actgaggcta 240  
 tttcctatta ccattactga accagataag actattcaaa ctttcgagag agttatagat 300  
 gtctttgcag atttcaaatt agtccgcaaa gaaccagtat tagtcgtggg tggcgggttta 360  
 attacagatg ttgtcggctt tgcttgttct acatatcgtc gcagcagcaa ttacatccgc 420  
 attcctacta cattgattgg attaatgat gccagtgtag caattaaggt agcagttaat 480  
 catcgcaaac tgaaaaaccg tttgggtgct tatcatgctt ctcgcaaagt attttttagat 540  
 ttctccttgt tgcgtactct ccctacagac caagtacgta acgggatggc ggaattggta 600  
 aaaatcgctg tagtagcgca tcaagaagtt tttgaattgt tggagaagta cggcgaagaa 660  
 ttactacgta ctcatTTTtg caatatagat gcaactccag agattaaaga aatagcccat 720  
 cgtttgactt acaaagctat ccataagatg ttggaattgg aagttcccaa cctgcatgag 780

ttagacctag ataggggtgat tgcttacggt cacacttgga gtcccacctt ggaacttgcg 840  
 cctcgtctac ccatgttcca cggacacgcc gttaatgtag atatggcttt ctcggaacg 900  
 atcgccgccc gtagaggata tattacaatt gcagaacgcg atcgtatfff aggattaatg 960  
 agtcgcgttg gtctatccct cgaccatccc atgttgata tagatatfff gtggcggtg 1020  
 actgaatcta tcacattaac tcgtgatggg ttgttaagag ctgctatgcc aaaaccatt 1080  
 ggtgattgtg tcttcgtcaa tgacctgaca agagaagaat tagcagccgc attagctgac 1140  
 cacaagaac tttgtaccag ttatccccgt ggtggtgaag gtgtggatgt gtatccggtt 1200  
 tatcaaaaag aattaatcgg gagtggtaaa taatgacttt tttgaattca aaatgcaaaa 1260  
 tactccacgg atacactgcg cgagcgcggt agcatttctg ttcgcggagc gtcccgtagg 1320  
 gaaagagaag gctacgcaa taatcggaca ctaattgtct ttaattttga attttgaatt 1380  
 ttgaattttg aattggagcg aagcgacttg acaaatgtga ttgtccaacc aacagctaga 1440  
 cctgttacac cattgggaat tttaaccaag cagttagaag ccatagtcca agaggtttaag 1500  
 caacatccag atttacctgg ggaattgata gcaaacatcc atcaggcttg gcgttttagcc 1560  
 gcaggatatag acccttatff ggaagaatgc accactccag aatctcctga actcgtgca 1620  
 ttggcaaaaa ccacagccac cgaagcctgg ggagaacact tccacggagg tacaaccgtc 1680  
 cgtcctctag aacaagagat gctttctggg catatcgaag gacaaacctt aaagatgttt 1740  
 gttcacatga ccaaagctaa aaaagtctta gaaattggga tgtttaccgg ttattcggcg 1800  
 ctggcgatgg cggaagcatt accagaggat ggactgcttg tggcttgga agttgaccct 1860  
 tacgcggcgg aaattggaca gaaagcctff caacaatctc cccacggtgg aaagattcgt 1920  
 gtggaattgg atgcagcctt agcaactctt gataagttag cagaagctgg ggagcttttt 1980  
 gacttggtat ttatcgacgc agataaaaaa gagtatgtag cctatffftca caagttgcta 2040  
 ggtagcagtt tgtttagcacc agatggctff atttgtgtag ataacacctt attacaaggg 2100  
 gaagtttatc taccagcaga ggaacgtagc gtcaatgggtg aagcgatcgc gcaatttaat 2160  
 catacagtag ctatagaccc ccgtgtagaa caggttttgt tgccgttgcg agatggttta 2220  
 acaattatcc gcagaatata accttaattg tccaatcgac tatggcacia tcccttcccc 2280  
 tttcttcgcg acctgctaca ccgtctcttc cttcccagac gaaaatagcc gcaattatcc 2340  
 aaaatatctg cactttgggt ttgttattac tagcattgcc cattaatgcc accattgttt 2400  
 ttatatcctt gttagtcttc cgaccgcaaa aggtcaaagc agcaaaccoc caaacattc 2460  
 ttatcagtg cggtaaagat accaaagctt tacaactagc aaggtcattc cacgcggctg 2520

gacatagagt tgtcttggtg gaaaccata aatactgggt gactgggtcat cggtttttccc 2580  
 aagcagtgga taagtttttac acagtccccg caccocagga caatccccaa gcttacattc 2640  
 aggctttggt agatatcgtc aaacaagaaa acatcgatgt ttatatcccc gtcaccagtc 2700  
 cagtgggtag ctactacgac tcattagcca aaccagagtt atcccattat tgcgaagtgt 2760  
 ttcactttga cgcagatatt acccaaagt tggatgataa atttgcggtg acacaaaaag 2820  
 cgcgatcgct tggtttatca gtacccaaat cctttaaaat tacctcacca gaacaagtca 2880  
 tcaacttcga tttttctgga gagacacgta aatacatcct caaaagcatt ccctacgact 2940  
 cagtgcggcg gttggactta accaaactcc cctgtgctac tccagaggaa acagcagcat 3000  
 tcgtcagaag tttgccatt actcccgaaa aaccgtggat tatgcaggaa tttatccccg 3060  
 gtaaggaatt ctgcacccat agcaccggtc ggaatgggga actcagactg cattgctggt 3120  
 gcgaatcttc agccttccaa gttaattatg agaatgtaaa taaccgcaa attaccgaat 3180  
 ggggtacagca ttttgtcaag gaactgaaac tgacaggaca gatttccttt gactttatcc 3240  
 aagccgaaga cggaacagtt tacgccatcg agtgaaccc ccgcacacat tcagcaatta 3300  
 ccacatttta cgaccacccc caggtagcag aagcgtactt gagtcaagca ccgacgactg 3360  
 aaaccataca accactaacg acaagcaagc ctacctattg gacttatcac gaagtgtggc 3420  
 gtttaactgg tatccgttct ttcacccagt tgcaaagatg gctggggaat atttggcgcg 3480  
 ggactgatgc gatttatcag ccagatgacc ccttaccgtt tttgatggta catcattggc 3540  
 aaattcccct actgttattg aataatttgc gtcgtcttaa aggttggacg cggatagatt 3600  
 tcaatattgg gaagtgggtg gaattggggg gagattag 3638

<210> 88  
 <211> 1099  
 <212> PRT  
 <213> Anabaena variabilis ATCC29413

<400> 88  
 Met Ser Ile Val Gln Ala Lys Phe Glu Ala Lys Glu Thr Ser Phe His  
 1 5 10

Val Glu Gly Tyr Glu Lys Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly  
 15 20 25  
 30

Ile Phe Glu Ile Gln Asn Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe  
 35 40  
 45

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

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

545 550 555  
 560  
 Ala Ser Val Ala Ile Lys Val Ala Val Asn His Arg Lys Leu Lys Asn  
 565 570  
 575  
 Arg Leu Gly Ala Tyr His Ala Ser Arg Lys Val Phe Leu Asp Phe Ser  
 580 585  
 590  
 Leu Leu Arg Thr Leu Pro Thr Asp Gln Val Arg Asn Gly Met Ala Glu  
 595 600  
 605  
 Leu Val Lys Ile Ala Val Val Ala His Gln Glu Val Phe Glu Leu Leu  
 610 615 620  
 Glu Lys Tyr Gly Glu Glu Leu Leu Arg Thr His Phe Gly Asn Ile Asp  
 625 630 635  
 640  
 Ala Thr Pro Glu Ile Lys Glu Ile Ala His Arg Leu Thr Tyr Lys Ala  
 645 650  
 655  
 Ile His Lys Met Leu Glu Leu Glu Val Pro Asn Leu His Glu Leu Asp  
 660 665  
 670  
 Leu Asp Arg Val Ile Ala Tyr Gly His Thr Trp Ser Pro Thr Leu Glu  
 675 680  
 685  
 Leu Ala Pro Arg Leu Pro Met Phe His Gly His Ala Val Asn Val Asp  
 690 695 700  
 Met Ala Phe Ser Ala Thr Ile Ala Ala Arg Arg Gly Tyr Ile Thr Ile  
 705 710 715  
 720  
 Ala Glu Arg Asp Arg Ile Leu Gly Leu Met Ser Arg Val Gly Leu Ser  
 725 730  
 735  
 Leu Asp His Pro Met Leu Asp Ile Asp Ile Leu Trp Arg Gly Thr Glu  
 740 745  
 750  
 Ser Ile Thr Leu Thr Arg Asp Gly Leu Leu Arg Ala Ala Met Pro Lys  
 755 760  
 765  
 Pro Ile Gly Asp Cys Val Phe Val Asn Asp Leu Thr Arg Glu Glu Leu  
 770 775 780  
 Ala Ala Ala Leu Ala Asp His Lys Glu Leu Cys Thr Ser Tyr Pro Arg

|                                                                 |      |      |      |      |
|-----------------------------------------------------------------|------|------|------|------|
| 785                                                             |      | 790  |      | 795  |
|                                                                 |      | 800  |      |      |
| Gly Gly Glu Gly Val Asp Val Tyr Pro Val Tyr Gln Lys Glu Leu Ile |      |      |      |      |
|                                                                 | 805  |      |      | 810  |
|                                                                 | 815  |      |      |      |
| Gly Ser Val Lys Met Thr Asn Val Ile Val Gln Pro Thr Ala Arg Pro |      |      |      |      |
|                                                                 |      | 820  |      | 825  |
|                                                                 | 830  |      |      |      |
| Val Thr Pro Leu Gly Ile Leu Thr Lys Gln Leu Glu Ala Ile Val Gln |      |      |      |      |
|                                                                 |      | 835  |      | 840  |
|                                                                 | 845  |      |      |      |
| Glu Val Lys Gln His Pro Asp Leu Pro Gly Glu Leu Ile Ala Asn Ile |      |      |      |      |
|                                                                 | 850  |      | 855  | 860  |
| His Gln Ala Trp Arg Leu Ala Ala Gly Ile Asp Pro Tyr Leu Glu Glu |      |      |      |      |
| 865                                                             |      | 870  |      | 875  |
|                                                                 |      | 880  |      |      |
| Cys Thr Thr Pro Glu Ser Pro Glu Leu Ala Ala Leu Ala Lys Thr Thr |      |      |      |      |
|                                                                 | 885  |      |      | 890  |
|                                                                 | 895  |      |      |      |
| Ala Thr Glu Ala Trp Gly Glu His Phe His Gly Gly Thr Thr Val Arg |      |      |      |      |
|                                                                 |      | 900  |      | 905  |
|                                                                 | 910  |      |      |      |
| Pro Leu Glu Gln Glu Met Leu Ser Gly His Ile Glu Gly Gln Thr Leu |      |      |      |      |
|                                                                 |      | 915  |      | 920  |
|                                                                 | 925  |      |      |      |
| Lys Met Phe Val His Met Thr Lys Ala Lys Lys Val Leu Glu Ile Gly |      |      |      |      |
|                                                                 | 930  |      | 935  | 940  |
| Met Phe Thr Gly Tyr Ser Ala Leu Ala Met Ala Glu Ala Leu Pro Glu |      |      |      |      |
| 945                                                             |      | 950  |      | 955  |
|                                                                 |      | 960  |      |      |
| Asp Gly Leu Leu Val Ala Cys Glu Val Asp Pro Tyr Ala Ala Glu Ile |      |      |      |      |
|                                                                 | 965  |      |      | 970  |
|                                                                 | 975  |      |      |      |
| Gly Gln Lys Ala Phe Gln Gln Ser Pro His Gly Gly Lys Ile Arg Val |      |      |      |      |
|                                                                 |      | 980  |      | 985  |
|                                                                 | 990  |      |      |      |
| Glu Leu Asp Ala Ala Leu Ala Thr Leu Asp Lys Leu Ala Glu Ala Gly |      |      |      |      |
|                                                                 |      | 995  |      | 1000 |
|                                                                 | 1005 |      |      |      |
| Glu Ser Phe Asp Leu Val Phe Ile Asp Ala Asp Lys Lys Glu Tyr Val |      |      |      |      |
|                                                                 | 1010 |      | 1015 | 1020 |
| Ala Tyr Phe His Lys Leu Leu Gly Ser Ser Leu Leu Ala Pro Asp Gly |      |      |      |      |
| 1025                                                            |      | 1030 |      | 1040 |
|                                                                 |      |      | 1035 |      |

Phe Ile Cys Val Asp Asn Thr Leu Leu Gln Gly Glu Val Tyr Leu Pro  
                             1045                            1050                            1055

Ala Glu Glu Arg Ser Val Asn Gly Glu Ala Ile Ala Gln Phe Asn His  
                             1060                            1065                            1070

Thr Val Ala Ile Asp Pro Arg Val Glu Gln Val Leu Leu Pro Leu Arg  
                             1075                            1080                            1085

Asp Gly Leu Thr Ile Ile Arg Arg Ile Gln Pro  
                             1090                            1095

<210> 89  
 <211> 1233  
 <212> DNA  
 <213> Anabaena variabilis ATCC29413

<400> 89  
 atgagtatcg tccaagcaaa gtttgaagct aaggaaacat cttttcatgt agaaggttac 60  
 gaaaagattg agtatgattt ggtgtatgta gatggtattt ttgaaatcca gaattctgca 120  
 ctagcagatg tatatcaagg ttttggacga tgcttggcga ttgtagatgc taacgtcagt 180  
 cggttgtatg gtaatcaaatt tcaggcatat ttccagtatt atggtataga actgaggcta 240  
 tttcctatta ccattactga accagataag actattcaaa ctttcgagag agttatagat 300  
 gtctttgcag atttcaaatt agtccgcaaa gaaccagtat tagtcgtggg tggcggttta 360  
 attacagatg ttgtcggctt tgcttgttct acatatcgtc gcagcagcaa ttacatccgc 420  
 attcctacta cattgattgg attaattgat gccagtgtag caattaaggt agcagttaat 480  
 catcgcaaac tgaaaaaccg tttgggtgct tatcatgctt ctcgcaaagt attttttagat 540  
 ttctccttgt tgcgtactct ccctacagac caagtacgta acgggatggc ggaattggta 600  
 aaaatcgctg tagtagcgca tcaagaagtt tttgaattgt tggagaagta cggcgaagaa 660  
 ttactacgta ctcatTTTTGG caatatagat gcaactccag agattaaaga aatagcccat 720  
 cgtttgactt acaaagctat ccataagatg ttggaattgg aagttcccaa cctgcatgag 780  
 ttagacctag ataggggtgat tgcttacggg cacacttggg gtcccacctt ggaacttgcg 840  
 cctcgtctac ccatgttcca cggacacgcc gttaatgtag atatggcttt ctgggcaacg 900  
 atcgccgccc gtagaggata tattacaatt gcagaacgcg atcgtatTTT aggattaatg 960  
 agtcgcgttg gtctatccct cgaccatccc atgttggata tagatatTTT gtggcggtgg 1020  
 actgaatcta tcacattaac tcgtgatggg ttgttaagag ctgctatgcc aaaaccatt 1080  
 ggtgattgtg tcttcgtcaa tgacctgaca agagaagaat tagcagccgc attagctgac 1140

cacaaagaac tttgtaccag ttatccccgt ggtggtgaag gtgtggatgt gtatccccgtt 1200  
 tatcaaaaag aattaatcgg gagtggttaa taa 1233

<210> 90  
 <211> 410  
 <212> PRT  
 <213> Anabaena variabilis ATCC29413

<400> 90  
 Met Ser Ile Val Gln Ala Lys Phe Glu Ala Lys Glu Thr Ser Phe His  
 1 5 10

15  
 Val Glu Gly Tyr Glu Lys Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly  
 20 25  
 30

Ile Phe Glu Ile Gln Asn Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe  
 35 40  
 45

Gly Arg Cys Leu Ala Ile Val Asp Ala Asn Val Ser Arg Leu Tyr Gly  
 50 55 60

Asn Gln Ile Gln Ala Tyr Phe Gln Tyr Tyr Gly Ile Glu Leu Arg Leu  
 65 70 75  
 80

Phe Pro Ile Thr Ile Thr Glu Pro Asp Lys Thr Ile Gln Thr Phe Glu  
 85 90  
 95

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

Val Leu Val Val Gly Gly Gly Leu Ile Thr Asp Val Val Gly Phe Ala  
 115 120  
 125

Cys Ser Thr Tyr Arg Arg Ser Ser Asn Tyr Ile Arg Ile Pro Thr Thr  
 130 135 140

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

His Arg Lys Leu Lys Asn Arg Leu Gly Ala Tyr His Ala Ser Arg Lys  
 165 170  
 175

Val Phe Leu Asp Phe Ser Leu Leu Arg Thr Leu Pro Thr Asp Gln Val  
 180 185  
 190

Arg Asn Gly Met Ala Glu Leu Val Lys Ile Ala Val Val Ala His Gln

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

&lt;210&gt; 91

&lt;211&gt; 840

&lt;212&gt; DNA

&lt;213&gt; Anabaena variabilis ATCC29413

&lt;400&gt; 91

ttgacaaatg tgattgtcca accaacagct agacctgtta caccattggg aattttaacc 60

aagcagttag aagccatagt ccaagagggtt aagcaacatc cagattttacc tggggaattg 120  
atagcaaaca tccatcaggc ttggcggttta gccgcaggta tagaccctta tttggaagaa 180  
tgcaccactc cagaatctcc tgaactcgct gcattggcaa aaaccacagc caccgaagcc 240  
tggggagaaac acttccacgg aggtacaacc gtccgtcctc tagaacaaga gatgctttct 300  
ggatcatatcg aaggacaaaac cttaaagatg tttgttcaca tgaccaaagc taaaaaagtc 360  
ttagaaattg ggatgtttac cggttattcg gcgctggcga tggcggaagc attaccagag 420  
gatggactgc ttgtggcttg tgaagttgac cttacgcgg cggaattgg acagaaagcc 480  
tttcaacaat ctccccacgg tggaaagatt cgtgtggaat tggatgcagc cttagcaact 540  
cttgataagt tagcagaagc tggggagtct tttgacttgg tatttatcga cgcagataaa 600  
aaagagtatg tagcctatct tcacaagttg ctaggtagca gtttgtagc accagatggc 660  
tttattttgtg tagataacac cttattacaa ggggaagttt atctaccagc agaggaacgt 720  
agcgtcaatg gtgaagcgat cgcgcaattt aatcatacag tagctataga cccccgtgta 780  
gaacagggtt tgttgccgtt gcgagatggt ttaacaatta tccgcagaat acaaccttaa 840

<210> 92  
<211> 410  
<212> PRT  
<213> Anabaena variabilis ATCC29413

<400> 92  
Met Ser Ile Val Gln Ala Lys Phe Glu Ala Lys Glu Thr Ser Phe His  
1 5 10  
15  
Val Glu Gly Tyr Glu Lys Ile Glu Tyr Asp Leu Val Tyr Val Asp Gly  
20 25  
30  
Ile Phe Glu Ile Gln Asn Ser Ala Leu Ala Asp Val Tyr Gln Gly Phe  
35 40  
45  
Gly Arg Cys Leu Ala Ile Val Asp Ala Asn Val Ser Arg Leu Tyr Gly  
50 55 60  
Asn Gln Ile Gln Ala Tyr Phe Gln Tyr Tyr Gly Ile Glu Leu Arg Leu  
65 70 75  
80  
Phe Pro Ile Thr Ile Thr Glu Pro Asp Lys Thr Ile Gln Thr Phe Glu  
85 90  
95  
Arg Val Ile Asp Val Phe Ala Asp Phe Lys Leu Val Arg Lys Glu Pro

## 41505

[illegible]

Arg Ala Ala Met Pro Lys Pro Ile Gly Asp Cys Val Phe Val Asn Asp  
355 360

365

Leu Thr Arg Glu Glu Leu Ala Ala Ala Leu Ala Asp His Lys Glu Leu  
370 375 380

Cys Thr Ser Tyr Pro Arg Gly Gly Glu Gly Val Asp Val Tyr Pro Val  
385 390 395  
400

Tyr Gln Lys Glu Leu Ile Gly Ser Val Lys  
405 410

```
<210>      93
<211>      840
<212>      DNA
<213>      Anabaena variabilis ATCC29413
```

|            |            |            |            |             |            |     |
|------------|------------|------------|------------|-------------|------------|-----|
| <400>      | 93         |            |            |             |            |     |
| ttgacaaatg | tgattgtcca | accaacagct | agacctgtta | caccattggg  | aattttaacc | 60  |
| aagcagttag | aagccatagt | ccaagaggtt | aagcaacatc | cagattttacc | tggggaattg | 120 |
| atagcaaaca | tccatcaggc | ttggcggtta | gccgcaggta | tagaccctta  | tttggaagaa | 180 |
| tgcaccactc | cagaatctcc | tgaactcgct | gcatttggca | aaaccacagc  | caccgaagcc | 240 |
| tggggagaa  | acttcacgg  | aggtacaacc | gtccgtcctc | tagaacaaga  | gatgctttct | 300 |
| ggtcatatcg | aaggacaaac | cttaaagatg | tttgttcaca | tgaccaaagc  | taaaaaagtc | 360 |
| ttagaaattg | ggatgtttac | cggttattcg | gcgctggcga | tggcggaagc  | attaccagag | 420 |
| gatggactgc | ttgtggcttg | tgaagttgac | ccttacgcgg | cggaaattgg  | acagaaagcc | 480 |
| tttcaacaat | ctccccacgg | tggaaagatt | cgtgtggaat | tggatgcagc  | cttagcaact | 540 |
| cttgataagt | tagcagaagc | tggggagtct | tttgacttgg | tattttatcg  | cgcagataaa | 600 |
| aaagagtatg | tagcctattt | tcacaagttg | ctaggttagc | gtttgttagc  | accagatggc | 660 |
| tttatttgtg | tagataaac  | cttattacaa | ggggaagttt | atctaccagc  | agaggaacgt | 720 |
| agcgtcaatg | gtgaagcgat | cgcgcaattt | aatcatacag | tagctataga  | ccccctgta  | 780 |
| gaacaggttt | tgttgccgtt | gcgagatgg  | ttaacaatta | tccgcagaat  | acaaccttaa | 840 |

|       |                               |
|-------|-------------------------------|
| <210> | 94                            |
| <211> | 279                           |
| <212> | PRT                           |
| <213> | Anabaena variabilis ATCC29413 |

<400> 94  
Met Thr Asn Val Ile Val Gln Pro Thr Ala Arg Pro Val Thr Pro Leu

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

Asp Pro Arg Val Glu Gln Val Leu Leu Pro Leu Arg Asp Gly Leu Thr  
 260 265  
 270

Ile Ile Arg Arg Ile Gln Pro  
 275

<210> 95  
 <211> 2667  
 <212> DNA  
 <213> *Anabaena variabilis* ATCC29413

<400> 95  
 atgcagacta tagatttttaa tattcgtaag ttacttgtag agtggaacgc gacccacaga 60  
 gattatgatac tttcccagag tttacatgaa ctaattgtag ctcaagtaga acgaacacct 120  
 gaggcgatcg ctgtcacctt tgacaagcaa caactaactt atcaagaact aaatcataaa 180  
 gcaaaccagc taggacatta tttacaaaca ttaggagtcc agccagaaac cctggtaggc 240  
 gtttggttag aacgttcctt agaaatgggt atctgtcttt taggaatcct caaagctggg 300  
 ggtgcttatg ttcctattga ccctgaatat cctcaagaac gcatagctta tatgctagaa 360  
 gattctcagg tgaaggtact actaactcaa gaaaaattac tcaatcaaat tccccaccat 420  
 caagcacaaa ctatctgtgt agatagggaa tgggagaaaa tttccacaca agctaatacc 480  
 aatccccaaa gtaatatataa aacggataat cttgcttatg taatttacac ctctggttcc 540  
 actggtaaac caaaagggtgc aatgaacacc caciaaggta tctgtaatcg cttattgtgg 600  
 atgcaggaag cttatcaaat cgattccaca gatagcattt tacaaaaaac cccctttagt 660  
 tttgatgttt ccgtttggga gttcttttgg actttattaa ctggcgcacg tttggtaata 720  
 gccaaaccag gcggacataa agatagtgtc tacctcatcg atttaattac tcaagaacaa 780  
 atcactacgt tgcattttgt ccctcaatg ctgcaagtgt ttttacaaaa tcgccatgta 840  
 agcaaataga gctctctaaa aagagttatt tgtagcgggtg aagctttatc tatagattta 900  
 caaaatagat ttttccagca tttgcaatgt gaattacata acctctatgg cccgacagaa 960  
 gcagcaattg atgtcacatt ttggcaatgt agaaaagata gtaatttaaa gagtgtacct 1020  
 attggtcgtc ccattgctaa tactcaaatt tatattcttg atgccgattt acaaccagta 1080  
 aatattggtg tcaactgggtga aatttatatt ggtggtgtag gggttgctcg tggttatttg 1140  
 aataaagaag aattgaccaa agaaaaattt attattaatc cttttcccaa ttctgagttt 1200  
 aagcgacttt ataaaacagg tgatttagct cgttattttac ccgatggaaa tattgaatat 1260  
 cttggtagaa cagattatca agtaaaaatt cgggggttata gaattgaaat tggcgagatt 1320

gaaaatgttt tatcttcaca cccacaagtc agagaagctg tagtcatagc gcgggatgat 1380  
 aacgctcaag aaaaacaaat catcgcttat attacctata actccatcaa acctcagctt 1440  
 gataatctgc gtgatttcct aaaagcaagg ctacctgatt ttatgattcc agccgctttt 1500  
 gtgatgctgg agcatcttcc ttttaactccc agtggttaaag tagaccgtaa ggcattacct 1560  
 aagcctgatt tatttaatta tagtgaacat aattcctatg tagcgcctcg gaatgaagtt 1620  
 gaagaaaaat tagtacaaat ctggtcgaat attctgcatt tacctaaagt aggtgtgaca 1680  
 gaaaactttt tcgctattgg tggttaattcc ctcaaagctc tacatttaatt ttctcaaatt 1740  
 gaagagttat ttgctaaaga gatatcctta gcaacacttt taacaaatcc agtaattgca 1800  
 gatttagcca aggttattca agcaaacaac caaatccata attcaccctt agttccaatt 1860  
 caaccacaag gtaagcagca gcctttcttt tgtatacatc ctgctggtgg tcatgtttta 1920  
 tgctatttta aactcgcaca atatatagga actgaccaac cattttatgg cttacaagct 1980  
 caaggatttt atggagatga agcacccttg acgcgagttg aagatatggc tagtctctac 2040  
 gtcaaaaacta ttagagaatt tcaaccccaa gggccttata gtgtcggggg gtggtcattt 2100  
 ggtggagtcg tagcttatga agtagcacag cagttacata gacaaggaca agaagtatct 2160  
 ttactagcaa tattagattc ttacgtaccg attctgctgg ataaacaaaa acccattgat 2220  
 gacgtttatt tagttggtgt tctctccaga gtttttggcg gtatgtttgg tcaagataat 2280  
 ctagtcacac ctgaagaaat agaaaattta actgtagaag aaaaaattaa ttacatcatt 2340  
 gataaagcac ggagcgctag aatattcccg cctggtgtag aacgtcaaaa taatcgccgt 2400  
 attcttgatg ttttggtggg aactttaaaa gcaacttatt cctatataag acaaccatat 2460  
 ccaggaaaag tcaactgtatt tcgagccagg gaaaaacata ttatggctcc tgaccgcacc 2520  
 ttagtttggg tagaattatt ttctgtaatg gcggctcaag aaattaagat tattgatgtc 2580  
 cctggaaaacc attattcggt tggtctagaa ccccatgtac aggttttagc acagcgttta 2640  
 caagattgtc tggaaaataa ttcataa 2667

<210> 96  
 <211> 888  
 <212> PRT  
 <213> Anabaena variabilis ATCC29413

<400> 96  
 Met Gln Thr Ile Asp Phe Asn Ile Arg Lys Leu Leu Val Glu Trp Asn  
 1 5 10

15

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

## 41505

[illegible]

Lys Val Asp Arg Lys Ala Leu Pro Lys Pro Asp Leu Phe Asn Tyr Ser  
 515 520  
 525  
 Glu His Asn Ser Tyr Val Ala Pro Arg Asn Glu Val Glu Glu Lys Leu  
 530 535 540  
 Val Gln Ile Trp Ser Asn Ile Leu His Leu Pro Lys Val Gly Val Thr  
 545 550 555  
 560  
 Glu Asn Phe Phe Ala Ile Gly Gly Asn Ser Leu Lys Ala Leu His Leu  
 565 570  
 575  
 Ile Ser Gln Ile Glu Glu Leu Phe Ala Lys Glu Ile Ser Leu Ala Thr  
 580 585  
 590  
 Leu Leu Thr Asn Pro Val Ile Ala Asp Leu Ala Lys Val Ile Gln Ala  
 595 600  
 605  
 Asn Asn Gln Ile His Asn Ser Pro Leu Val Pro Ile Gln Pro Gln Gly  
 610 615 620  
 Lys Gln Gln Pro Phe Phe Cys Ile His Pro Ala Gly Gly His Val Leu  
 625 630 635  
 640  
 Cys Tyr Phe Lys Leu Ala Gln Tyr Ile Gly Thr Asp Gln Pro Phe Tyr  
 645 650  
 655  
 Gly Leu Gln Ala Gln Gly Phe Tyr Gly Asp Glu Ala Pro Leu Thr Arg  
 660 665  
 670  
 Val Glu Asp Met Ala Ser Leu Tyr Val Lys Thr Ile Arg Glu Phe Gln  
 675 680  
 685  
 Pro Gln Gly Pro Tyr Arg Val Gly Gly Trp Ser Phe Gly Gly Val Val  
 690 695 700  
 Ala Tyr Glu Val Ala Gln Gln Leu His Arg Gln Gly Gln Glu Val Ser  
 705 710 715  
 720  
 Leu Leu Ala Ile Leu Asp Ser Tyr Val Pro Ile Leu Leu Asp Lys Gln  
 725 730  
 735  
 Lys Pro Ile Asp Asp Val Tyr Leu Val Gly Val Leu Ser Arg Val Phe  
 740 745  
 750

Gly Gly Met Phe Gly Gln Asp Asn Leu Val Thr Pro Glu Glu Ile Glu  
 755 760  
 765

Asn Leu Thr Val Glu Glu Lys Ile Asn Tyr Ile Ile Asp Lys Ala Arg  
 770 775 780

Ser Ala Arg Ile Phe Pro Pro Gly Val Glu Arg Gln Asn Asn Arg Arg  
 785 790 795  
 800

Ile Leu Asp Val Leu Val Gly Thr Leu Lys Ala Thr Tyr Ser Tyr Ile  
 805 810  
 815

Arg Gln Pro Tyr Pro Gly Lys Val Thr Val Phe Arg Ala Arg Glu Lys  
 820 825  
 830

His Ile Met Ala Pro Asp Pro Thr Leu Val Trp Val Glu Leu Phe Ser  
 835 840  
 845

Val Met Ala Ala Gln Glu Ile Lys Ile Ile Asp Val Pro Gly Asn His  
 850 855 860

Tyr Ser Phe Val Leu Glu Pro His Val Gln Val Leu Ala Gln Arg Leu  
 865 870 875  
 880

Gln Asp Cys Leu Glu Asn Asn Ser  
 885

<210> 97  
 <211> 1233  
 <212> DNA  
 <213> Nostoc punctiforme ATCC29133

<400> 97  
 atgagtaatg ttcaagcatc gtttgaagca acggaagctg aattccgcgt ggaagggttac 60  
 gaaaaaattg agtttagtct tgtctatgta aatggtgcat ttgatatcag taacagagaa 120  
 attgcagaca gctatgagaa gtttggccgt tgtctgactg tgattgatgc taatgtcaac 180  
 agattatatg gcaagcaa at caagtcatat ttagacact atggtattga tctgaccgta 240  
 gttcccattg tgattactga gcctactaaa acccttgcaa cctttgagaa aattggtgat 300  
 gctttttctg actttgggtt aatccgcaag gaaccagtat tagtagtggg tgggtggttta 360  
 accactgatg tagctgggtt tgcgtgtgct gcttaccgtc gtaagagtaa ctatattcgg 420  
 gttccgacaa cgctgattgg tctgattgat gcaggtgtag cgattaaggt tgcagtcaac 480  
 catcgcaagt taaaaaatcg cttgggtgca tatcatgctc ctttgaaagt catcctcgat 540

```

ttctccttct tgcaaacatt accaacggct caagttcgta atgggatggc agagttggtc 600
aaaattgctg ttgtggcgaa ctctgaagtc tttgaattgt tgtatgaata tggcgaagag 660
ttgctttcca ctcactttgg atatgtgaat ggtacaaagg aactgaaagc gatcgcacat 720
aaactcaatt acgaggcaat aaaaactatg ctggagttgg aaactccaaa cttgcatgag 780
ttagacctcg atcgcgtcat tgccctacggc cacacttgga gtccgacctt agaattagct 840
ccgatgatac cgttgtttca cggtcatgcc gtcaatatag acatggcttt gtctgcaact 900
attgcgggcaa gacgtggcta cattacatca ggagaacgcg atcgcathtt gagcttgatg 960
agtcgtatag gtttatcaat cgatcatcct ctactagatg gcgatttgct ctggtatgct 1020
acccaatcta ttagcttgac gcgagacggg aaacaacgcg cagctatgcc taaaccatt 1080
ggtgagtgtc tctttgtcaa cgatttcacc cgtgaagaac tagatgcagc tttagctgaa 1140
cacaaacgtc tgtgtgctac ataccctcgt ggtggagatg gcattgacgc ttacatagaa 1200
actcaagaag aatccaaact attgggagtg tga 1233

```

```

<210>      98
<211>      410
<212>      PRT
<213>      Nostoc punctiforme ATCC29133

```

```

<400>      98
Met Ser Asn Val Gln Ala Ser Phe Glu Ala Thr Glu Ala Glu Phe Arg
  1                               5                               10
                        15

```

```

Val Glu Gly Tyr Glu Lys Ile Glu Phe Ser Leu Val Tyr Val Asn Gly
                        20                               25
                30

```

```

Ala Phe Asp Ile Ser Asn Arg Glu Ile Ala Asp Ser Tyr Glu Lys Phe
  35                               40
                        45

```

```

Gly Arg Cys Leu Thr Val Ile Asp Ala Asn Val Asn Arg Leu Tyr Gly
  50                               55                               60

```

```

Lys Gln Ile Lys Ser Tyr Phe Arg His Tyr Gly Ile Asp Leu Thr Val
  65                               70                               75
                        80

```

```

Val Pro Ile Val Ile Thr Glu Pro Thr Lys Thr Leu Ala Thr Phe Glu
                        85                               90
                95

```

```

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

```

Val Leu Val Val Gly Gly Gly Leu Thr Thr Asp Val Ala Gly Phe Ala  
 115 120  
 125  
 Cys Ala Ala Tyr Arg Arg Lys Ser Asn Tyr Ile Arg Val Pro Thr Thr  
 130 135 140  
 Leu Ile Gly Leu Ile Asp Ala Gly Val Ala Ile Lys Val Ala Val Asn  
 145 150 155  
 160  
 His Arg Lys Leu Lys Asn Arg Leu Gly Ala Tyr His Ala Pro Leu Lys  
 165 170  
 175  
 Val Ile Leu Asp Phe Ser Phe Leu Gln Thr Leu Pro Thr Ala Gln Val  
 180 185  
 190  
 Arg Asn Gly Met Ala Glu Leu Val Lys Ile Ala Val Val Ala Asn Ser  
 195 200  
 205  
 Glu Val Phe Glu Leu Leu Tyr Glu Tyr Gly Glu Glu Leu Leu Ser Thr  
 210 215 220  
 His Phe Gly Tyr Val Asn Gly Thr Lys Glu Leu Lys Ala Ile Ala His  
 225 230 235  
 240  
 Lys Leu Asn Tyr Glu Ala Ile Lys Thr Met Leu Glu Leu Glu Thr Pro  
 245 250  
 255  
 Asn Leu His Glu Leu Asp Leu Asp Arg Val Ile Ala Tyr Gly His Thr  
 260 265  
 270  
 Trp Ser Pro Thr Leu Glu Leu Ala Pro Met Ile Pro Leu Phe His Gly  
 275 280  
 285  
 His Ala Val Asn Ile Asp Met Ala Leu Ser Ala Thr Ile Ala Ala Arg  
 290 295 300  
 Arg Gly Tyr Ile Thr Ser Gly Glu Arg Asp Arg Ile Leu Ser Leu Met  
 305 310 315  
 320  
 Ser Arg Ile Gly Leu Ser Ile Asp His Pro Leu Leu Asp Gly Asp Leu  
 325 330 335  
 Leu Trp Tyr Ala Thr Gln Ser Ile Ser Leu Thr Arg Asp Gly Lys Gln  
 340 345  
 350  
 Arg Ala Ala Met Pro Lys Pro Ile Gly Glu Cys Phe Phe Val Asn Asp

365 355 360  
 Phe Thr Arg Glu Glu Leu Asp Ala Ala Leu Ala Glu His Lys Arg Leu 380  
 370 375  
 Cys Ala Thr Tyr Pro Arg Gly Gly Asp Gly Ile Asp Ala Tyr Ile Glu  
 385 390 395  
 400  
 Thr Gln Glu Glu Ser Lys Leu Leu Gly Val  
 405 410

<210> 99  
 <211> 834  
 <212> DNA  
 <213> Nostoc punctiforme ATCC29133

<400> 99  
 atgaccagta ttttaggacg agataccgca agaccaataa cgccacatag cattctggta 60  
 gcacagctac aaaaaaccct cagaatggca gaggaaagta atattccttc agagatactg 120  
 acttctctgc gccaaagggtt gcaattagca gcaggtttag atccctatct ggatgattgc 180  
 actaccctg aatcgaccgc attgacagca ctagcgcaga agacaagcat tgaagactgg 240  
 agtaaacgct tcagtgatgg tgaaacagtg cgtcaattag agcaagaaat gctctcagga 300  
 catcttgaag gacaaacact aaagatgttt gtgcatatca ctaaggctaa aagcatccta 360  
 gaagtgggaa tggtcacagg atattcagct ttggcaatgg cagaggcggt accagatgat 420  
 gggcgactga ttgcttgtga agtagactcc tatgtggccg agtttgctca aacttgcttt 480  
 caagagtctc cccacggccg caagattgtt gtagaagtgg cacctgcact agagacgctg 540  
 cacaagctgg tggctaataa agaatccttt gatctgatct tcattgatgc ggataaaaag 600  
 gagtatatag aatacttcca aattatcttg gatagccatt tactagctcc cgacggatta 660  
 atctgtgtag ataatacttt gttgcaagga caagtttacc tgccatcaga acagcgtaca 720  
 gccaatggtg aagcgatcgc tcaattcaac cgcattgtcg ccgcagatcc tcgtgtagag 780  
 caagttctgt taccatacag agatgggtata accctgatta gacgcttggt ataa 834

<210> 100  
 <211> 277  
 <212> PRT  
 <213> Nostoc punctiforme ATCC29133

<400> 100  
 Met Thr Ser Ile Leu Gly Arg Asp Thr Ala Arg Pro Ile Thr Pro His  
 1 5 10  
 15

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

260

265

270

Ile Arg Arg Leu Val  
275

<210> 101  
<211> 1386  
<212> DNA  
<213> Nostoc punctiforme ATCC29133

<400> 101  
atggcacaat caatctcttt atctcttcct caatccacaa cgccatcaaa ggggtgtgagg 60  
ctaaaaatag cagctctact gaagactatc gggactctaa tactactgct gatagccttg 120  
ccgcttaatg ctttgatagt attgatatct ctgatgtgta ggccggtttac aaaaaaacct 180  
gccgtagcca ctcatcccca gaatatcttg gtcagtggcg gcaaaatgac caaagcattg 240  
caacttgccc gtcctttcca tgcagccgga cacagagtta ttctgattga aggtcataaa 300  
tactggttat cagggcatag attctcaaatt tctgtgagtc gtttttatac agttcctgca 360  
ccacaagacg acccagaagg ctatacccaa gcgctattgg aaattgtcaa acgagagaag 420  
attgacgttt atgtaccggt atgcagccct gtagctagtt attacgactc tttggcaaag 480  
tctgcactat cagaatattg tgagggtttt cactttgatg ctgatataac caagatgctg 540  
gatgataaat ttgcctttac cgatcgggcy cgatcgcttg gtttatcagc cccgaaatct 600  
tttaaaatta ccgatcctga acaagttatc aacttcgatt ttagtaaaga gacgcgcaaa 660  
tatattctta agagtatttc ttacgactca gttcgccgct taaatttaac caaacttcct 720  
tgtgataccc cagaagagac agcagcgttt gtcaagagtt taccatcag cccagaaaaa 780  
ccttggaatta tgcaagaatt tattcctggg aaagaattat gcacccatag cacagtccga 840  
gacggcgaat taagggttgca ttgctgttca aattcttcag cgtttcagat taactatgaa 900  
aatgtcgaaa atccccaaat tcaagaatgg gtacaacatt tcgtcaaaag tttacggctg 960  
actggacaaa tatctcttga ctttatccaa gctgaagatg gtacagctta tgccattgaa 1020  
tgtaatcctc gcacccattc ggcgatcaca atgttctaca atcaccagc tgttgagaa 1080  
gcctatcttg gtaaaactcc tctagctgca cttttggaac ctcttgagaa tagcaagccc 1140  
acttactgga tatatcacga aatctggcga ttgactggga ttcgctctgg acaacaatta 1200  
caaacttggg ttgggagatt agtcagaggt acagatgcca tttatcgctt ggacgatccg 1260  
ataccatttt taactttgca ccattggcag attactttac ttttgctaca aaatttgcaa 1320  
cgactcaaag gttgggtaaa gatcgatttt aatatcggtt aactcgtgga attagggggc 1380

gactaa

1386

<210> 102  
 <211> 461  
 <212> PRT  
 <213> Nostoc punctiforme ATCC29133

<400> 102  
 Met Ala Gln Ser Ile Ser Leu Ser Leu Pro Gln Ser Thr Thr Pro Ser  
 1 5 10

15  
 Lys Gly Val Arg Leu Lys Ile Ala Ala Leu Leu Lys Thr Ile Gly Thr  
 20 25  
 30

Leu Ile Leu Leu Leu Ile Ala Leu Pro Leu Asn Ala Leu Ile Val Leu  
 35 40  
 45

Ile Ser Leu Met Cys Arg Pro Phe Thr Lys Lys Pro Ala Val Ala Thr  
 50 55 60

His Pro Gln Asn Ile Leu Val Ser Gly Gly Lys Met Thr Lys Ala Leu  
 65 70 75  
 80

Gln Leu Ala Arg Ser Phe His Ala Ala Gly His Arg Val Ile Leu Ile  
 85 90  
 95

Glu Gly His Lys Tyr Trp Leu Ser Gly His Arg Phe Ser Asn Ser Val  
 100 105  
 110

Ser Arg Phe Tyr Thr Val Pro Ala Pro Gln Asp Asp Pro Glu Gly Tyr  
 115 120  
 125

Thr Gln Ala Leu Leu Glu Ile Val Lys Arg Glu Lys Ile Asp Val Tyr  
 130 135 140

Val Pro Val Cys Ser Pro Val Ala Ser Tyr Tyr Asp Ser Leu Ala Lys  
 145 150 155  
 160

Ser Ala Leu Ser Glu Tyr Cys Glu Val Phe His Phe Asp Ala Asp Ile  
 165 170  
 175

Thr Lys Met Leu Asp Asp Lys Phe Ala Phe Thr Asp Arg Ala Arg Ser  
 180 185  
 190

Leu Gly Leu Ser Ala Pro Lys Ser Phe Lys Ile Thr Asp Pro Glu Gln

-132-

445 435 440

Asp Phe Asn Ile Gly Lys Leu Val Glu Leu Gly Gly Asp  
450 455 460

<210> 103  
<211> 1047  
<212> DNA  
<213> Nostoc punctiforme ATCC29133

<400> 103  
atgccagttac ttaatatcct tcatttagtt gggctctgcac acgataagtt ttactgtgat 60  
ttatcacgtc tttatgcca agactgttta gctgcaacag cagatccatc gctttataac 120  
tttcaaattg catatatcac acccgatcgg cagtggcgat ttcctgactc tctcagtcga 180  
gaagatattg ctcttaccaa accgattcct gtgtttgatg ccatacaatt tctaacaggc 240  
caaaacattg acatgatggt accacaaatg ttttgtattc ctggaatgac tcagtaccgt 300  
gccctattcg atctgctcaa gatcccttat ataggaaata cccagatat tatggcgatc 360  
gcggcccaca aagccagagc caaagcaatt gtcgaagcag caggggtaaa agtgcctcgt 420  
ggagaattgc ttcgccaagg agatattcca acaattacac ctccagcagt cgtcaaacct 480  
gtaagttctg acaactcttt aggagtagtc ttagttaaag atgtgactga atatgatgct 540  
gccttaaaga aagcatttga atatgcttcg gaggtcatcg tagaagcatt catcgaactt 600  
ggtcgagaag tcagatgcgg catcattgta aaagacggtg agctaataagg tttaccctt 660  
gaagagtatc tggtagacct acacgataaa cctatccgta actatgctga taaactccaa 720  
caaactgacg atggcgactt gcatttgact gctaaagata atatcaaggc ttggatttta 780  
gaccctaacg acccaatcac ccaaagggt cagcaagtgg ctaaaagggtg tcatcaggct 840  
ttgggttgtc gccactacag tttatttgac ttccgaatcg atccaaaggg acaaccttgg 900  
ttcttagaag ctggattata ttgttctttt gcccccaaaa gtgtgatttc ttctatggcg 960  
aaagcagccg gaatccctct aaatgattta ttaataaccg ctattaatga aacattgggt 1020  
agtaataaaa aggtgttaca aaattga 1047

<210> 104  
<211> 348  
<212> PRT  
<213> Nostoc punctiforme ATCC29133

<400> 104  
Met Pro Val Leu Asn Ile Leu His Leu Val Gly Ser Ala His Asp Lys

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

Ala Trp Ile Leu Asp Pro Asn Asp Pro Ile Thr Gln Lys Val Gln Gln  
 260 265

270

Val Ala Lys Arg Cys His Gln Ala Leu Gly Cys Arg His Tyr Ser Leu  
 275 280

285

Phe Asp Phe Arg Ile Asp Pro Lys Gly Gln Pro Trp Phe Leu Glu Ala  
 290 295

300

Gly Leu Tyr Cys Ser Phe Ala Pro Lys Ser Val Ile Ser Ser Met Ala  
 305 310 315

320

Lys Ala Ala Gly Ile Pro Leu Asn Asp Leu Leu Ile Thr Ala Ile Asn  
 325 330 335

Glu Thr Leu Gly Ser Asn Lys Lys Val Leu Gln Asn  
 340

345